



UNIVERSIDADE
ESTADUAL de LONDRINA

EVERTON GERALDO CAPOTE FERREIRA

**“INFERÊNCIA FUNCIONAL DE GENES HIPOTÉTICOS DA
ILHA SIMBIÓTICA DA ESTIRPE CPAC 15 DE *Bradyrhizobium*
japonicum”**

EVERTON GERALDO CAPOTE FERREIRA

**“INFERÊNCIA FUNCIONAL DE GENES HIPOTÉTICOS DA
ILHA SIMBIÓTICA DA ESTIRPE CPAC 15 DE *Bradyrhizobium*
japonicum”**

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

Orientador: Dr. Fernando Gomes Barcellos.
Co-orientadora: Dra. Mariangela Hungria.

Londrina
2016

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UEL

Ferreira, Everton Geraldo Capote.

Inferência Funcional de Genes Hipotéticos da Ilha Simbiótica da Estirpe CPAC 15 de *Bradyrhizobium japonicum* / Everton Geraldo Capote Ferreira. - Londrina, 2016.
108 f.

Orientador: Fernando Gomes Barcellos.

Coorientador: Mariangela Hungria .

Dissertação (Mestrado em Genética e Biologia Molecular) - Universidade Estadual de Londrina, Centro de Ciências Biológicas, Programa de Pós-Graduação em Genética e Biologia Molecular, 2016.

Inclui bibliografia.

1. Fixação Biológica de Nitrogênio - Teses. 2. Rizóbios - Teses. 3. Análise da Expressão - Teses. 4. Ilha Simbiótica - Teses. I. Gomes Barcellos, Fernando . II. , Mariangela Hungria. III. Universidade Estadual de Londrina. Centro de Ciências Biológicas. Programa de Pós-Graduação em Genética e Biologia Molecular. IV. Título.

EVERTON GERALDO CAPOTE FERREIRA

**“INFERÊNCIA FUNCIONAL DE GENES HIPOTÉTICOS DA ILHA
SIMBIÓTICA DA ESTIRPE CPAC 15 DE *Bradyrhizobium japonicum*”**

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

BANCA EXAMINADORA

Orientador: Prof. Dr. Fernando Gomes Barcellos
Universidade Estadual de Londrina – UEL

Co-orientadora: Dra. Mariangela Hungria
Universidade Estadual de Londrina – UEL

Prof^ª. Dra. Jesiane Stêfania da Silva Batista
Universidade Estadual de Ponta Grossa – UEPG

Prof^ª Dra. Elisete Pains Rodrigues
Universidade Estadual de Londrina – UEL

Londrina, 29 de fevereiro de 2016.

Dedico
Sempre, sempre e sempre, as mulheres da minha vida, Zeni, mãe.
Cris, irmã
Anair, vó

AGRADECIMENTOS

Eu, que sempre fui fadado às palavras, a poesia, as coisas que a vida nos dá e nos tira, sempre esperei pelo momento de fazer os devidos agradecimentos a todos que de alguma forma foram importantes nessa jornada concluída. Poderia escrever cinco páginas tranquilamente de agradecimentos, mas vou ser sucinto, o que é até difícil, afinal, as palavras sempre escapam de mim sem permissão.

Queria agradecer, primeiramente, a Deus, ao Universo, as forças que nos regem, a aquilo que cada ser humano registra e coloca como ser divino em suas vidas, com toda a certeza, isso faz a diferença. Nesse mesmo parágrafo, agradeço a minha família, que acreditou que o menino doido com sonhos gigantes seria alguém, muito obrigado. À mãe, pelo suporte, carinho e principalmente pelos puxões de orelha. À minha irmã, que nunca mediu esforços para me ajudar e a minha vó, que é a minha doçura, a minha força para viver, a luz dos meus olhos.

À orientação e carinho, do meu querido orientador Fernando, obrigado por ser sempre essa pessoa atenciosa, querida, generosa, de fala mansa, obrigado, por fazer essa jornada prazerosa, por sempre, me orientar e auxiliar no que eu mais precisava e pelas conversas amigas.

À Mariangela Hungria, minha co-orientadora, por ser essa pessoa extremamente iluminada, sempre será meu espelho, a sua dedicação ao seu trabalho é admirável, a sua força e foco nos seus objetivos, com toda a certeza, são incríveis e dignos de nota. Obrigado por me entender, por sempre puxar esse menino que vive no mundo da Lua, com uma mochila cheia de sonhos, “Ton, você precisa ter foco”, eu sei, é complicado, sou sagitariano, é culpa do signo, é o que dizem. Obrigado por assim como eu, ser alguém de alma, e pessoas de alma, se reconhecem. Apesar de ser objetiva, é muito bonito isso de ser abstrato, e você é os dois na medida certa. Meu muito obrigado, do menino que às vezes, viaja na maionese, mas de bom coração.

À todos os integrantes e amigos do Laboratório de Biotecnologia de Solos, todos incríveis, queridos, amáveis. Em especial, ao Douglas, que me auxiliou em cada etapa desse trabalho, sem dúvida, esse trabalho é tanto meu quanto seu. Ao pessoal do Laboratório de Biotecnologia Vegetal, pessoas incríveis que tive o prazer de conhecer nessa jornada.

À CAPES pela concessão da bolsa e à todos os meus amigos de Londrina, que arrancaram de mim, o meu melhor, os meus sorrisos, alegrias, histórias e vivências, meu muito obrigado.

“Nas pequenas lutas, fazem-se grandes ações. Há bravuras persistentes e ignoradas que se defendem corajosamente na sombra contra a invasão fatal das necessidades e torpezas. Nobres e misteriosos triunfos que nenhum olhar vê, que nenhum renome paga, que nenhuma fanfarra saúda. A vida, a desgraça, o isolamento, o abandono, a pobreza, são campos de batalha que também têm seus heróis obscuros, maiores talvez que muito herói ilustre.”

— Victor Hugo, no livro “Os miseráveis”

FERREIRA, Everton Geraldo Capote. **Inferência Funcional de Genes Hipotéticos da Ilha Simbiótica da Estirpe CPAC 15 de *Bradyrhizobium japonicum***. 2016. 108p. Dissertação (Mestrado em Genética e Biologia Molecular) – Universidade Estadual de Londrina, Londrina, 2016.

RESUMO

A fixação biológica do nitrogênio (FBN) na soja (*Glycine max* (L.) Merr) ocorre através da simbiose com bactérias (rizóbios), como a espécie *Bradyrhizobium japonicum*, seu principal simbiote. Recentemente após o sequenciamento do genoma da estirpe CPAC 15 de *Bradyrhizobium japonicum*, utilizada como inoculante comercial em soja, observou-se a presença de aproximadamente 50% de genes codificando para proteínas hipotéticas. Neste contexto, o estudo teve como objetivo a análise da expressão de genes localizados dentro da ilha simbiótica e que codificam para proteínas hipotéticas e que possam estar relacionadas à FBN e a atividade saprofítica. A inferência funcional foi realizada com base em busca de similaridades com diversos bancos de dados e programas de predições de estruturas como peptídeos sinais, localização celular, domínios conservados, entre outros. A avaliação da expressão gênica foi realizada através de RT-qPCR com indução com genisteína em meio de cultivo com a bactéria, por aproximadamente, 48 horas. As análises de similaridade com o banco de dados RefSeq demonstrou que todas as proteínas hipotéticas estavam distribuídas em vários genomas de diferentes espécies do gênero *Bradyrhizobium*. Os resultados obtidos por RT-qPCR demonstraram que todos os genes hipotéticos selecionados obtiveram elevados níveis de expressão em presença de genisteína em relação ao controle, o metanol. A maioria dos genes codificando para proteínas hipotéticas inferidos funcionalmente estava relacionada ao transporte de substâncias, vias metabólicas de aminoácidos e degradação de compostos, e possivelmente então, podendo ser úteis na atividade saprofítica da estirpe e na FBN em si.

Palavras-chave: Proteínas hipotéticas. Genisteína. Ilha simbiótica. FBN. Anotação funcional. RT-qPCR.

FERREIRA, Everton Geraldo Capote. **Functional inference of hypothetical genes of the symbiotic island of the strain CPAC 15 of *Bradyrhizobium japonicum***. 2016. 108p. Dissertation (Master's degree in Genetics and Molecular Biology) – Universidade Estadual de Londrina, Londrina, 2016.

ABSTRACT

The biological nitrogen fixation (BNF) in soybean (*Glycine max* (L.) Merr) occurs through symbiosis with bacteria (rhizobia), as *Bradyrhizobium japonicum* species, its main symbiont. Recently, after the sequencing of the genome of strain CPAC 15 *B. japonicum* utilized as inoculant for soybean was observed the presence of approximately 50% of genes coding for putative proteins. In this context, the study aimed to analyze the expression of genes located within the symbiotic island and which encode hypothetical proteins that are possibly associated with BNF. The functional inference was performed based on the search for similarities with various databases and structural prediction programs such as signal peptides, among others. The evaluation of gene expression was performed by RT-qPCR, in this assay, genistein induction was used in culture media with the bacteria for approximately 48 hours. The analysis of similarity with the *RefSeq* database showed that all putative proteins were distributed in several genomes of different species of the genus *Bradyrhizobium*. The results obtained by RT-qPCR showed that all selected hypothetical genes had high levels of expression in the presence of genistein compared to control (methanol). Most hypothetical genes that have been inferred were functionally related to the transport of substances, metabolic pathways of amino acids and degradation compounds, leading to the assumption that these genes may be useful in saprophytic activity of the strain and FBN process.

Keywords: Hypothetical proteins. Genistein. Symbiotic Island BNF. Functional annotation. RT-qPCR.

LISTA DE FIGURAS

CAPÍTULO 1

- Figura 1** - Etapas do estabelecimento da simbiose entre leguminosas e rizóbios 18

CAPÍTULO 2

- Figura 1** - Localização do gene *bjs08317* no genoma da estirpe CPAC 15 de *B. japonicum*. Em preto, representado o gene *bjs08317*, em amarelo o gene *nopP* que está relativamente próximo no contexto genômico e que tem função conhecida, e os números em azul representam a identificação do gene no NCBI. 56
- Figura 2** - Localização do gene *bjs08216* no genoma da estirpe CPAC 15 de *B. japonicum*. Em preto, representando o gene *bs08216*, em amarelo o gene de fixação *fixW*, em azul o número de identificação do gene no NCBI. 57
- Figura 3** - Localização dos genes *bjs08251*, *bjs08254*, *bjs08258*, *bjs08261* e *bjs08267* no genoma da estirpe CPAC 15 de *B. japonicum*. O gene *bjs08258* foi anotado no NCBI como o gene *aspC*. Em preto estão representados os genes selecionados, em amarelo o gene de nodulação *nodU*, anotado no NCBI como *bjs08257*, em azul o número de identificação de cada gene no NCBI. 58
- Figura 4** - Integridade das amostras de RNA extraídas dos inóculos. A primeira banda corresponde ao 23S rRNA, a segunda banda ao 16S rRNA e o arraste presente entre as bandas e após elas correspondem aos RNA mensageiros de cada amostra. A última e mais forte banda corresponde ao excedente de tampão de amostra utilizado na corrida eletroforética. 65
- Figura 5** - Gel de agarose com as amostras de RNA tratadas com DNase..... 66
- Figura 6** - Níveis de expressão de genes codificando proteínas hipotéticas de *B. japonicum* Estirpe CPAC 15 após crescimento até fase exponencial em presença de genisteína. Dados de expressão mostrados para cada gene representam a média das três réplicas biológicas. Dados foram normalizados em relação ao controle endógeno (16S rRNA). Todos os

genes obtiveram uma expressão estaticamente significativa ao nível de 5% determinada pelo software *REST2009*. 68

Figura 7 - Distribuição das proteínas hipotéticas em diferentes espécies do género *Bradyrhizobium*. Os dados separados por gene e utilizados para a criação dos gráficos encontram-se no Material Suplementar..... 70

LISTA DE TABELAS

CAPÍTULO 2

Tabela 1 - Sequências dos <i>primers</i> dos genes utilizados no RT-qPCR.....	53
Tabela 2 - Tamanho dos genes hipotéticos selecionados da ilha simbiótica da estirpe CPAC 15.....	55
Tabela 3 - Quantificação da concentração de RNA extraído via Nanodrop.....	65
Tabela 4 - Quantificação da concentração de RNA tratado com DNase via Nanodrop.....	66

SUMÁRIO

1	INTRODUÇÃO	13
2	REVISÃO BIBLIOGRÁFICA	15
2.1	A FIXAÇÃO BIOLÓGICA DE NITROGÊNIO NA CULTURA DA SOJA.....	15
2.1.1	A Importância Econômica e Ambiental da Fixação Biológica de Nitrogênio em Soja no Cenário Nacional.....	19
2.2	AS ESTIRPES RECOMENDADAS PARA INOCULANTES COMERCIAIS EM SOLOS BRASILEIROS.....	20
2.3	CARACTERÍSTICAS GERAIS DAS ESTIRPES DE <i>Bradyrhizobium japonicum</i>	21
2.4	GENOMA DA ESTIRPE CPAC 15 DE <i>Bradyrhizobium japonicum</i>	23
2.5	GENÔMICA FUNCIONAL.....	25
2.6	ANÁLISES TRANSCRICIONAIS EM <i>Bradyrhizobium japonicum</i> E OUTRAS ESTIRPES MICROSSIMBIONTES DA SOJA.....	27
2.7	ANOTAÇÃO FUNCIONAL DE PROTEÍNAS HIPOTÉTICAS.....	30
3	OBJETIVOS	37
3.1	OBJETIVO GERAL.....	37
3.2	OBJETIVOS ESPECÍFICOS.....	37
4	REFERÊNCIAS BIBLIOGRÁFICAS	38
CAPÍTULO 2		47
INTRODUÇÃO.....		49
MATERIAL E MÉTODOS.....		50
ESTIRPE BACTERIANA E CONDIÇÕES DE CULTIVO.....		50
EXTRAÇÃO E QUANTIFICAÇÃO DO RNA TOTAL.....		50
ENSAIO DA EXPRESSÃO DOS GENES EM RESPOSTA A GENISTEÍNA.....		51
TRATAMENTO COM DNase-I E SÍNTESE DO DNA COMPLEMENTAR (cDNA).....		52
SELEÇÃO E ANOTAÇÃO FUNCIONAL DAS SEQUÊNCIAS ALVO CODIFICADORAS DE PROTEÍNAS HIPOTÉTICAS DA ILHA SIMBIÓTICA.....		52
DESENHO DOS <i>PRIMERS</i> PARA AS SEQUENCIAS ALVO SELECIONADAS.....		53
ANÁLISE DE EXPRESSÃO POR RT-qPCR (PCR QUANTITATIVO EM TEMPO REAL).....		54
RESULTADOS.....		55
ANÁLISE IN SILICO DOS GENES CODIFICANDO PARA PROTEÍNAS HIPOTÉTICAS.....		55

EXTRAÇÃO DE RNA, DIGESTÃO COM DNase E VERIFICAÇÃO DA QUALIDADE, INTEGRIDADE E PUREZA.....	64
EXPRESSÃO RELATIVA DOS GENES HIPOTÉTICOS DA ILHA SIMBIÓTICA.....	67
CONSERVAÇÃO DAS PROTEÍNAS HIPOTÉTICAS EM OUTROS GENOMAS DE ESTIRPES DE <i>Bradyrhizobium</i>	69
DISCUSSÃO.....	70
CONSIDERAÇÕES FINAIS.....	84
REFERÊNCIAS BIBLIOGRÁFICAS	85
7. MATERIAL SUPLEMENTAR.....	95

1 INTRODUÇÃO

Diversas espécies de plantas da família Leguminosae (= Fabaceae) são capazes de estabelecer simbiose com um grupo de bactérias, grande parte, da ordem Rhizobiales, no qual estão incluídos espécies do gênero *Allorhizobium*, *Azorhizobium*, *Mesorhizobium*, *Sinorhizobium*, *Rhizobium* e *Bradyrhizobium*. (WILLENS, 2006). Estas bactérias coletivamente chamadas de rizóbios conseguem fixar o nitrogênio atmosférico (N₂) no interior de estruturas altamente especializadas, os nódulos, e assim o disponibilizando de uma forma assimilável (amônia-NH₃) para as plantas (BUCKER MORAES et al., 2010). Desta forma, a Fixação Biológica de Nitrogênio (FBN) constitui-se como uma alternativa sustentável ao manejo do solo, pois ao substituir parcialmente ou totalmente (dependendo da cultura agrícola) a adubação mineral nitrogenada, acarreta na minimização de danos ambientais, além de contribuir para a diminuição nos custos de produção e aumento da produtividade (HUNGRIA et al., 2007).

A FBN é um componente essencial para a sustentabilidade agrícola, tanto que inoculantes microbianos vêm sendo aplicados frequentemente como biofertilizantes, caracterizam-se como uma tecnologia barata, custo de U\$ 2-3 ha⁻¹ em simbiose com leguminosas (HUNGRIA et al., 2006a). Globalmente, estima-se que a FBN contribui com 65% da incorporação de N na biosfera, sendo o maior provedor desse nutriente para a manutenção da vida no nosso planeta (HUNGRIA et al., 2001).

A principal utilização do grão de soja é na alimentação humana e animal, devido ao elevado teor proteico nos grãos. Por consequência, a soja requer uma demanda elevada de nitrogênio para seu desenvolvimento e para sua produtividade. O processo de FBN pode fornecer toda a demanda de nitrogênio requerida pela soja, desde que respeitados os procedimentos para uma boa nodulação (HUNGRIA et al., 2006a).

Existe uma grande variabilidade entre as estirpes que nodulam a soja, quanto à eficiência do processo simbiótico e à capacidade competitiva com bactérias estabelecidas no solo (ARAÚJO; HUNGRIA, 1999; HUNGRIA et al., 2006; SILVA BATISTA et al., 2007; CHEN et al., 2015). Somado a isto, fatores como o constante aprimoramento das práticas agrícolas, lançamento de cultivares mais produtivas e adaptadas as distintas condições ambientais, além do estabelecimento de populações de bactérias em solos previamente inoculados, fazem com que seja extremamente necessário uma constante

busca por genótipos de *Bradyrhizobium* mais competitivos no solo e eficientes no processo simbiótico com a soja.

Atualmente, recentes técnicas de sequenciamento em larga escala, *next generation sequencing* (NGR) empregadas em projetos genoma e transcriptoma vêm produzindo um enorme volume de sequências de genes e proteínas com funções ainda desconhecidas. As proteínas hipotéticas identificadas a partir de sequências nucleotídicas (*ORF – Open Reading Frame*) nestes estudos têm sua existência predita, sem evidência experimental a nível transcricional ou proteômico (METZKER, 2010). O processo de anotação caracteriza-se como a atribuição de funções a essas sequências, em um processo que requer a integração de uma variedade de métodos de predição de funções, além do conhecimento biológico de proteínas e do organismo em estudo, para uma anotação confiável (SOUZA, 2014).

Recentemente, a estirpe CPAC 15 de *Bradyrhizobium japonicum* teve o seu genoma sequenciado (SIQUEIRA et al., 2014), e durante a anotação deste genoma observou-se a existência de aproximadamente 50% de genes hipotéticos, isto é, sem validações experimentais, apenas predições, incluindo genes localizados na “ilha simbiótica”. Uma região ou ilha simbiótica corresponde a um agrupamento de genes simbióticos que se integraram no genoma após eventos como transferência gênica horizontal a partir de linhagens diferentes. Regiões simbióticas são primariamente passíveis da ocorrência de rearranjos de DNA. No entanto, ao longo do processo evolutivo, na ausência de rearranjos genômicos, esta região do genoma acabou por conter a maioria dos genes que são envolvidos na simbiose em diferentes rizóbios (KANEKO et al., 2000; GÖTTTFERT et al., 2001). Ao passo que após o sequenciamento da estirpe CPAC 15, 50% dos genes são hipotéticos, intensos estudos são necessários tanto a nível transcricional quanto a nível proteômico a fim de elucidar as funções que estes genes desempenham nesta estirpe, e desta forma, lançar perspectivas para um aumento da compreensão das características superiores que este rizóbio apresenta.

Portanto, este estudo ao analisar a expressão gênica por RT-qPCR de genes codificando para proteínas hipotéticas da ilha simbiótica de *B. japonicum* CPAC 15 é de grande relevância, ao proporcionar avanços no conhecimento genético desta estirpe, lançando perspectivas de identificação de novos genes com funções interessantes para este rizóbio, que possam estar relacionadas as suas características superiores de competitividade no solo.

2 REVISÃO BIBLIOGRÁFICA

2.1 A FIXAÇÃO BIOLÓGICA DE NITROGÊNIO NA CULTURA DA SOJA

A soja (*Glycine max* (L). Merrill) é uma planta pertence à família *Fabaceae* (=Leguminosae) e é de grande importância nutricional e econômica no mundo. É considerada uma das plantas cultivadas mais antigas no mundo, com relatos da literatura chinesa sobre a cultura datando de 2.500 anos A.C. A região central da China seria o provável centro genético primário da soja e, com a movimentação da população para a Manchúria, esta região tornou-se o centro genético secundário (QIU e CHANG, 2010).

A soja tornou-se uma importante *commodity* (planta de importância comercial agrícola produzida em larga escala, principalmente para exportação) devido a sua grande adaptação geográfica, composição química, alto valor nutricional, benefícios funcionais à saúde e aplicações industriais, assim como o grande potencial para produção e utilização da mesma e de seus derivados nos alimentos, como forragem e aplicações farmacêuticas por todo o planeta. Durante a primeira metade do século XX, a China foi o maior produtor e exportador de soja no mundo. Já na década de 1950 a produção de soja avançou rapidamente nos Estados Unidos, atualmente o maior produtor mundial de grãos dessa leguminosa (ALI, 2000).

Em um panorama nacional, embora a soja seja uma cultura exótica introduzida no Brasil há pouco mais de um século, hoje ela está estabelecida em quase todo o território nacional, inclusive na região Norte e Nordeste, tornando o país o segundo maior produtor mundial de grãos dessa leguminosa (DEAGRO, 2015). A cultura representa atualmente o principal produto agrícola de exportação nacional. De acordo com a Companhia Nacional de Abastecimento (CONAB) (2016), a área plantada com soja no país teve um acréscimo de 5,7% (1,73 milhão de hectares) nesta safra (2014/15) em relação a safra de 2013/14. A estimativa da safra nacional de soja para 2014/15 é de 95.070,2 milhões de toneladas, um recorde na produção nacional, (contra 86.120,8 milhões de toneladas em 2013/14), representando um aumento de 11,5%. No Paraná, houve o incremento de 4,1% na produção do estado na safra de 2014/15 (37.052,7 milhões de toneladas) em relação à 2013/14 (35.592,2 milhões de toneladas). Ressaltando que o estado caracteriza-se como o segundo maior produtor, perdendo apenas para o estado do Mato Grosso (CONAB, 2016).

O teor proteico elevado dos grãos de soja, cerca de 40%, faz com que a cultura possua grande importância, principalmente, nos países que apresentam carências nutricionais (ALVES et al., 2003). Esta característica proteica faz com que a demanda por nitrogênio pela planta seja alta, sendo então esse nutriente, em geral, um dos principais fatores limitantes à produtividade da cultura (HUNGRIA et al., 2006a). O nitrogênio é constituinte de aminoácidos, proteínas, bases nitrogenadas, ácidos nucleicos, hormônios, clorofila, entre outras moléculas, isto é, moléculas que são fundamentais para os processos biológicos, tornando assim, este elemento imprescindível para o desenvolvimento das plantas (KUSANO et al., 2011). Mesmo o grande reservatório de N encontrando-se na atmosfera, que apresenta em sua composição cerca de 80% deste elemento, os organismos eucariontes são incapazes de absorver o N_2 e convertê-lo a uma forma assimilável, devido à ausência do aparato enzimático necessário a clivagem da tripla ligação existente entre os átomos de N_2 ($N\equiv N$), uma das mais fortes que se tem conhecimento na natureza (BRECHENMACHER et al., 2008).

Contudo, procariotos denominados diazotróficos são capazes de assimilar o N_2 atmosférico em um processo conhecido como fixação biológica do nitrogênio (FBN). Esses micro-organismos possuem um complexo enzimático, chamado nitrogenase, que é capaz de romper a tríplice ligação do N_2 , resultando na síntese de amônia (NH_3), que será disponibilizada, inicialmente, na forma de amônio (NH_4^+). Já este é imediatamente incorporado em esqueletos de carbono a fim de serem utilizados por determinadas plantas, suprimindo o N necessário ao desenvolvimento e à produtividade (FERGUSON et al., 2010).

A nitrogenase é bastante sensível ao oxigênio, pois ele provoca desnaturação irreversível de ambos os componentes proteicos da enzima. Devido a isto, os micro-organismos fixadores de N apresentam mecanismos capazes de proteger a nitrogenase da ação do oxigênio (DROZDOWICZ, 1997). Neste contexto, no caso da soja e de outras leguminosas, a FBN ocorre dentro de nódulos (estruturas especializadas formadas nas raízes) de plantas hospedeiras, em simbiose com essas bactérias. Estas são conhecidas como fixadoras simbióticas, como exemplo, estão as pertencentes aos gêneros *Rhizobium*, *Bradyrhizobium*, *Sinorhizobium* (= *Ensifer*), entre outras (DROZDOWICZ, 1997; ARAÚJO e CARVALHO, 2006).

Os nódulos estão presentes em raízes de plantas leguminosas e são formados como resultado da infecção provocada pelas bactérias fixadoras simbióticas. A formação dos nódulos é um processo complexo, que ocorre em várias etapas e engloba mudanças morfológicas e fisiológicas, tanto na célula hospedeira quanto na bactéria. As

alterações na bactéria visam, especialmente, o recebimento de fontes de carbono da planta hospedeira, para prover o ATP e poder redutor, que são necessários para o processo de fixação, ao passo que as mudanças na planta hospedeira visam assimilar a amônia sintetizada pelas bactérias (HUNGRIA et al., 1993).

O estabelecimento da associação simbiótica entre rizóbios e a planta hospedeira inicia-se com a liberação, pelas raízes e sementes da leguminosa, de compostos fenólicos, que atuam como moléculas sinalizados nesse processo, entre essas moléculas, estão os flavonoides que atuam como substâncias quimiotáticas, estimuladoras da multiplicação dos rizóbios e indutoras de genes rizobianos responsáveis pela nodulação (genes *nod*, *nol* e *noe*). Os principais flavonoides exsudatos pelas raízes e sementes da soja que ativam a transcrição destes genes de nodulação são a genisteína e a daidzeína (HUNGRIA e STACEY, 1997; BRENCIC e WINANS, 2005).

Os genes de nodulação, como os genes *nod*, codificam a produção dos fatores de nodulação, denominados fatores Nod, que são lipo-quito-oligossacarídeos (LCOs) específicos secretados pelos rizóbios, que de forma sucinta, interagem com a planta, desencadeando como primeiro efeito a curvatura do pelo radicular; logo após, promovendo a divisão das células corticais e conduzindo à formação dos nódulos. As bactérias, então, penetram o pelo da raiz, são internalizadas em um compartimento circundado por uma membrana chamado de simbiossomo, onde passam a ser chamadas de bacteroides, ou seja, as formas capazes de fixar N (Figura 1) (HIRSCH et al., 2003; STACEY et al., 2006).

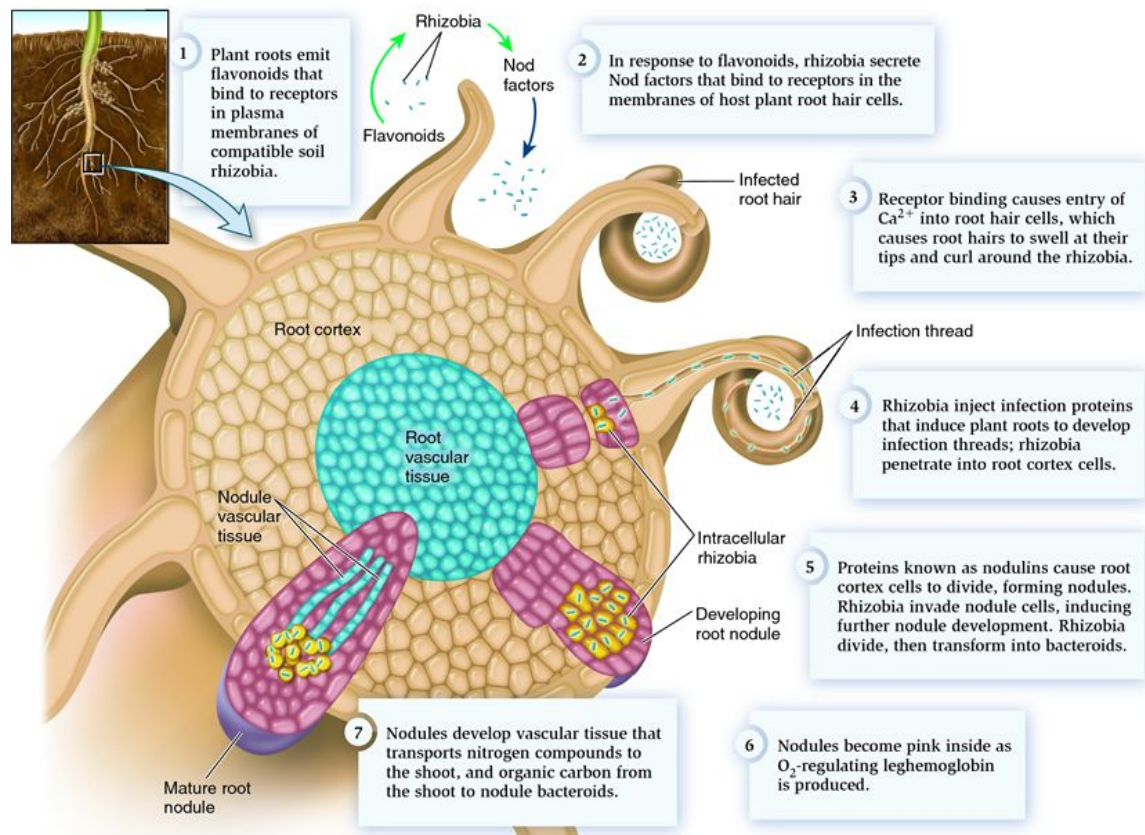


Figura 1 - Etapas do estabelecimento da simbiose entre leguminosas e rizóbios. (Fonte: biology-forums.com, 2012).

A soja pode formar nódulos radiculares fixadores de nitrogênio com algumas espécies de rizóbios, incluindo *B. japonicum* (JORDAN, 1984), *B. diazoefficiens* (DELAMUTA et al., 2013), *B. liaoningense* (XU et al., 1995), *B. elkanii* (KUYKENDALL et al., 1992), *B. yuanmingense* (YAO et al., 2002), *B. canariense* (VINUESA et al., 2005), *B. huanghuaihaiense* (ZHANG et al., 2012), *B. daqingense* (WANG et al., 2013), *Sinorhizobium fredii* (CHEN et al., 1988), *Sinorhizobium sojae* (LI et al., 2011), *Mesorhizobium tianshanense* (CHEN et al., 1995). Contudo, não são todas essas espécies que são utilizadas na formulação de inoculantes comerciais.

Atualmente no Brasil, em substituição aos fertilizantes nitrogenados, quatro estirpes de *Bradyrhizobium* são recomendadas para a utilização em inoculantes comerciais na cultura da soja. Duas são pertencentes à espécie *B. elkanii*, a SEMIA 587 (SEMIA, —Seção de Microbiologia Agrícola) e a SEMIA 5019 (=29W), recomendadas desde 1979, e duas estirpes de *Bradyrhizobium* selecionadas pela Embrapa Cerrados, a CPAC 15 (=SEMIA 5079, *B. japonicum*) e a CPAC 7 (=SEMIA 5080, *B. diazoefficiens*), que vêm sendo utilizadas intensamente em inoculantes comerciais desde 1992 (HUNGRIA et al., 2006a; SIQUEIRA et

al, 2014). A CPAC 15 é uma variante natural da SEMIA 566, obtida a partir do reisolamento de estirpes de nódulos de soja em solos do Distrito Federal, vários anos após a última inoculação, objetivando estirpes mais adaptadas as condições do bioma, e com elevada capacidade de FBN e competitividade (HUNGRIA et al., 1996).

2.1.1 A Importância Econômica e Ambiental da Fixação Biológica de Nitrogênio em Soja no Cenário Nacional

Dentre as vantagens obtidas pela FBN, destaca-se a economia no emprego de fertilizantes nitrogenados. A inoculação das sementes de soja com bactérias específicas e comprovadamente eficientes é capaz de dispensar a aplicação de fertilizantes nitrogenados, gerando uma economia estimada em quase US\$ 7 bilhões anualmente para o país (HUNGRIA et al., 2007).

Neste cenário, estimativas apontam que o custo de produção da soja nacional é cerca de 40% inferior ao dos Estados Unidos (EUA), 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 (HUNGRIA et al., 2006). Enquanto no maior produtor mundial, os EUA, 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).

O emprego de fertilizantes nitrogenados implica em custos elevados de produção, tendo em vista que a síntese química dos mesmos necessita de grande quantidade de energia para o rompimento da tripla ligação (fortemente estável) do nitrogênio gasoso para transformá-lo em uma forma livre assimilável pelas plantas. Além disso, há o baixo aproveitamento dos fertilizantes nitrogenados pelas plantas (em torno de 50%) (MERCANTE, 2005), fora o fato de que o uso de fertilizantes nitrogenados implica em poluição ambiental, pois estes acabam percolando pelo solo e poluindo os lençóis freáticos, rios, lagos e reservatórios de água, além de serem perdidos pelos processos de lixiviação e desnitrificação. (JENSEN e HAUGGAARD-NIELSEN, 2003).

2.2 AS ESTIRPES RECOMENDADAS PARA INOCULANTES COMERCIAIS EM SOLOS BRASILEIROS

No Brasil, os solos são originalmente isentos de estirpes de *Bradyrhizobium* capazes de nodular eficientemente a soja, no entanto, foram sendo selecionadas estirpes adaptadas às cultivares e condições ambientais brasileiras. Desde o período de expansão comercial da soja, essas estirpes selecionadas são recomendadas para produção de inoculantes comerciais para soja, considerando principalmente a sua capacidade competitiva, produção de nódulos e taxa de FBN (VARGAS e HUNGRIA, 1997; HUNGRIA et al., 2007).

Embora esse processo simbiótico seja conhecido desde o século XIX, foi necessário um grande esforço das instituições de pesquisa brasileiras a fim de desenvolver e selecionar bactérias adaptadas às condições do país, especialmente na região do Cerrado. A produção de inoculantes visando culturas comerciais no Brasil iniciou-se na Seção de Microbiologia Agrícola (SEMIA) do Rio Grande do Sul, com culturas líquidas e sobre ágar. Em 1956, a primeira indústria aprovada foi estabelecida no Rio Grande do Sul, com assistência do grupo FEPAGRO (Fundação Estadual de Pesquisa Agropecuária). A lista de estirpes recomendadas para uso em inoculantes comerciais foi primeiramente publicada em 1956 (HUNGRIA et al., 2005). Porém, somente a partir de 1997, o Ministério da Agricultura, Pecuária e Abastecimento (MAPA) tornou a fiscalização na produção de inoculantes mais rigorosa. (FREIRE e VERNETTI, 1999).

As estirpes microssimbiontes da soja pertencentes às espécies *B. japonicum* e *B. elkanii* foram introduzidas nos solos brasileiros, principalmente a partir das décadas de 1960 e 1970, com as primeiras inoculações na Região Sul (HUNGRIA et al., 1994). A inoculação em maior escala iniciou-se com a estirpe de *B. japonicum* SEMIA 566, que teve seu isolamento no Rio Grande do Sul em 1966 a partir de um inoculante americano. Em 1970, iniciaram-se os programas de seleção de estirpes para a inoculação de soja nos Cerrados. Neste âmbito, a SEMIA 566 foi empregada nos primeiros inoculantes utilizados nos solos dos Cerrados, conseguindo estabelecer-se no solo devido a sua alta capacidade saprofítica (VARGAS; HUNGRIA, 1997).

No Brasil, a SEMIA 566 foi a primeira estirpe “selecionada”, isolada de um nódulo de soja cultivar *Hardee*, nodulava eficientemente todos os cultivares de soja da época. Uma segunda estirpe, de *B. elkanii*, SEMIA 587, foi isolada em 1967, sendo recomendada

para inoculação em conjunto com a SEMIA 566 e SEMIA 543 de 1968 a 1975. Várias combinações de estirpes foram sugeridas, de 1976 a 1978, incluindo as SEMIA 566, SEMIA 527 e SEMIA 532. Todavia, o desempenho foi insatisfatório. Provavelmente, devido às condições do bioma, tais como, altos períodos de seca, elevada temperatura dos solos (superior a 40° C), baixa umidade do solo, pouca fertilidade, acidez do solo e toxicidade por alumínio (HUNGRIA et al., 2006b).

Com o surgimento de novos genótipos de soja de maior produtividade, e consequentemente que necessitavam de maiores demandas de N, necessitou-se de uma nova etapa de seleção de estirpes que tivessem excelente capacidade de FBN, combinada com tolerância às condições estressantes do Cerrado. Após vários ensaios, a estirpe CPAC 15 (= SEMIA 5079) foi identificada e mostrou ter capacidade de aumentar a produtividade da soja em comparação com as estirpes SEMIA 587 e 29W (PERES et al., 1993). A CPAC 15 é uma variante natural da SEMIA 566, obtida a partir do reisolamento, em solos do Distrito Federal de estirpes vários anos após a última inoculação, buscando aquelas mais adaptadas, com elevada capacidade de FBN e competitividade (HUNGRIA et al., 2006b).

Novas seleções e extensivos testes de campo levaram à identificação da estirpe CPAC 7 (= SEMIA 5080), uma variante da CB 1809, que possui uma maior capacidade competitiva. Posteriormente, em 1992, a CPAC 7 e a CPAC 15 foram incluídas na lista de estirpes autorizadas para uso em inoculantes comerciais para a soja. (VARGAS et al., 1992; PERES et al., 1993).

2.3 CARACTERÍSTICAS GERAIS DAS ESTIRPES DE *Bradyrhizobium japonicum*

Estirpes pertencentes a espécie *B. japonicum* são aeróbias, a maioria é quimio-organotróficas, devido a utilizarem pentoses como fonte de carbono. Como fonte de N, estas bactérias utilizam sais de amônia (geralmente aminoácidos e nitratos). Ainda de acordo com Santos (1998), as células da espécie *B. japonicum* são gram-negativas, não formadoras de esporo, em forma de bastonete curto, com tamanho de 0,5 a 0,9 μM por 1,2 a 3,0 μM e são, comumente, pleomórficas sob diversas condições de crescimento. De um modo geral, possuem grânulos de poli- β -hidroxibutirato que são refringentes por microscopia por contraste de fase. Possuem mobilidade por um flagelo polar único, subpolar ou dois a seis

flagelos peritríqueos (SOMASEGARAM e HOBEN, 1994). As colônias são circulares, opacas, raramente translúcidas, brancas e convexas e tendendo a serem granular em textura, crescendo em meio com carboidrato usualmente associado à secreção extracelular. Ainda de acordo com o autor, este gênero caracteriza-se também por apresentar crescimento lento, com tempo de geração de 7 a 13 horas. Muitas estirpes crescem em meios com sais minerais contendo extrato de levedura, glicose, galactose, gliconato, glicerol, frutose, arabinose ou manitol. Neste contexto, a maltose é utilizada por cerca de 10% das estirpes, já a lactose, ramnose, rafinose, trealose, sacarose, dulcitol e dextrina são raramente utilizadas. Ácidos orgânicos como fumarato, malato, succinato, citrato e piruvato são consumidos, desde que o meio possua Ca^{2+} e Mg^{2+} suficientes para sobrepujar o efeito quelante desses ácidos. Por fim, celulose e amido não são utilizados (JORDAN, 1984)

As condições ótimas de temperatura e pH para o crescimento são de 20 a 30°C e pH 6 a 7. Usualmente são tolerantes à acidez, muitas estirpes crescendo em pH 4.5, mais de 30% das estirpes crescem em pH 4.0 e poucas em pH inferior a 3.5. Não ocorre crescimento em pH superior a 9.0. Não crescem em meio contendo 2% de cloreto de sódio (NaCl) e não produzem sulfeto de hidrogênio (H_2S). A produção de penicilinase é comum (JORDAN, 1984).

Baseado nas análises de sequências de rRNA 16S, observou-se que a estirpe CPAC 15 difere em apenas um par de base da SEMIA 566 (CHUEIRE et al., 2003). Porém, ao comparar as duas estirpes, a CPAC 15 apresenta maior produção de muco (NISHI et al., 1996; BODDEY e HUNGRIA, 1997), aumento do número de pelos radiculares (fenótipo Hai) (HUNGRIA et al., 1998), resistência a vários antibióticos, alta síntese de ácido indolacético (AIA) *in vitro* e fenótipo negativo da hidrogenase (Hup^-) em relação a SEMIA 566.

Recentemente, Delamuta e colaboradores (2013) ao utilizarem diversas estirpes de *Bradyrhizobium*, dentre elas: SEMIA 6059, SEMIA 5060, CPAC 7 e USDA 110, realizaram análises do rRNA 16S, de sequencias parciais dos genes *atpD*, *glnII*, *recA*, *gyrB*, *rpoB*, e *dnaK* e da região ITS (*Intergenic Transcribed Spacer*) do rRNA 16S-23S, e com os resultados obtidos e baseados em todas as evidências genotípicas e fenotípicas, propuseram a reclassificação do “grupo Ia” de *B. japonicum* em uma nova espécie denominada *Bradyrhizobium diazoefficiens*.

Da mesma maneira que a estirpe CPAC 15, a estirpe de *B. diazoefficiens* CPAC 7 também pertence ao gênero *Bradyrhizobium*. No entanto, ao contrário da CPAC 15, a estirpe CPAC 7 apresenta fenótipo Hup^+ (NISHI et al., 1996), alta eficiência no processo de

fixação de N (DÖBEREINER *et al.*, 1970; NEVES *et al.*, 1985), competitividade (HUNGRIA *et al.*, 1998; HUNGRIA e VARGAS, 2000) e também é produtora de muco (NISHI *et al.*, 1996; BODDEY e HUNGRIA, 1997).

Ainda de acordo com NISHI e colaboradores (1996), fora as características morfológicas comuns às colônias de *Bradyrhizobium*, as estirpes CPAC 15 e CPAC 7 apresentam bordas lisas e produzem álcalis quando incubadas em meio YMA (VINCENT, 1970) a 28°C. A estirpe CPAC 7 é resistente aos antibióticos ($\mu\text{g/ml}$): tetraciclina, 100; ácido nalidíxico, 50; rifampicina, 500; streptomina, 100. Além destes antibióticos mencionados acima, a estirpe CPAC 15 também é resistente à carbenicilina, 500; e eritromicina, 250; contudo não sendo resistente a cloranfenicol, 500. Por fim, ao comparar a estirpe CPAC 7 com a CPAC 15 pode-se destacar que enquanto a CPAC 7 possui maior eficiência no processo de fixação de nitrogênio (mg de N fixado por grama de nódulo) a CPAC 15 possui maior competitividade e capacidade saprofítica (SANTOS *et al.*, 2011).

2.4 GENOMA DA ESTIRPE CPAC 15 DE *Bradyrhizobium japonicum*

A estirpe de *B. japonicum* tem um genoma de 9,582,287-bp composto de um cromossomo circular com dois operons ribossomais e um conteúdo G+C de 63.54% (SIQUEIRA *et al.*, 2014). De acordo com a anotação realizada, foram preditas 4.203 ORFs com função atribuída (48.6%) e 4.445 ORFs hipotéticas (51.4%), totalizando um total de 8.648 genes preditos. Em relação ao genoma da estirpe CPAC 7 de *B. diazoefficiens*, o genoma da estirpe CPAC 15 é 500.000 pb maior do que da CPAC 7. Já em relação a similaridade do genoma da CPAC 15 com o de organismos relacionados, observou-se que 80.9% das ORFs da CPAC 15 tem alta similaridade com as ORFs da estirpe USDA 6 de *B. japonicum*, 8.2% de similaridade com as ORFs da estirpe USDA 110 de *B. diazoefficiens* e 1.53% de similaridade com as da estirpe *Bradyrhizobium* sp. S23321.

Em resumo, considerando o genoma inteiro da CPAC 15, 12.1% deste genoma foi representado por genes hipotéticos exclusivos desta estirpe. Essas ORFs variaram de 70 a 3.270 pb, com uma média de 194 pb e estando distribuídas ao longo do genoma. A maioria das ORFs hipotéticas tem ortólogos em pelo menos um dos organismos relacionados, como na estirpe USDA 6. Os genes preditos foram encaixados em 18 categorias funcionais pelo KEGG (*Kyoto Encyclopedia of Genes and Genomes*). A principal categoria na qual as

ORFs tiveram a sua função atribuída foi a de metabolismo de aminoácidos (15.95%), seguida da degradação de xenobióticos e metabolismo (13.24%).

O início da ilha simbiótica da CPAC 15 foi demarcada em um gene de DNA-recombinase e o final da ilha simbiótica foi definida em um gene para valina tRNA. Dentro da ilha simbiótica há um gene para metionina tRNA. O tamanho da ilha simbiótica é de 700.213 pb, com um total de 569 ORFs, sendo que destas, 247 são hipotéticas (43%), além de 90 elementos móveis. A ilha simbiótica da CPAC 15 engloba todos os genes de nodulação e fixação do N₂ que são requeridos para a simbiose, uma provável operon não funcional de hidrogenase, citocromos para fornecimento de energia, transportadores do tipo ABC, operons para sistemas de secreção, ressaltando-se que todos esses genes são rodeados por muitos genes hipotéticos e elementos genéticos móveis, que podem estar relacionados a transferência gênica horizontal.

Transportadores do tipo ABC representaram 7.8% do genoma inteiro da CPAC 15, transportadores relacionados a absorção de opina, fosfonato e poliaminas. Mais genes relacionados ao transporte fósforo nesta estirpe poderiam estar relacionados com a característica que contribui para a alta capacidade saprofítica e de competitividade nos solos do Cerrado desta estirpe, assim como os transportadores de opina, pois opinas são produzidas pelas raízes das plantas e representam fontes de carbono (C) e nitrogênio (N), sugerindo-se que a utilização destas moléculas por essa estirpe poderia melhorar a sua competitividade contra outras estirpes presentes nos solos.

Além disso, foram identificadas outras 12 ilhas genômicas no genoma da CPAC 15 consideradas como sendo provenientes transferência gênica horizontal. Genes que podem conferir capacidades de adaptação como sistema de secreção do tipo I e de tolerância a estresse foram identificados nessas ilhas, ressaltando-se que 68% das ORFs dessas ilhas são hipotéticas.

Através de comparações de genomas entre *B. japonicum* e *B. diazoefficiens* foi possível verificar que os genomas foram sistêmicos, no entanto, rearranjos foram observados tanto entre e dentro de espécies, conferindo alta plasticidade genômica, que poderia conferir alta adaptabilidade ambiental e conferir diferenças na performance simbiótica. O escasso conhecimento sobre esse importante rizóbio foi enfatizado pela observação de que não foi possível atribuir função para aproximadamente 50% dos genes preditos do genoma da estirpe da CPAC 15, incluindo genes da ilha simbiótica. A grande proporção de ORFs relacionadas ao metabolismo de aminoácidos e carboidratos neste rizóbio

poderia estar relacionado a ampla adaptabilidade desta estirpe aos solos tropicais com baixo conteúdo de N e C. Destacando-se que a CPAC 15, que é excelente na capacidade saprofítica e na competitividade, tem mais genes em todas as categorias de metabólitos secundários, transportadores de nutrientes, aquisição de ferro e metabolismo de AIA (ácido indolacético) do que a estirpe CPAC 7.

2.5. GENÔMICA FUNCIONAL

Na era pós-genômica ou era pós-sequenciamento, ficou evidente que o conhecimento das sequências de todos os genes de um organismo não é suficiente para se elucidar todos os mecanismos moleculares de uma célula. Os projetos genoma envolvem o sequenciamento dos conjuntos de genes de um organismo inteiro ou de parte dele. Contudo, o conhecimento da sequência de nucleotídeos não é suficiente para entender a complexidade de um organismo, pois esses projetos não revelam dados sobre a expressão destes genes, níveis de expressão, e o funcionamento dos seus produtos. O fato de as células de um organismo possuírem o mesmo genoma, mas exibirem as mais variadas funções e morfologias, resultado dos mecanismos de regulação que levam a diferentes composições de proteínas expressas, reforça e ilustra a importância de analisar não só a sequência dos genes mas também a sua expressão, a fim de obter uma melhor compreensão das suas funções biológicas (JOAQUIM e EL-HANI, 2010).

A genômica funcional caracteriza-se como um campo da biologia molecular que descreve a função de genes e proteínas. Ao contrário da genômica estrutural, a genômica funcional se concentra nos aspectos dinâmicos, tais como a transcrição de um gene, a tradução e interações entre proteínas e metabólitos, em oposição aos aspectos estáticos da genômica estrutural, como o sequenciamento de genomas e estudos estruturais. Neste campo, a genômica funcional busca responder questões sobre a funcionalidade dos genes, transcrição de RNA e síntese de proteínas. Os estudos de genômica funcional são caracterizados por abordagens que geralmente envolvem métodos *high-throughput* (grande geração de dados), sendo as principais abordagens de avaliação nesta área, a transcriptômica, proteômica e a metabolômica (FURLAN et al., 2007).

A proteômica é definida como a caracterização em larga escala do conjunto de proteínas expressas por um organismo ou tecido em um determinado momento e sob uma condição específica (ROCHA et al., 2005). O proteoma não é somente a soma dos produtos

traduzidos a partir das sequências genômicas, mas inclui também proteínas resultantes de processos pós-transcricionais e pós-traducionais, bem como complexos formados por essas biomoléculas (AHRENS et al., 2010).

A transcriptômica é baseada no estudo da expressão dos genes. Transcriptoma é o termo dado ao conjunto de RNAs mensageiros (RNAm) que codificam as informações que a célula necessita em um determinado momento (NAKANISHI e NUREKI, 2005). Neste caso, podem-se obter informações precisas sobre a regulação da transcrição. Contudo, deve-se enfatizar que devido aos mecanismos de regulação pós-traducionais, a quantidade de proteína expressa não é necessariamente proporcional à quantidade de seu RNAm correspondente (CALSA JUNIOR et al., 2004).

O Metabolômica é a composição de todas as pequenas moléculas presentes em um organismo. É a tecnologia voltada para o fornecimento de uma visão geral compreensiva qualitativa e quantitativa dos metabólitos presentes em um organismo (HALL, 2006). Esta área surgiu devido ao fato de que alterações nos níveis de RNAm nem sempre resultavam em alterações nos níveis de proteínas (GYGI et al., 1999), e uma vez traduzida, uma proteína podia estar ou não enzimaticamente ativa (SUMNER et al., 2003). Assim, as alterações observadas no transcriptoma e no proteoma nem sempre correspondiam a alterações fenotípicas. O nível dos metabólitos representa uma informação integrativa da função celular em nível molecular, definindo assim o fenótipo de uma célula ou tecido em resposta a alterações ambientais ou genéticas. A identificação do nível dos metabólicos é, desse modo, um complemento fundamental na determinação da função gênica.

Várias metodologias vêm sendo empregadas em estudos de expressão gênica, como o SAGE (*Serial Analysis of Gene Expression*) (VELCULESCU et al., 1995), os microarranjos (ou *microarrays*) (SCHENA et al., 1995), a hibridização subtrativa supressiva (DIATCHENKO et al., 1996) e o sequenciamento de RNA (RNA Seq) (WANG et al., 2009). Além destas, as técnicas de PCR quantitativo em tempo real (RT-qPCR) (HEID et al., 1996) e *Northern blot* (ALWINE et al., 1977) possibilitam análises na expressão de genes pontuais.

Métodos para análise diferencial da expressão de genes utilizam como alvo de estudo, principalmente os transcritos (moléculas de RNAm), que são sintetizados pelo organismo durante determinado tratamento, em comparação com aqueles produzidos na ausência do mesmo (PEREIRA, 2008).

Mesmo apresentando a possibilidade de análises de expressão de centenas de genes simultaneamente, técnicas de análise em larga escala como os microarranjos de

DNA, apresentam baixa sensibilidade comparativa. Atualmente a técnica de PCR quantitativo em tempo real (RT-qPCR) apresenta-se como a metodologia mais precisa para quantificar níveis de expressão de um determinado gene, oferecendo ainda rápidos e reprodutíveis dados (GINZINGER, 2002). De modo geral, análises globais como as realizadas por microarranjos e RNAseq fornecem um panorama global da expressão gênica, sendo posteriormente os genes selecionados para estudo sendo validados por RT-qPCR. Em contraste com a PCR convencional, na qual se obtêm quantificações nos ciclos finais da reação (*end points*), a PCR em tempo real monitora a reação ciclo a ciclo, associando a amplificação do alvo em cada ciclo com a emissão de uma determinada quantidade de fluorescência. Sendo a intensidade da fluorescência emitida proporcional à quantidade de DNA alvo amplificado e aumentando numa ordem exponencial em cada ciclo de amplificação. Desta maneira, possibilitando monitorar a quantidade de produto gerada em cada ciclo (KUBISTA et al., 2006).

Na técnica de RT-qPCR, durante a fase exponencial de amplificação é possível mensurar um valor de intensidade de fluorescência, no qual todas as amostras podem ser comparadas. Este valor é denominado limiar ou *threshold*, sendo calculado em função da quantidade de fluorescência basal (*background*). Neste ponto, o sinal de fluorescência gerado pela amostra é significativamente maior que a fluorescência basal. A quantidade de ciclos de PCR que são requeridos para que cada amostra emita fluorescência suficiente para alcançar este ponto é definido *cycle threshold* ou Ct. O Ct é específico para cada amostra e é inversamente proporcional à quantidade inicial do alvo presente na reação. Este valor é a base para a quantificação baseada em PCR quantitativo (KUBISTA et al., 2006; MARCELINO, 2006).

2.6 ANÁLISES TRANSCRICIONAIS EM *Bradyrhizobium japonicum* E OUTRAS ESTIRPES MICROSSIMBIONTES DA SOJA

Diversas análises transcricionais e proteômicas vêm sendo empregadas em rizóbios, tanto a níveis globais quanto a níveis pontuais. Estes estudos fornecem informações valiosas sobre os mecanismos biológicos envolvidos tanto na simbiose quanto na sobrevivência destes rizóbios sob diferentes condições. Ao longo desse capítulo demonstra-se como o emprego destas metodologias vem sendo utilizadas para elucidar funções de genes e proteínas durante a simbiose e outros processos biológicos importantes.

Lang e colaboradores (2008) através da técnica de microarranjo analisaram o perfil transcricional de *B. japonicum* em resposta a indução com genisteína, uma isoflavona secretada pelas raízes de soja. Neste trabalho, aproximadamente 100 genes foram induzidos, incluindo, todos os genes de nodulação, um *cluster* de genes para a síntese de flagelo e genes que provavelmente estão envolvidos em processos de transporte. Vale à pena mencionar que dos 100 genes induzidos por genisteína, 34 genes estavam dentro da ilha simbiótica.

Delmotte e colaboradores (2010) realizam um estudo proteômico das proteínas que estavam sendo expressas nos bacteroides de *B. japonicum* e com dados obtidos anteriormente do perfil transcricional dos genes expressos nos bacteroides sob as mesmas condições, conseguiram compilar um perfil global de expressão de *B. japonicum* durante a simbiose com a soja, que compreende 3587 genes/proteínas. Através de classificações funcionais e análises de vias bioquímicas, foi possível observar que a maioria das proteínas das vias metabólicas do nitrogênio e do carbono foram expressas, incluindo o conjunto completo de enzimas do ciclo do ácido tricarboxílico, enzimas da via pentose-fosfato e da via gliconeogênese, além de diversas proteínas que anteriormente não foram consideradas em estarem presentes durante a simbiose.

Em outro estudo, Gao et al., (2015) analisaram a expressão dos genes *nodA*, *nodD1* e *nodD2* em estirpe de *B. liaoningense* estirpe CCBAU05525 induzida por materiais húmicos hidrossolúveis (WSHM) e genisteína e observaram que a expressão relativa destes genes alcançou o mais alto nível após 3 horas da adição de WSHM aos meios de cultura. O mesmo padrão foi observado no tratamento com genisteína, no entanto, a indução por este flavonoide foi significativamente menor do que a obtida por WSHM

Pessi e colaboradores (2007) avaliaram o perfil global transcricional de *B. japonicum* sobre três diferentes condições fisiológicas (aeróbica, micro aeróbica e bacteroide) e quatro diferentes tempos de desenvolvimento do nódulo e bacteroide (10, 13, 21 e 31 dias pós-inoculação [dpi]). Neste trabalho, ao comparar o padrão de transcritos em bacteróides de soja (2780 genes) em relação às células crescidas em meio de cultura sob condição aeróbica (5439 genes), um menor número de genes foi expresso apenas em bacteroides (411 genes). Dos 2780 genes expressos em bacteroide, um notável número de genes (2369) são expressos tanto em bacteroide quanto em condição aeróbica.

Bortolan e colaboradores (2009) avaliaram a expressão de genes de nodulação *nodC* e *nodW* e do gene *nodP* em *B. japonicum* CPAC 15 em dois experimentos distintos utilizando a técnica de RT-qPCR. No primeiro experimento observaram-se maiores

níveis de expressão logo após o contato com a genisteína (15 minutos) do que em tempos maiores de incubação (1,4 e 8 horas). Em um segundo experimento, avaliou-se a expressão destes genes nas bactérias crescidas por 48 horas em presença de genisteína e exsudatos de sementes de soja, sendo observado que a expressão dos três genes foi maior na presença de genisteína do que exsudatos de sementes. Em comparação aos dois experimentos, os maiores níveis de indução sendo alcançados no experimento por 48 horas. Os resultados obtidos confirmaram a funcionalidade dos três genes na estirpe CPAC 15, com ênfase para o *nopP*, cuja funcionalidade em *B. japonicum* foi descrita pela primeira vez. Já em estudo desenvolvido por Gomes e colaboradores (2014), houve a seleção de 10 proteínas hipotéticas obtidas através de mapa de referência 2-D de *B. diazoefficiens* estirpe CPAC 7 e destas, 6 tiveram a sua expressão gênica *up*-regulada (ou seja, níveis de expressão diferencialmente aumentados) em resposta a genisteína por RT-qPCR.

Wei e colaboradores (2008) compararam o perfil global de expressão gênica de *Bradyrhizobium japonicum* USDA 110 em resposta a indução com extratos de sementes de soja (SSE) e genisteína (5 μ M). Neste estudo, SSE induziu mais *loci* gênicos do que a genisteína, acentuadamente induzindo regiões genômicas dentro da ilha simbiótica, que incluem genes *tts* (sistema de secreção do tipo III), vários genes *nod*, genes para poligalacturonase (degradação da parede celular), assim como genes que estão fora da ilha simbiótica, sendo estes os de síntese de exopolissacarídeos (EPS), ACC (1-aminociclopropano-1-carboxilato) deaminase, família de proteínas ribossomais e do metabolismo de energia, além de muitos genes codificando para proteínas hipotéticas. Neste mesmo trabalho, os autores validaram a expressão de 12 genes expressos pelo microarranjo por RT-qPCR. Dos genes validados, nove estavam localizados dentro da ilha simbiótica e 3 fora da ilha.

Já Mesa e colaboradores (2009) a fim de elucidar a sensibilidade ao estresse oxidativo de FixK₂ (este gene é ativado pelo sistema de dois componentes FixLJ em resposta a um moderado decréscimo da tensão de oxigênio) realizaram uma análise global por microarranjo em *B. japonicum* em resposta a H₂O₂ (peróxido de hidrogênio). Observando assim, que a maioria dos genes do regulon FixK₂ diminuíram a sua expressão após tratamento com H₂O₂. Gourion e colaboradores (2009) também realizaram análises de microarranjo e observaram que os dois reguladores PhyR e ECF σ factor σ^{EcfG} em *B. japonicum* fazem parte da mesma cascata de sinalização, sendo ainda identificados no estudo um grande número de genes com funções desconhecidas e que estão presentes no regulon PhyR/ σ^{EcfG} .

Takeshima e colaboradores (2013) encontraram uma região amplamente separada da ilha simbiótica em *B. japonicum* e que foi perceptivamente induzida dentro de minutos após a adição de genisteína. O *locus* em questão, nomeado de BjG30, contém genes para uma bomba de efluxo multidroga e reguladores transcricionais da família TetR.

Diversas análises vêm sendo empregadas para a obtenção do perfil global de expressão gênica em *B. japonicum* sob diferentes condições. A maioria das análises globais envolve técnicas como microarranjo, biblioteca subtrativa suprimida (SSH) e RNA-seq, estas análises fornecem um grande número de genes que são expressos diferencialmente, principalmente, ao comparar duas situações (controle e tratamento), e fornecem aos pesquisadores ótimas fontes para a bioprospecção de genes que sejam úteis para elucidar e compreender os mecanismos genéticos e as vias interligadas no genoma da bactéria frente a diversas condições. Como nos exemplos citados anteriormente ao longo do texto, análises do transcriptoma em determinadas situações podem ser úteis desde para obter genes que são dependentes de algum regulador transcricional ou de outros genes (através do perfil transcricional de mutantes e estirpes selvagens) quanto para descobrir novos genes e vias que sejam super expressas em presença de exsudatos de sementes, raízes e flavonoides isolados sob diferentes tempos de indução, assim como em diferentes condições de cultivo, envolvendo diferentes compostos nutricionais e temperaturas. Cada vez mais análises transcricionais vêm sendo empregadas no intuito de elucidar os genes responsivos a diferentes estresses, como o oxidativo e térmico. De modo geral, as análises do perfil global são muito úteis e vantajosas, sendo que o refinamento e validação da expressão gênica são realizados por RT-qPCR. Vale à pena mencionar que análises transcriptômicas e proteômicas vêm sendo empregadas concomitantemente ou complementando uma a outra e vêm fornecendo um panorama mais real da funcionalidade dos genes/proteínas que estejam envolvidas na FBN.

2.7. ANOTAÇÃO FUNCIONAL DE PROTEÍNAS HIPOTÉTICAS

Atualmente, recentes técnicas de sequenciamento em larga escala, *next-generation sequencing* (NGS) (METZKER, 2010), empregadas em projetos genoma e transcriptoma, vêm produzindo um enorme volume de sequências de genes e proteínas com funções ainda desconhecidas, e por consequente, criando a necessidade de desenvolver metodologias e ferramentas eficientes com a finalidade de dar suporte computacional a esses

projetos de sequenciamento. Nestes projetos, pesquisadores visam identificar sequências de DNA e/ou RNA. As funções, juntamente com outras características biológicas, são determinadas na fase de anotação (SOUZA, 2014).

O processo de anotação caracteriza-se como a atribuição de funções as sequências identificadas no organismo em estudo. Devido à geração de um grande volume de dados não caracterizados, a anotação requer automação, uma tarefa complexa que inclui diferentes metodologias e estratégias de anotação para a predição de funções dos genes e das proteínas codificadas por eles (PANDEY et al., 2006). A anotação e caracterização das sequências de proteínas não é algo trivial, e requer a integração de uma variedade de métodos de predição de funções, além do conhecimento biológico de proteínas, para que a anotação seja confiável. Certos métodos de anotação extensivamente utilizados são baseados em: (i) homologia, investigando similaridade de sequências de nucleotídeos ou aminoácidos como motivos e domínios conservados; (ii) informações recuperadas de características estruturais das proteínas, como padrões espaciais de dobramento e sítios ativos de enzimas; e (iii) predição de função, investigando conservação em organismos relacionados, sendo grande parte destas informações obtidas de bancos de dados continuamente atualizados (SOUZA, 2014).

As consultas por similaridades de sequências permitem inferir proteínas ou genes “homólogos”. Indicativos de homologia, por exemplo, são sequências similares que apresentam aproximadamente a mesma sequência de aminoácidos e que possivelmente foram herdadas de um ancestral comum. Este processo tem como hipótese o fato de que os organismos herdam características de seus ancestrais no processo evolutivo, preservando informações genéticas de sua linhagem, em particular funções biológicas importantes. A busca por similaridades é um método extensivamente utilizado para predição de função de proteínas não caracterizadas, sendo conhecido como transferência de anotação baseada em homologia (PEARSON, 2013).

A metodologia de busca por similaridades apresenta algumas limitações que devem ser consideradas, sendo a mais significativa decorrendo da evolução divergente em resposta à pressão seletiva (GERLT e BABBITT, 2000) no qual uma duplicação de um determinado gene pode desencadear uma nova função, desta forma, não garantindo que a predição de função por homologia esteja correta. Por esta razão, estratégias buscam combinar resultados de diversas fontes para concretizar uma predição por função mais acurada, tais como a investigação de conjuntos (*clusters*) de sequências com funções similares e pesquisar

categorias funcionais do *Gene Ontology* (GO), dentre outras fontes. Abordagens baseadas em vizinhança gênica (*clusters*) são baseadas na hipótese de que as proteínas, cujos genes correspondentes são localizados “próximos” uns dos outros em um genoma, podem estar relacionadas funcionalmente, por exemplo, genes codificando para proteínas que poderiam participar da mesma via metabólica. Portanto, torna-se uma estratégia viável para inferir associações funcionais entre genes e suas proteínas correspondentes (SOUZA, 2014).

Algumas ferramentas e bancos de dados utilizados para buscas por similaridade empregadas são:

- I. *Basic Local Alignment Search Tool* (BLAST): esta ferramenta busca por alinhamentos locais entre sequências de nucleotídeos e proteínas, fazendo a comparação e calculando a significância estatística de cada sequência em um banco de dados (*e-value*). O cálculo de similaridade é auxiliado por uma matriz de substituição, que modela probabilidades de ocorrência de mutações e de substituições entre aminoácidos similares. Por exemplo, uma matriz padrão bastante utilizada é a BLOSUM62 (MADDEN, 2013).
- II. *Hmmer*: emprega perfis de modelos ocultos de Markov (*profile Hidden Markov Models* – *profile-HMM*) para executar a comparação de sequências, sendo utilizado para consulta por homólogos em bancos de dados de sequências de proteínas ou DNA. *Hmmer* visa ser significativamente preciso, além de ser capaz de detectar homólogos remotos, devido aos seus modelos (EDDY, 1998, 2010).
- III. *UniProtKB*: é um consórcio que visa fornecer anotação relevante e curada de proteínas. Baseia-se em dados de proteômica (principalmente), genômica e transcriptômica. É um banco de dados dividido em sub-bancos, sendo estes: *Swiss-Prot*, que é um banco de dados de sequências de proteínas que apresentam anotações curadas, isto é, revistas manualmente, e de alta qualidade, contendo nomenclaturas padronizadas, *links* integrados para outros bancos de dados especializados, fornecendo informações sobre o *gene ontology* (GO) e mínima redundância. Já o *TrEMBL* é um banco suplementar ao *Swiss-Prot*, sendo formado por sequências de proteínas que possuem anotações automáticas (BOECKMANN, 2003).
- IV. *RefSeq*: é um banco não redundante de sequências de genes, transcritos e proteínas, provido e curado pelo *National Center for Biotechnology Information* (NCBI). As sequências anotadas contêm regiões de codificação, não são redundantes, domínios

conservados, referências cruzadas entre bancos, entre outras características (PRUITT et al., 2005).

- V. *Pfam*: é um banco de famílias de proteínas construídas com HMM, onde as famílias são estabelecidas por um conjunto de proteínas que compartilham um nível significativo de similaridade, sugerindo assim homologia. As famílias Pfam devem representar unidades funcionais (domínios) que, quando combinados de diferentes maneiras, podem gerar proteínas com funções únicas. As famílias podem conter domínios não caracterizados, identificados no banco como “*Domain of Unknown Function*” (DUF). Assim como as famílias não caracterizadas são identificadas no *Pfam* como “*Uncharacterized Protein Family*” (UPF) (PUNTA, et al., 2012).
- VI. *KEGG*: é uma coleção de bancos de dados online que ligam genomas com vias metabólicas. O *KEGG* integra muitos bancos, tais como o *KEGG Pathway* que é uma coleção de mapas desenhados manualmente representando interações moleculares e vias de reações para metabolismo, informação genética, processos celulares, sistemas organizacionais e doenças humanas. *KEGG Orthology (KO)* é uma coleção de grupos ortólogos definidos manualmente. Já o *KEGG Enzyme* é uma implementação da Nomenclatura de Enzimas (número EC) produzido pela IUBMB/IUPAC *Biochemical Nomenclature Committee*. O *KEGG Reaction* por sua vez, contém todas as reações adotadas a partir do *KEGG enzyme* e reações adicionais obtidas a partir do *KEGG Pathway*. (KANEHISA et al., 2012).
- VII. *Rhizobase*: é um banco de dados curados de genomas de rizóbios, no qual curadores tem extraído nome, símbolo, produtos e funções de cada gene reportado na literatura. Este banco fornece uma maneira simples de acessar todos os dados de sequência e de anotações através de mapas de imagem, buscas por palavras chaves e por listas de gene (FUJISAWA et al., 2014).

Proteínas hipotéticas (HPs) são definidas como proteínas identificadas por análises computacionais a partir de sequências nucleotídicas (*ORF - Open Reading Frame*) obtidas principalmente, em sequenciamentos de genomas. Portanto, tem sua existência predita, mas não existe evidência experimental da sua expressão, estas proteínas são geralmente caracterizadas pela baixa identidade com as proteínas conhecidas anotadas em bancos de dados. Uma "proteína hipotética conservada" descreve uma parcela de genes nos genomas sequenciados que são homólogos a genes encontrados em outros organismos, mas que também

não foram funcionalmente caracterizados e descritos no nível químico da proteína (DESLER et al, 2009).

Alguns estudos vêm sugerindo que a maioria destas proteínas ditas hipotéticas não existe, isto é, podem ser resultados da anotação de pseudogenes (genes que não são mais funcionais) ou ainda sendo consequências de erros que possam ter surgido durante o sequenciamento ou no processo de anotação funcional (DESLER et al, 2009, POPTSOVA e GOGARTEN, 2010). Em uma análise realizada no genoma de vários organismos foi verificado que muitas proteínas hipotéticas podem ser pseudogenes. Geralmente, pseudogenes consistem de sequências curtas, em contraste com as proteínas conservadas que em geral são compostas por um maior número de aminoácidos (LIPMAN et al, 2002). Isto vem sendo evidenciado em estudos, como o desenvolvido por Qiao e colaboradores (2013) que analisaram todos os genes do genoma de *Synechocystis* sp. PCC 6803 e evidenciaram que as ORFs com número entre 150 e 200 aminoácidos estavam em maiores índices como proteínas hipotéticas, sendo que entre estas, as com aproximadamente um tamanho de 150 aminoácidos quase não foram expressas em análises proteômicas sob diferentes condições, sugerindo assim, que estas sejam oriundas de pseudogenes. Outro estudo ao utilizar análise da expressão gênica em combinação com análise proteômica, também resultou na eliminação de 255 (7,5%) proteínas e adição de 127 (3,7%) novas proteínas na anotação do genoma do *Desulfovibrio vulgaris*, sugerindo que grande parte das proteínas anotadas como hipotéticas são pseudogenes, tendo em vista a ausência de expressão destes genes em nível transcricional ou traducional (PRICE et al., 2011). Vale ressaltar também, que a re-anotação de genomas também levam à conclusão de que uma parcela de proteínas hipotéticas são erros de anotação (VONGSANGNAK et al., 2008; DESLER et al, 2009; MARCELLIN et al, 2013).

Por outro lado, existem vários estudos que contribuem para a confirmação da autenticidade e caracterização de proteínas hipotéticas. Em um estudo foram selecionados 10 genes anotados como codificados de proteínas hipotéticas em *Trypanosoma cruzi*, cujas expressões foram obtidas por RNA-seq. Essas proteínas foram caracterizadas molecularmente através de vetores de expressão e análises de bioinformática (RAMOS, 2014). Outro estudo observou que cerca de 60% dos genes do genoma de *Paracoccidioides* spp. codificavam para proteínas anotadas como hipotéticas; estudos transcricionais revelaram que muitas proteínas hipotéticas foram expressas nas fases de micélio e levedura, a caracterização funcional destas proteínas hipotéticas foram realizadas utilizando métodos de comparações de similaridade de sequência, presença de peptídeos sinais, presença de domínios de proteínas conhecidas e

também uma classificação funcional baseada no *Gene Ontology*. Além da análise de seis proteínas hipotéticas a nível transcricional por RT-qPCR que demonstrou níveis de expressão para os genes codificando para essas proteínas. (SILVA, 2014).

Genes codificando para proteínas hipotéticas vêm sendo encontradas na grande maioria dos estudos de perfil global a nível transcricional sob uma diversidade de condições em rizóbios. Lang et al. (2008) analisaram o perfil global transcricional de *B. japonicum* em resposta a genisteína e observaram a expressão de vários genes codificando para proteínas hipotéticas. Outro perfil global transcricional de *B. japonicum* sob estresse oxidativo mediado pelo herbicida *paraquat* evidenciou que uma parcela significativa dos genes diferencialmente expressos durante o estresse são anotados como hipotéticos (DONATI et al., 2011). Várias análises transcricionais sob diferentes condições foram realizadas e compiladas para obter um *pool* de transcritos reiterados que foram expressos durante a FBN em *B. japonicum*, deste *pool*, a maioria dos transcritos reiterados com expressão aumentada nos bacteróides foram genes hipotéticos, sugerindo assim, muitos processos envolvidos na simbiose que ainda não foram identificados (FRANCK et al., 2014). Recentemente, em um estudo realizado por López-Fuentes et al. (2015) demonstrou-se que mutantes em genes codificando para proteínas hipotéticas localizados em uma região de genes de transferência conjugativa em *Rhizobium etli* CFN42 afetaram a conjugação de diferente formas dependendo da estirpe e do gene hipotético mutado.

A nível proteômico, várias análises também demonstram a presença de proteínas hipotéticas em rizóbios sob diferentes condições. Silva Batista et al. (2010) atribuíram funções para 28 proteínas hipotéticas diferencialmente expressas em *B. japonicum* CPAC 15 identificadas por análises proteômicas em relação a outras duas estirpes relacionadas. Outro estudo proteômico em *B. japonicum* CPAC 15 identificou proteínas que foram diferencialmente expressas em presença de genisteína, entre as 45 proteínas que foram fortemente induzidas pelo flavonóide, encontravam-se várias hipotéticas (SILVA BATISTA e HUNGRIA, 2012). Ao comparar os perfis proteômicos de duas estirpes de *B. diazoefficiens* que são contrastantes na competitividade para nodulação em resposta aos exsudatos de raízes de soja, Liu et al., (2015) identificaram várias proteínas hipotéticas presentes exclusivamente em cada uma das estirpes. Gomes et al. (2014) ao analisarem o perfil proteômico de *B. diazoefficiens* CPAC 7 sob condições de vida livre também observaram a presença de proteínas hipotéticas, destas, selecionou-se 10 para a verificação da expressão gênica via RT-qPCR em resposta a genisteína, sendo seis *up*-reguladas por este flavonóide.

A presença de genes codificando para proteínas hipotéticas é onipresente em todos os genomas, apesar de uma parcela destes genes serem oriundos tanto de erros no processo de sequenciamento e montagem de *contigs* quanto de anotações, a outra parcela corresponde a genes que ainda não possuem validação experimental, tanto a nível transcricional como a nível proteômico, mas que provavelmente são funcionais. Diante deste panorama, torna-se evidente que o conhecimento biológico atual destes genes e subseqüentes proteínas é ainda escasso e requer estudos mais aprofundados para elucidação das funções e processos biológicos que estas proteínas participam. No que se refere à FBN, em especial, a simbiose entre *B. japonicum* e a soja, a literatura vem mostrando que proteínas hipotéticas sem funções conhecidas ou com funções previstas por análises de similaridade e sem validação experimental representam uma parcela significativa nos diferentes estudos envolvendo a FBN. Siqueira et al. (2014) ao realizarem o sequenciamento do genoma de duas estirpes elites empregadas nos inoculantes para a soja (*B. japonicum* CPAC 15 e *B. diazoefficiens* CPAC 7) observaram que aproximadamente 50% dos genes presentes nestes genomas codificaram para proteínas hipotéticas. Sendo então necessários intensos estudos tanto a nível transcricional quanto a nível proteômico com o intuito de elucidar as funções que estes genes desempenham nessas estirpes, e desta forma, lançando perspectivas para um aumento da compreensão das características superiores que estes rizóbios apresentam.

Portanto, este estudo é de grande relevância, ao proporcionar avanços no conhecimento genético desta estirpe, lançando perspectivas de identificação de novos genes e de suas funções durante a FBN, e por conseqüente, permitindo uma maximização deste processo na soja e facilitando a seleção de estirpes mais eficientes e competitivas.

3 OBJETIVOS

3.1 OBJETIVO GERAL

O presente estudo teve como objetivo a inferência funcional e análise da expressão de genes hipotéticos localizados na ilha simbiótica da estirpe CPAC 15 de *B. japonicum* em resposta a genisteína.

3.2 OBJETIVOS ESPECÍFICOS

- 1) Selecionar genes (“ORFs – *Open Reading Frames*”) que codificam proteínas hipotéticas e que estejam localizados na “Ilha Simbiótica” do genoma da estirpe CPAC 15 de *Bradyrhizobium japonicum*.
- 2) Analisar a similaridade das proteínas hipotéticas codificadas pelo genes em outros genomas através da busca de similaridade com o uso do programa BLASTp (NCBI) e com sequências depositadas no banco de dados RefSeq.
- 3) Realizar a anotação funcional, ou seja, a inferência das funções das proteínas preditas destes genes, através de análises de similaridade em bancos de dados especializados e curados, como o *UniProtKB*, *KEGG*, *Pfam*, *Gene Ontology*, *Rhizobase*, entre outros. Com isso, obter informações de possíveis domínios de famílias de proteínas e apontamentos das funções. Também prever a localização celular através de programas como o PSORT-B, PSL-pred, SignalP 4.1 e SecretomeP 2.0.
- 4) Analisar a expressão dos genes codificando para as proteínas hipotéticas através da técnica de RT-qPCR utilizando sistema de indução com genisteína em presença da bactéria.

4 REFERÊNCIAS BIBLIOGRÁFICAS

AHRENS C. H.; BRUNNER, E.; QELI, E.; BASLER, K.; AEBERSOLD, R.; Generating and navigating proteome maps using mass spectrometry. **Nature Reviews Molecular Cell Biology**, v. 11, n. 11, p. 789-801, 2010.

ALI, N. Soybean processing technology and promotional efforts for its food uses. In: Proceedings of the international soybean processing and utilization conference (ISPUC-III), 2000, Tokyo. **Annals of Proceedings of the Internacional Soybean Processing and Utilization Conference (ISPUC-III)**. Tokyo: Korin Publishing, p. 327–328, 2000.

ALVES, B. J. R.; BODDEY, R. M.; URQUIAGA, S. The success of BNF in soybean in Brazil. **Plant and Soil**, v.252, p.1–9, 2003.

ALWINE, J. C.; KEMP, D. J.; STARK, G. R. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. **Proceedings of the National Academy of Sciences of the United States of America**, v. 74, p. 5350-5354, 1977.

ARAÚJO, A. S. F.; CARVALHO, E. M. S. Fixação biológica do nitrogênio em leguminosas. **Comunicado Técnico**. Universidade Federal do Piauí, n.11, p.1-4, 2006.

ARAÚJO, F. F.; HUNGRIA, M. Nodulação e rendimento de soja co-infectada com *Bacillus subtilis* e *Bradyrhizobium japonicum/Bradyrhizobium elkanii*. **Pesquisa Agropecuária Brasileira**, Brasília, v. 34, p. 1633-1643, 1999.

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.

BOECKMANN, B. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. **Nucleic Acids Research**, v. 31, n. 1, p. 365–370, 2003.

BORTOLAN, S.; BARCELLOS, F. G.; MARCELINO, F. C.; HUNGRIA, M. Expressão dos genes *nodC*, *nodW* e *nopP* em *Bradyrhizobium japonicum* estirpe CPAC 15 avaliada por RT-qPCR. **Pesquisa Agropecuaria Brasileira**, v. 44, n. 11, p. 1491–1498, 2009.

BRECHENMACHER, L.; KIM, M. Y.; BENITEZ, M.; LI, M.; JOSHI, T.; CALLA, B.; LEE, M. P.; LIBAULT, M.; VODKIN, L. O.; XU, D.; LEE, S. H.; CLOUGH, S. J.; STACEY, G. Transcription profiling of soybean nodulation by *Bradyrhizobium japonicum*. **Molecular Plant-Microbe Interactions**, v.21, p. 631-645, 2008.

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: MMBR**, v. 69, n. 1, p. 155–194, 2005.

BUCKER MORAES, W.; MARTINS FILHO, S.; GARCIA, G. O., CAETANO, S. P.; BUCKER MORAES, W.; CARRARA COSMI, F. Avaliação da fixação biológica do

nitrogênio em genótipos de feijoeiros tolerantes a seca. **Idesia (Arica)**, v. 28, n. 1, p. 61-68, 2010.

CALSA JUNIOR, T.; BENEDITO, V. A.; FIGUEIRA, A. V. Análise serial da expressão gênica. **Biotecnologia Ciência e Desenvolvimento**, v. 33, p. 88-100, 2004.

CHEN, W. X.; WANG, E. T.; WANG, S. Y.; LI, Y. B.; CHEN, X. D.; LI, Y. Characteristics of *Rhizobium tianshanense* sp. nov., a moderately and slowly growing root nodule bacterium isolated from an arid saline environment in xinjiang, people's republic of China. **International Journal of Systematic Bacteriology**, n. 45, p. 153-159, 1995.

CHEN, W. X.; YAN, G. H.; LI, J. L. Numerical taxonomic study of fast-growing soybean rhizobia and a proposal that *Rhizobium fredii* be assigned to *Sinorhizobium* gen. nov. **International Journal of Systematic Bacteriology**, n. 38, p. 392-397, 1988.

CHEN, Y. X., ZHOU, T., PENTTINEN, P., ZOU, L., WANG, K., CUI, Y. Q., et al. (2015). Symbiotic matching, taxonomic position, and field assessment of symbiotically efficient rhizobia isolated from soybean root nodules in Sichuan, China. **Biology and Fertility of Soils**, v. 51, n. 6, p. 707-718, 2015.

CHUEIRE, L.M.O.; BANGEL, E. V.; MOSTASSO, F. L.; CAMPO, R. J.; PEDROSA, F. O.; HUNGRIA, M. Classificação taxonômica das estirpes de rizóbio recomendadas para as culturas da soja e do feijoeiro baseada no sequenciamento do gene 16S rRNA. **Revista Brasileira de Ciência do Solo**, v.27, p.833-840, 2003.

CONAB - Acompanhamento da Safra Brasileira – Grãos – Safra 2015/2016 – Sétimo Levantamento, v. 7, p. 1-158, 2016.

DELAMUTA, J. R. M.; RIBEIRO, R. A.; ORMEÑO-ORRILLO, E.; MELO, I. S.; MARTÍNEZ-ROMERO, E.; HUNGRIA, M.; Polyphasic evidence supporting the reclassification of *Bradyrhizobium japonicum* Group Ia strain as *Bradyrhizobium diazoefficiens* sp. nov. **International Journal of Systematic and Evolutionary Microbiology**, v. 63, p. 3342-3351, 2013.

DELMOTTE, N.; AHRENS, C. H.; KNIEF, C.; et al. An integrated proteomics and transcriptomics reference data set provides new insights into the *Bradyrhizobium japonicum* bacteroid metabolism in soybean root nodules. **Proteomics**, v. 10, n. 7, p. 1391–1400, 2010.

DESLER, C.; SURAVAJHALA, P.; SANDERHOFF, M.; RASMUSSEN, M.; RASMUSSEN, L. In Silico screening for functional candidates amongst hypothetical proteins. **BMC Bioinformatics**, v. 10, n. 1, p. 289, 2009

DIATCHENKO, L.; LAU, Y. F.; CAMPBELL, A. P.; CHENCHIK, A.; MOQADAM, F.; HUANG, B.; LUKYANOV, S.; LUKYANOV, K.; GURSKAYA, N.; SVERDLOV, E. D.; SIEBERT, P. D. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. **Proceedings of the National Academy of Sciences of the United States of America**, v. 93, p. 6025-6030, 1996.

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

EDDY, S. Profile hidden Markov models. **Bioinformatics**, v. 14, n. 9, p. 755–763, 1998.

EDDY, S. R. HMMER User's Guide. **Hmmer Manual**, n. March, p. 0–93, 2010.

FERGUSON, B. J.; INDRASUMUNAR, A.; HAYASHI, S.; LIN, M. H.; LIN, Y. H.; REID, D. E.; GRESSHOFF, P. M. Molecular analysis of legume nodule development and autoregulation. **Journal of Integrative Plant Biology**. v. 52, p. 61–76, 2010.

FRANCK, S.; FRANCK, W. L.; BIRKE, S. R.; et al. Comparative transcriptomic analysis of symbiotic *Bradyrhizobium japonicum*. **Symbiosis**, v. 63, n. 3, p. 123–135, 2014.

FREIRE, J. R. J.; VERNETTI, F. J. A pesquisa com soja, a seleção de rizóbio e a produção de inoculantes no Brasil. **Pesquisa Agropecuária Gaúcha**, n. 5, p. 117-126, 1999.

FUJISAWA, T.; OKAMOTO, S.; KATAYAMA, T.; et al. CyanoBase and RhizoBase: databases of manually curated annotations for cyanobacterial and rhizobial genomes. **Nucleic Acids Research**, v. 42, n. D1, p. D666–D670, 2014.

FURLAN, L. R.; FERRAZ, A. L. J.; BORTOLOSSI, J. C. A genômica funcional no âmbito da produção animal: estado da arte e perspectivas. **Revista Brasileira de Zootecnia**, v. 36, p. 331–341, 2007.

GAO, T. G.; XU, Y. Y.; JIANG, F.; LI, B. Z.; YANG, J. S.; WANG, E. T.; YUAN, H. L. GERLT, J. A; BABBITT, P. C. Can sequence determine function? **Genome biology**, v. 1, n. 5, p. 1-10, 2000

GINZINGER, D. G. Gene quantification using real-time quantitative PCR: an emerging technology hits the mainstream. **Experimental Hematology**, v. 30, p. 503-512, 2002.

GOMES, D. F.; BATISTA, J. S.; ROLLA, A. A.; et al. Proteomic analysis of free-living *Bradyrhizobium diazoefficiens*: highlighting potential determinants of a successful symbiosis. **BMC genomics**, v. 15, n. 1, p. 643, 2014.

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, p. 1405–1412, 2001.

HEID, C. A.; STEVENS, J.; LIVAK, K. J.; WILLIAMS, P. M. Real time quantitative PCR. **Genome Research**, v. 6, p. 986-994, 1996.

HIRSCH, A. N.; BAUER, W. D.; BIRD, D. M.; CULLIMORE, J.; TYLER, B.; YODER, J. I. Molecular signals and receptors: controlling rhizosphere interactions between plants and other organisms. **Ecology**, v.8 4, p. 858-868, 2003.

HUNGRIA, M.; CAMPO, J. R.; MENDES, I. C. **Fixação biológica do nitrogênio na cultura da soja**. Londrina: EMBRAPA-CNPSO, 2001. 48p.

HUNGRIA, M.; CAMPO, R. J.; MENDES, I. C. A importância do processo de Fixação biológica do nitrogênio para a cultura da soja: componente essencial para a competitividade do produto brasileiro. Londrina: Embrapa Soja, 2007. 48p. (**Embrapa Soja. Documentos, 283**).

HUNGRIA, M.; CAMPO, R.J.; MENDES, I. C.; GRAHAM, P. H. Contribution of biological nitrogen fixation to the N nutrition of grain crops in the tropics: the success of soybean (*Glycine max* L. Merr.) in South America. In: SINGH, R. P.; SHANKAR, N.; JAIWAL, P. K., eds. **Nitrogen nutrition and sustainable plant productivity**. Houston, Texas: Studium Press, LLC, p.43-93, 2006a.

HUNGRIA, M.; CHUEIRE, L.; MEGIAS, M.; LAMRABET, Y.; PROBANZA, A.; GUTTIERREZ-MAÑERO, F. J.; CAMPO, R. J. Genetic diversity of indigenous tropical fast-growing rhizobia isolated from soybean nodules. **Plant Soil**, n. 288, p. 343-356, 2006b.

HUNGRIA, M.; PHILLIPS, D. A. Effects of a seed color mutation on rhizobial nod-gene-inducing flavonoids and nodulation in common bean. **Molecular Plant-Microbe Interactions**, v. 6, p. 418-422, 1993.

JENSEN, E.S; HAUGGAARD-NIELSEN, H. How can increased use of biological N₂ fixation in agriculture benefit the environment? **Plant and Soil**, v. 252, p. 177–186, 2003.

JOAQUIM, L. M.; EL-HANI, C. N.; A genética em transformação: crise e revisão do conceito de gene. **Scientiæ Zudia**, v. 8, n. 1, p. 93-128, 2010.

JORDAN, D.C. Rhizobiaceae Conn 1938. In: KRIEG, N. R.; HOLT, J.G., eds. **Bergey's manual of systematic bacteriology**. Baltimore, London: Williams & Wilkins, 1984. p. 235-244.

KANEHISA, M.; GOTO, S.; SATO, Y.; FURUMICHI, M.; TANABE, M. KEGG for integration and interpretation of large-scale molecular data sets. **Nucleic Acids Research**, v. 40, n. D1, p. D109–D114, 2012.

KANEKO, T. et al. Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. **DNA Research**, v. 7, n. 6, p. 321-338, 2000.

KUBISTA, M.; ANDRADE J. M.; BENGTSSON, M.;FOROOTAN, A.; JONÁK, J.; LIND, K.; SINDELKA, R.; SJÖBACK, R.; SJÖGREEN, B.; STRÖMBOM, L.; STÅHLBERG, A.; ZORIC, N. The real-time polymerase chain reaction. **Molecular Aspects of Medicine**, v. 27, p. 95-125, 2006.

KUSANO, M.; FUKUSHIMA, A.; REDESTIG, H.; SAITO, K. Metabolomic approaches toward understanding nitrogen metabolism in plants. **Journal of Experimental Botany**, v. 62, p. 1439-1453, 2011.

KUYKENDALL, L.; SAXENA, B.; DEVINE, T. Genetic diversity in *Bradyrhizobium japonicum* Jordan 1982 and a proposal for *Bradyrhizobium elkanii* sp. nov. **Canadian Journal of Microbiology**, v. 38, p. 501-505, 1992.

LANG, K.; LINDEMANN, A.; HAUSER, F.; GÖTTFERT, M. The genistein stimolon of *Bradyrhizobium japonicum*. **Molecular Genetics and Genomics**, v. 279, n. 3, p. 203–211, 2008.

LI, Q. Q.; WANG, E. T.; CHANG, Y. L.; ZHANG, Y. M.; SUI, X. H.; CHEN, W. F.; CHEN, W. X. *Ensifer sojae* sp. nov., isolated from root nodules of *Glycine max* grown in saline-alkaline soils. **International Journal of Systematic and Evolutionary Microbiology**, v. 61, p. 1981-1988, 2011.

LIPMAN, D. J.; SOUVOROV, A.; KOONIN, E. V.; PANCHENKO, A. R.; TATUSOVA, T. A. The relationship of protein conservation and sequence length. **BMC Evolutionary Biology**, v. 2, n. 1, p. 20, 2002.

LÓPEZ-FUENTES, E.; TORRES-TEJERIZO, G.; CERVANTES, L.; BROM, S. Genes encoding conserved hypothetical proteins localized in the conjugative transfer region of plasmid pRet42a from *Rhizobium etli* CFN42 participate in modulating transfer and affect conjugation from different donors. **Frontiers in Microbiology**, v. 5, p. 1–10, 2015.

MADDEN, T. The BLAST Sequence Analysis Tool. 2013. The NCBI Handbook [Internet]. 2nd Edition. Disponível em: <<http://www.ncbi.nlm.nih.gov/books/NBK153387/>> Acesso em: 10/12/2015.

MARCELINO, F. C. **Avaliação de resíduos de transgênicos em alimentos no Brasil e desenvolvimento de metodologias de análise**. 145f. Tese (Doutorado em Genética e Melhoramento), Universidade Federal de Viçosa, Viçosa, 2006.

MARCELLIN, E.; LICONA-CASSANI, C.; MERCER, T. R.; PALFREYMAN, R. W.; NIELSEN, L. K. Re-annotation of the *Saccharopolyspora erythraea* genome using a systems biology approach. **BMC Genomics**, v. 14, n. 1, p. 699, 2013.

MERCANTE, F. M. **Uso de inoculante garante economia de 3 bilhões de dólares na cultura soja no país**. 2005. Disponível em: <http://www.embrapa.br/noticias/artigos/folder.20050202.1550581232/artigo.2005-12-05.0506770395/mostra_artigo>. Acesso em: 20/09/2015

MESA, S.; REUTIMANN, L.; FISCHER, H. M.; HENNECKE, H. Posttranslational control of transcription factor FixK2, a key regulator for the *Bradyrhizobium japonicum*-soybean symbiosis. **Proceedings of the National Academy of Sciences of the United States of America**, v. 106, n. 51, p. 21860–21865, 2009.

METZKER, M. L. Sequencing technologies - the next generation. **Nature reviews. Genetics**, v. 11, n. 1, p. 31–46, 2010.

NAKANISHI, K.; NUREKI, O. Recent progress of structural biology of tRNA processing and modification. **Molecules and Cells**, v. 19, p. 157-166, 2005.

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.

Nodulation Characterization and proteomic profiling of *Bradyrhizobium liaoningense* CCBAU05525 in response to water-soluble humic materials. **Scientific Reports**, n. 10836, 2015.

PANDEY, G.; KUMAR, V.; STEINBACH, M.; **Computational Approaches for Protein Function Prediction: A Survey**. Technical report, Department of Computer Science and Engineering, University of Minnesota, Twin Cities, 2006.

PEARSON W. R. An introduction to sequence similarity ("homology") searching. **Current protocols in bioinformatics** / editorial board, Baxevanis, A. D. [et al.], Capítulo 3: Unidade t3.1, 2013.

PEREIRA, S. S. **Análise da expressão de fatores de transcrição relacionados com tolerância à seca em raízes de soja**. 127f. Dissertação (Mestrado em Genética e Biologia Molecular) – Universidade Estadual de Londrina, Londrina, 2008.

PERES, J.R.R.; MENDES, I.C.; SUHET, A.R.; VARGAS, M.A. T. Eficiência e competitividade de estirpes de rizóbio para soja em solos de cerrado. **Revista Brasileira de Ciência do Solo**, v. 17, p. 357-363, 1993.

PESSI, G.; AHRENS, C. H.; REHRAUER, H.; et al. Genome-wide transcript analysis of *Bradyrhizobium japonicum* bacteroids in soybean root nodules. **Molecular plant-microbe interactions: MPMI**, v. 20, n. 11, p. 1353–1363, 2007.

POPTSOVA, M.S., GOGARTEN, J.P., Using comparative genome analysis to identify problems in annotated microbial genomes. **Microbiology**, v. 156, 1909–1917, 2010.

PRICE, M. N.; DEUTSCHBAUER, A. M.; KUEHL, J. V.; et al. Evidence-based annotation of transcripts and proteins in the sulfate-reducing bacterium *Desulfovibrio vulgaris* Hildenborough. **Journal of Bacteriology**, v. 193, n. 20, p. 5716–5727, 2011.

PRUITT, K. D. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. **Nucleic Acids Research**, v. 33, p. D501–D504, 2005.

PUNTA, M.; FINN, R. D.; MISTRY, J.; et al. The Pfam protein families database. **Nucleic Acids Research**, v. 40, n. November 2011, p. D211–D222, 2012.

QIAO, J., SHAO, M., CHEN, L., WANG, J., WU, G., TIAN, X., LIU, J., HUANG, S., ZHANG, W. Systematic characterization of hypothetical proteins in *Synechocystis* sp. PCC 6803 reveals proteins functionally relevant to stress responses. **Gene**, v. 1 p. 6-15, 2013.

QIU e CHANG. The origin and history of soybean. In: SINGH, G. **The Soybean: Botany, Production and uses**. CABI, UK. Cap. 1, p. 1-23. 2010.

RAMOS, B. D. **Caracterização molecular de proteínas hipotéticas diferencialmente expressas em tripomastigota metacíclico**. 99f. Dissertação (Mestrado em Biociências e Biotecnologia), Instituto Carlos Chagas, 2014.

ROCHA, T. L.; COSTA, P. H. A.; MAGALHÃES, J. C. C.; EVARISTO, R. G. S.; VASCONCELOS, E. A. R.; COUTINHO, M. V.; PAES, N. S.; SILVA, M. C. M.; GROSSI-DE-SÁ, M. F. Eletroforese bidimensional e análise de proteomas. Brasília: Embrapa Recursos Genéticos e Biotecnologia, 2005. 12p. (Embrapa Recursos Genéticos e Biotecnologia. **Comunicado Técnico**, 136).

SANTOS, M. A. **Caracterização genotípica e fenotípica de *Bradyrhizobium noduladoras de soja (Glycine max (L.) Merrill)* isolado da região dos cerrados**. Dissertação (Mestrado em Microbiologia) - Departamento de Microbiologia – CCB, Universidade Estadual de Londrina, Londrina. 1998.

SANTOS, M. H. R.; BITTENCOURT, J. V. M.; SEVERO, B. B.; FRANGO, V. Avaliação do desempenho de inoculante à base de *Bradyrhizobium japonicum* em cultura de soja na cidade de palmeira – Paraná. Anais do VIII EPEGE - **Encontro Paranaense de Empreendedorismo e Gestão Empresarial**, Paraná, 2011.

SCHENA, M.; SHALON, D.; DAVIS, R. W.; BROWN, P. O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. **Science**, v. 270, p. 467–470, 1995.

SILVA BATISTA, J. S.; HUNGRIA, M. Proteomics reveals differential expression of proteins related to a variety of metabolic pathways by genistein-induced *Bradyrhizobium japonicum* strains. **Journal of Proteomics**, v. 75, n. 4, p. 1211–1219, 2012.

SILVA BATISTA, J. S.; HUNGRIA, M.; BARCELLOS, F. G.; FERREIRA, M. C.; MENDES, I. C. Variability in *Bradyrhizobium japonicum* and *B. elkanii* seven years after introduction of both the exotic microsymbiont and the soybean host in a cerrados soil. **Microbial Ecology**. v. 53, p. 270–284, 2007.

SILVA BATISTA, J. S.; TORRES, A. R.; HUNGRIA, M. Towards a two-dimensional proteomic reference map of *Bradyrhizobium japonicum* CPAC 15: Spotting “hypothetical proteins”. **Proteomics**, n. 10, p. 3176-3189, 2010.

SILVA, P. F. F. **Caracterização funcional de proteínas hipotéticas do fungo patogênico humano *Paracoccidioides* sp.** 133f. Dissertação (Mestrado em Genética e Biologia Molecular), Universidade Federal de Goiás, Instituto de Ciências Biológicas, Goiânia, 2014.

SIQUEIRA, A. F.; ORMEÑO-ORRILLO, E.; SOUZA, R. C.; et al. Comparative genomics of *Bradyrhizobium japonicum* CPAC 15 and *Bradyrhizobium diazoefficiens* CPAC 7: elite model strains for understanding symbiotic performance with soybean. **BMC genomics**, v. 15, p. 2 - 20, 2014.

SOUZA, D. S. **Uma ferramenta multiagente baseada em conhecimento para anotação de proteínas: um estudo de caso para o fungo *Saccharomyces cerevisiae***. 79f. Dissertação (Mestrado em Informática). Universidade de Brasília, Brasília, 2014.

STACEY, G.; LIBAULT, M.; BRECHENMACHER, L.; WAN, J.; MAY, G. D. Genetics and functional genomics of legume nodulation. **Current Opinion in Plant Biology**, v. 9, p. 110-121, 2006.

- TAKESHIMA, K.; HIDAKA, T.; WEI, M.; et al. Involvement of a novel genistein-inducible multidrug efflux pump of *Bradyrhizobium japonicum* early in the interaction with *Glycine max* (L.) Merr. **Microbes and Environments**, v. 28, n. 4, p. 414–421, 2013.
- 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.; DENARDIN, N. D.; VAN BERKUM, P. Tolerância de rizóbios nativos do feijoeiro a antibióticos e sua relação com fatores de acidez do solo. **Revista Brasileira de Ciência do Solo**, v.16, p.331-336, 1992.
- 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.
- VELCULESCU, V. E.; ZHANG, L.; VOGELSTEIN, B.; KINZLER, K. W. Serial analysis of gene expression. **Science**, v. 270, p. 484–487, 1995.
- VINUESA, P.; LEÓN-BARRIOS, M.; SILVA, C.; WILLEMS, A.; JARABO-LORENZO, A.; PÉREZ-GALDONA, R.; WERNER, D.; MARTÍNEZ-ROMERO, E. *Bradyrhizobium canariense* sp. nov., an acid-tolerant endosymbiont that nodulates endemic genistoid legumes (*Papilionoideae: Genisteae*) from the Canary Islands, along with *Bradyrhizobium japonicum* bv. *genistearum*, *Bradyrhizobium* genospecies alpha and *Bradyrhizobium* genospecies beta. **International Journal of Systematic and Evolutionary Microbiology**, v. 55, p. 569-575, 2005.
- VONGSANGNAK, W.; OLSEN, P.; HANSEN, K.; KROGSGAARD, S.; NIELSEN, J. Improved annotation through genome-scale metabolic modeling of *Aspergillusoryzae*. **BMC Genomics**, v. 9, n. 1, p. 245, 2008.
- WANG, J. Y.; WANG, R.; ZHANG, Y. M.; LIU, H. C.; CHEN, W. F.; WANG, E. T.; SUI, X. H.; CHEN, W. X. *Bradyrhizobium daqingense* sp. nov., isolated from soybean nodules. **International Journal of Systematic and Evolutionary Microbiology**, v. 63, p. 616-624, 2013.
- WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nature Reviews Genetics**, v. 10, p. 57-63, 2009.
- WEI, M.; YOKOYAMA, T.; MINAMISAWA, K.; et al. Soybean seed extracts preferentially express genomic loci of *Bradyrhizobium japonicum* in the initial interaction with soybean, *Glycine max* (L.) Merr. **DNA Research**, v. 15, n. 4, p. 201–14, 2008.
- WILLEMS, A. The taxonomy of rhizobia: an overview. **Plant and Soil**, v. 287, p. 3-14, 2006.
- XU, L. M.; GE, C.; LI, J.; FAN, H. *Bradyrhizobium liaonigense* sp. nov., isolated from the nodules of soybeans. **International Journal of Systematic and Evolutionary Microbiology**, n. 45, p. 706-711, 1995.
- YAO, Z. Y.; KAN, F. L.; WANG, E. T.; WEI, G. H.; CHEN, W. X. Characterization of rhizobia that nodulate legume species of the genus *Lespedeza* and description of

Bradyrhizobium yuamingense sp. nov. **International Journal of Systematic and Evolutionary Microbiology**, v. 55, p. 2219-2230, 2002.

Zhang, Y. M.; Li, Y. Jr.; Chen, W. F.; Wang, E. T.; Sui, X. H.; Li, Q. Q.; Zhang, Y. Z.; Zhou, Y. G.; Chen, W. X. *Bradyrhizobium huanghuaihaiense* sp. nov., an effective symbiotic bacterium isolated from soybean (*Glycine max* L.) nodules. **International Journal of Systematic and Evolutionary Microbiology**, n. 62, p. 1951-1957, 2012.

CAPÍTULO 2

INFERÊNCIA FUNCIONAL DE GENES HIPOTÉTICOS DA ILHA SIMBIÓTICA DA ESTIRPE CPAC 15 DE *Bradyrhizobium japonicum**

*Esse artigo será submetido à publicação na revista BMC Genomics

INFERÊNCIA FUNCIONAL DE GENES HIPOTÉTICOS DA ILHA SIMBIÓTICA DA ESTIRPE CPAC 15 DE *Bradyrhizobium japonicum*

Everton Geraldo Capote Ferreira¹, Fernando Gomes Barcellos¹, Mariangela Hungria²

¹Departamento de Biologia Geral, Universidade Estadual de Londrina, CCB, CEP 86051-970, Londrina, Paraná, Brasil.

²Laboratório de Biotecnologia dos Solos, Embrapa-Soja, Rodovia Carlos João Strass - Distrito de Warta, CEP 86001-970, Londrina, Paraná, Brasil.

RESUMO

A fixação biológica do nitrogênio (FBN) na soja (*Glycine max* (L.)) ocorre através da simbiose com bactérias (rizóbios), como a espécie *Bradyrhizobium japonicum*, seu principal simbiote. O presente estudo fornece a análise da expressão e inferência funcional de genes localizados dentro da ilha simbiótica da estirpe CPAC 15 de *Bradyrhizobium japonicum* e que codificam para proteínas hipotéticas. A inferência funcional foi realizada com base em busca de similaridades com diversos bancos de dados e programas de predições de estruturas como peptídeos sinais, localização celular, domínios conservados, entre outros. A avaliação da expressão gênica foi realizada através de RT-qPCR com indução com genisteína em meio de cultivo com a bactéria, por aproximadamente, 48 horas. As análises de similaridade com o banco de dados RefSeq demonstrou que todas as proteínas hipotéticas estavam distribuídas em vários genomas de diferentes espécies do gênero *Bradyrhizobium*. Os resultados obtidos por RT-qPCR demonstraram que todos os genes hipotéticos selecionados obtiveram elevados níveis de expressão em presença de genisteína em relação ao controle, o metanol. A maioria dos genes codificando para proteínas hipotéticas inferidos funcionalmente estava relacionada ao transporte de substâncias, vias metabólicas de aminoácidos e degradação de compostos, e possivelmente então, podendo ser úteis na atividade saprofítica da estirpe e na FBN em si.

Palavras -chaves: Proteínas Hipotéticas. Genisteína. Ilha Simbiótica. FBN. Anotação Funcional. RT-qPCR.

INTRODUÇÃO

Existe uma grande variabilidade entre as estirpes que nodulam a soja, quanto à eficiência do processo simbiótico e à capacidade competitiva com bactérias estabelecidas no solo (HUNGRIA et al., 2006; SILVA BATISTA et al., 2007; CHEN et al., 2015). Somado a isto, fatores como o constante aprimoramento das práticas agrícolas, lançamento de cultivares mais produtivas e adaptadas as distintas condições ambientais, além do estabelecimento de populações de bactérias em solos previamente inoculados, fazem com que seja extremamente necessário uma constante busca por genótipos de *Bradyrhizobium* mais competitivos no solo e eficientes no processo simbiótico com a soja.

Recentemente, a estirpe CPAC 15 de *Bradyrhizobium japonicum* teve o seu genoma sequenciado (SIQUEIRA et al., 2014). Esta estirpe tem um genoma de 9,582,287-bp composto de um cromossomo circular com dois operons ribossomais e um conteúdo G+C de 63.54% (SIQUEIRA et al., 2014). De acordo com a anotação realizada, foram preditas 4.203 ORFs com função atribuída (48.6%) e 4.445 ORFs hipotéticas (51.4%), totalizando um total de 8.648 genes preditos.

O início da ilha simbiótica da CPAC 15 foi demarcada em um gene de DNA-recombinase e o final da ilha simbiótica foi definida em um gene para valina tRNA. Dentro da ilha simbiótica há um gene para metionina tRNA. O tamanho da ilha simbiótica é de 700.213 pb, com um total de 569 ORFs, sendo que destas, 247 são hipotéticas (43%), além de 90 elementos móveis. A ilha simbiótica da CPAC 15 engloba todos os genes de nodulação e fixação do N₂ que são requeridos para a simbiose, uma provável operon não funcional de hidrogenase, citocromos para fornecimento de energia, transportadores do tipo ABC, operons para sistemas de secreção, ressaltando-se que todos esses genes são rodeados por muitos genes hipotéticos e elementos genéticos móveis, que podem estar relacionados a transferência gênica horizontal.

O escasso conhecimento sobre esse importante rizóbio foi enfatizado pela observação de que não foi possível atribuir função para aproximadamente 50% dos genes preditos do genoma da estirpe da CPAC 15, incluindo genes da ilha simbiótica. Intensos estudos são necessários tanto a nível transcricional quanto a nível proteômico a fim de elucidar as funções que estes genes desempenham nesta estirpe, e desta forma, lançar

perspectivas para um aumento da compreensão das características superiores que este rizóbio apresenta.

MATERIAL E MÉTODOS

ESTIRPE BACTERIANA E CONDIÇÕES DE CULTIVO

A estirpe CPAC 15 de *B. japonicum* utilizada foi proveniente da coleção de culturas denominada: “Coleção de Culturas de Bactérias Diazotróficas e Promotoras do Crescimento de Plantas da Embrapa Soja”, do Laboratório de Biotecnologia dos Solos da Embrapa Soja (Londrina – Paraná).

A estirpe foi retirada do estoque a -80°C , repicada em meio de cultura YM sólido (VINCENT, 1970), com adição de corante vermelho congo (0,025%), e crescida a 28°C por 48 horas. Posteriormente, colônias desta estirpe crescida em placa foram inoculadas em 10 mL de meio de cultura TY líquido (BERINGER, 1974), e esta cultura (pré-inóculo) foi mantida sob agitação (100 rpm) a 28°C por, aproximadamente, 24 horas, até atingir a densidade óptica de 0,6 ($\lambda = 600 \text{ nm}$). Posteriormente, 1 mL do pré-inóculo foi adicionado em cada uma das réplicas biológicas, cada uma com 100 mL de meio de cultura TY líquido (inóculo) acrescido com genisteína ou metanol.

EXTRAÇÃO E QUANTIFICAÇÃO DO RNA TOTAL

Para a extração de RNA empregou-se o protocolo básico usando Trizol, (*Invitrogen - Life Technologies*) com modificações. Primeiramente, foram centrifugados 30 mL das culturas a 10000 rpm a 4°C , descartado o sobrenadante e adicionado 1 mL de solução salina 0,85%, os precipitados foram então ressuspensos e transferidos para microtubos de 2 mL estéreis. A seguir, houve mais uma etapa de centrifugação nas mesmas condições descritas acima, após, novamente os sobrenadantes foram descartados e os *pellets* ressuspensos em 300 μL de tampão de lise celular (250 μL de TE 10:1; 40 μL de SDS 10%; 10 μL de lisozima a 5mg/mL) e incubados por 5 minutos à 40°C . Logo em seguida, adicionou-se 1,5 mL de trizol, as amostras passaram pelo vórtex por 15 segundos e foram incubadas por 5 minutos a temperatura ambiente. Atingindo os 5 minutos, as amostras foram centrifugadas a 14000 rpm por 10 minutos, a 4°C . Transferiram-se aproximadamente 800 μL

dos sobrenadantes para novos microtubos, nestes adicionaram-se 250 μL de clorofórmio, vórtex por 15 segundos, e mais uma incubação por 5 minutos em temperatura ambiente e centrifugação por 15 minutos a 14.000 rpm. As fases aquosas foram então transferidas cuidadosamente para novos tubos (400 μL) e foram adicionados 500 μL de isopropanol gelado, inverteram-se os microtubos cerca de 20 vezes para melhor precipitação do *pellet* e, a seguir, incubaram-se as amostras por 20 minutos no gelo. Centrifugaram-se as amostras por 15 minutos a 14000 rpm e 4°C, descartou-se o sobrenadante e os *pellets* foram ressuspensos em água ultra pura livre de RNase. Adicionaram-se 10 μL de acetato de sódio 3 M e 250 μL de etanol absoluto gelado, inverteram-se os microtubos cerca de 20 vezes para precipitação e procedeu-se a mais uma etapa de centrifugação por 15 minutos a 14000 rpm – 4°C. A seguir, descartou-se o sobrenadante e adicionaram-se 500 μL de etanol 70% gelado para lavagem dos *pellets* e procedeu-se a outra etapa de centrifugação nas mesmas condições da centrifugação anterior. Em seguida, os sobrenadantes foram descartados, os *pellets* foram deixados na bancada para secagem por cerca de 30 minutos, após, foram ressuspensos em 50 μL de água ultra pura livre de RNase e armazenados a -80°C.

O RNA extraído foi quantificado usando micro-espectrofotômetro (*Nanodrop[®] Spectrophotometer* ND-1000, Uniscienze), com leituras nos comprimentos de onda de 260 e 280 nm, utilizando água ultra pura livre de RNase como controle.

A verificação da integridade dos RNAs extraídos foi constatada por eletroforese em gel de agarose 1%, preparado com tampão SB (ácido bórico de sódio) (1x) (BRODY e KERN, 2004), corado com brometo de etídio. Adicionaram-se no gel, 2 μL de RNA com 1 μL de tampão de amostra (sacarose 40% mais azul de bromofenol). A corrida eletroforética foi conduzida em tampão SB (1x) durante 30 minutos a uma voltagem de 50 V. E os géis foram visualizados em transiluminador UV e fotografados.

ENSAIO DA EXPRESSÃO DOS GENES EM RESPOSTA A GENISTEÍNA

A solução estoque de genisteína (Sigma; Peso molecular = 270g) foi preparada na concentração final de 1000 μM , a partir da dissolução de 0,013 g da mesma diluída em 50 mL de metanol. Para a indução dos genes nas réplicas biológicas de *B. japonicum* foi utilizada a concentração de 2,5 μM .

O experimento seguiu um delineamento em blocos ao acaso com três réplicas como blocos, sendo o tratamento realizado com genisteína e o controle com metanol,

o tempo de indução foi de, aproximadamente, 48 horas (BORTOLAN et al., 2009). Os inóculos acrescidos com genisteína ou metanol cresceram por, aproximadamente, 48 horas, ou até atingirem a densidade ótica (D.O) que corresponde ao crescimento exponencial (D.O em torno de 0,5 em $\lambda = 600$ nm). Após atingirem o crescimento exponencial, os inóculos foram submetidos à extração de RNA.

TRATAMENTO COM DNase-I E SÍNTESE DO DNA COMPLEMENTAR (cDNA)

A digestão da DNase- I (*Invitrogen – Life Technologies*) foi realizada de acordo com as especificações do fabricante, com algumas modificações. A quantidade de RNA tratado foi de 5 μ g, utilizando 1,5 μ L de enzima (5 U/ μ L) e 1,5 μ L de tampão. A reação foi incubada à temperatura ambiente por 8 minutos. A seguir, adicionou-se 1 μ L de EDTA 25 mM e incubaram-se as amostras a 60°C durante 5 minutos.

O cDNA foi sintetizado com o kit *Super Script III Reverse Transcriptase* (*Invitrogen - Life Technologies*) utilizando *random primers*, conforme instruções do fabricante.

SELEÇÃO E ANOTAÇÃO FUNCIONAL DAS SEQUÊNCIAS ALVO CODIFICADORAS DE PROTEÍNAS HIPOTÉTICAS DA ILHA SIMBIÓTICA

Foram identificados genes hipotéticos (“ORFs” hipotéticas) localizados na ilha simbiótica da estirpe CPAC 15 de *B. japonicum*. A identificação dos genes foi realizada com base na anotação “automática” destes genomas disponibilizada pelo Laboratório de Bioinformática (LABINFO) do Laboratório Nacional de Computação Científica (LNCC), em Petrópolis – RJ, no endereço eletrônico: <http://www.labinfo.lncc.br/index.php/projects/biological-nitrogen-fixation> ou pela sequência referência no NCBI (NZ_CP007569.1).

A inferência funcional das proteínas hipotéticas codificadas por esses genes foi realizada com base na busca de similaridades entre as sequências de aminoácidos destes genes com sequências depositadas nos bancos de dados PFAM (<http://pfam.xfam.org/>), KEGG (<http://www.genome.jp/kegg/>), UniProtKB (<http://www.uniprot.org/>), Rhizobase (genome.microbedb.jp/rhizobase), NCBI – BLASTn e BLASTp

(<http://blast.ncbi.nlm.nih.gov/Blast.cgi>), Interpro (<https://www.ebi.ac.uk/interpro/>), Gene Ontology – GO (<http://geneontology.org/>).

DESENHO DOS *PRIMERS* PARA AS SEQUENCIAS ALVO SELECIONADAS

Foram utilizadas as sequências de genes hipotéticos localizados na ilha simbiótica, e utilizando o software *Primer Express* versão 3.0 (*Applied Biosystems*, Foster, CA, USA), procedeu-se à construção dos *primers* para o RT-qPCR, utilizando os parâmetros *default* para construção de *primers* para RT-qPCR, sendo os fragmentos amplificados por estes *primers* entre uma faixa de tamanho de 50 à 150 pares de bases (pb). Após a obtenção das sequências dos *primers*, os mesmos foram contrastados contra o genoma da estirpe CPAC 15 através da ferramenta BLASTn do NCBI e verificados quanto a sua especificidade, ou seja, evitando que as sequências anelassem em outras regiões do genoma, o que aumentaria assim, as chances de amplificação inespecífica. Na tabela 1 podem ser visualizadas as sequências empregadas neste estudo. As sequências dos *primers* para o gene normalizador *16S rRNA* e do gene utilizado como controle positivo (*nodC*) são as sequências de *B. japonicum* CPAC 15 sintetizadas e empregadas por Bortolan et al. (2009).

Tabela 1 - Sequências dos *primers* dos genes utilizados no RT-qPCR

Gene	Sequência Nucleotídica	Amplicon
<i>bjs08317</i>	F AAATAGTCCGGCTGACAAGG R CATAGGTCTTCGTGCAATCG	102 pb
<i>bjs08261</i>	F CGTGATCAATCTCGTCTTCG R TGAAGGAGACTTTCCGGATG	83 pb
<i>bjs08258</i>	F ATCCGAGCTGCCATTACATC R GCAATGTTGTTCCGGCTGTAG	91 pb
<i>bjs08254</i>	F GGTCGGCTATTTCAAGATCG R GATAGCCCGAAATGTTGACG	74 pb
<i>bjs08251</i>	F CAACGATGATACGGTTGCAG R GCATGACCTTGATTGCCTTC	131 pb
<i>bjs08216</i>	F GATCCACGGGATGTTTGAAG R CGCAGATCAATTCACGTAGC	137 pb
<i>bjs08267</i>	F TCCTGATCTGGATGTCAACG R AGAAGCATGGGATGAGCAAC	105 pb

ANÁLISE DE EXPRESSÃO POR RT-qPCR (PCR QUANTITATIVO EM TEMPO REAL)

As reações de RT-qPCR foram realizadas no termociclador ABI Prism 7500 *Sequence Detection System* (Applied Biosystems, Foster, CA, EUA). As reações foram realizadas em uma placa de 96 poços, sendo constituídas de três replicatas técnicas de cada repetição biológica, no caso da quantificação relativa e de duas replicatas técnicas de cada reação, na análise de eficiência dos *primers*. Para realizar as reações de RT-qPCR foi usado o *kit Platinum SYBR Green qPCR SuperMix-UDG* com ROX (Invitrogen – Life Technologies). Curvas de eficiência de amplificação para cada par de *primers* dos genes alvos e para o par de *primers* do gene normalizador e controle positivo foram realizadas. Para isto, partiu-se de um *mix* de cDNAs (5µL de cada uma das repetições biológicas da situação controle e tratamento) e deste, fez-se uma série de diluições, utilizando um fator de diluição 1:10.

Os valores de Ct (*cycle threshold*) obtidos em cada diluição foram plotados em função do logaritmo das diluições, e o índice de inclinação da reta foi utilizado para calcular a eficiência de cada sistema, de acordo com a fórmula $E=[10^{(-1/-slope)-1}]$. Os parâmetros de ciclagem utilizados para as reações de eficiência dos *primers* e quantificação relativa foram: 50°C por 2 minutos, 95°C por 10 minutos, seguidos de 45 ciclos de 95°C por 15 segundos, 60°C por 1 minuto, sendo os dados coletados neste último passo. Os parâmetros do programa de amplificação empregados são os recomendados pelo próprio fabricante do termociclador. A concentração dos *primers* com melhor desempenho na curva de eficiência foi selecionada para utilização na etapa de quantificação relativa. As reações de PCR constituíram-se de 4,25 µL de água ultra pura autoclavada; 6,25 µL de SYBR Green com ROX; 0,5 µL tanto do *primer forward* e do *primer reverse* (5 µM); 1 µL de cDNA, obtendo um volume final de reação de 12,5µL.

O cálculo da expressão e análise estatística foi realizada utilizando o *software* REST 2009 (PFAFFL et al., 2002). Esse *software* realiza a quantificação comparativa pelo método de “*Pair-Wise Fixed Reallocation Randomization Test*” (PFAFFL et al., 2002). Esse *software* realiza a quantificação relativa ao comparar um grupo de amostras com o grupo controle, a comparação sempre é realizada a partir de um gene normalizador (gene expresso na mesma quantidade e nas condições experimentais avaliadas) (NACHTIGALL, 2012).

RESULTADOS

ANÁLISE IN SILICO DOS GENES CODIFICANDO PARA PROTEÍNAS HIPOTÉTICAS

Vários critérios foram aplicados a fim de escolher os genes codificando para proteínas hipotéticas. Neste sentido, objetivou-se, primeiramente, escolher genes que estivessem numa faixa entre 600 pb até 2500 pb, como demonstrado na Tabela 2. Esse critério foi aplicado devido ao fato que cada vez mais estudos vêm mostrando que genes codificando para proteínas hipotéticas com tamanho menores que essa faixa tem grande probabilidade de serem pseudogenes ou serem recorrentes de erros de sequenciamento e anotação (DESLER et al, 2009, POPTSOVA e GOGARTEN, 2010).

Tabela 2 – Tamanho dos genes hipotéticos selecionados da ilha simbiótica da estirpe CPAC 15

Gene	NCBI ID da Proteína Hipotética	Locus	Tamanho em Pares de Bases (pb)
<i>bjs08216</i>	gi 627779337	AHY48554	1575 pb
<i>bjs08251</i>	gi 627779303	AHY48520	2100 pb
<i>bjs08258</i>	gi 627779297	AHY48514	1299 pb
<i>bjs08254</i>	gi 627779300	AHY48517	900 pb
<i>bjs08261</i>	gi 627779294	AHY48511	1035 pb
<i>bjs08267</i>	gi 627779288	AHY48505	1197 pb
<i>bjs08317</i>	gi 627779238	AHY48455	2244 pb

O segundo critério aplicado foi o da localização destes genes dentro da ilha simbiótica do genoma da estirpe de *B. japonicum*. Tentou-se escolher genes que estivessem localizados relativamente próximos de genes já validados e/ou de genes essenciais a atividade saprofítica e/ou FBN, como os de nodulação; sistemas de secreção, entre outros, tentando desta forma, aumentar a probabilidade destes genes estarem sob influência dos mesmos mecanismos de regulação gênica, desempenharem papéis importantes no mesmo processo biológico ou serem parte de algum operon ou *cluster* gênico. O gene *bjs08317* está próximo do gene *nopP* (Figura 2) que é um efetor do sistema de secreção do tipo III e desta forma,

desempenha funções no processo de infecção da bactéria ao hospedeiro vegetal (AUSMESS et al., 2004).

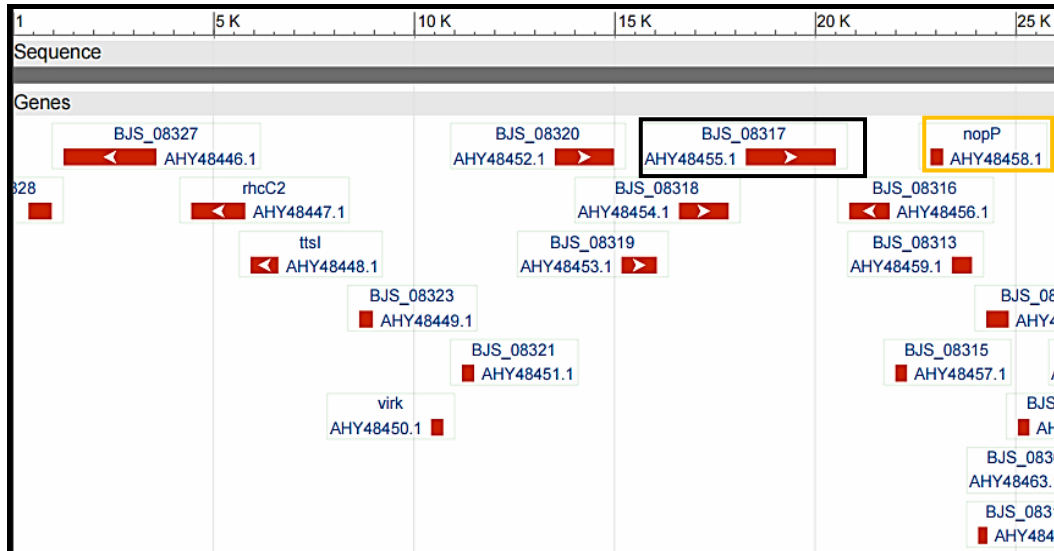


Figura 1 – Localização do gene *bjs08317* no genoma da estirpe CPAC 15 de *B. japonicum*. Em preto, representado o gene *bjs08317*, em amarelo o gene *nopP* que está relativamente próximo no contexto genômico e que tem função conhecida, e os números em azul representa a identificação do gene no NCBI.

O gene *bjs08216* está próximo do gene *fixW* (Figura 3), que codifica para uma oxidoreductase, necessária no processo de FBN (HONTELEZ et al., 1989).

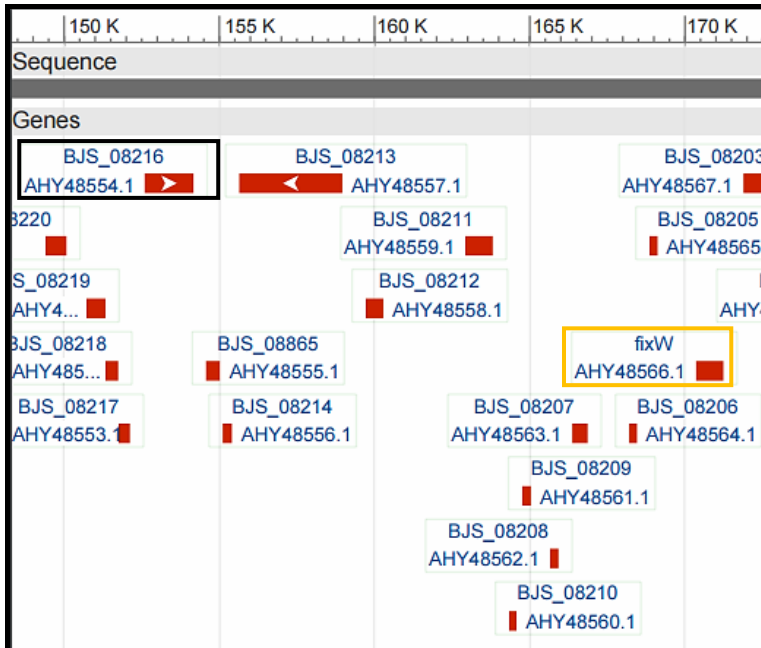


Figura 2 – Localização do gene *bjs08216* no genoma da estirpe CPAC 15 de *B. japonicum*. Em preto, representando o gene *bjs08216*, em amarelo o gene de fixação *fixW*, em azul o número de identificação do gene no NCBI.

Os genes *bjs08251*, *bjs08254*, *bjs09258*, *bjs08261* e *bjs08267* estão localizados próximos do gene *bjs08257*, de nodulação *nodU*, este gene por sua vez, está anotado no NCBI como o gene *nodU* ((Figura 4). Com destaque para o gene *bjs08258*, que é o mais próximo do gene *nodU*, este gene está anotado no NCBI como o gene *aspC*. No entanto, foi considerado como hipotético e realizada a inferência funcional do mesmo para confirmação da sua função anotada anteriormente. O gene *nodU* codifica para uma Carbamoil transferase, que atua na biossíntese dos fatores Nod, sendo essencial ao processo de nodulação (FREIBERG et al., 1997).

ESPADALER et al., 2009), no entanto, devido à escassez de informações na busca por homologia, foi utilizado o valor de 30%.

É importante frisar que enquanto algumas proteínas apresentaram altos níveis de identidade com proteínas encontradas em outros genomas bacterianos, outras apresentaram níveis baixos de identidade com proteínas conhecidas. Estas proteínas com baixo nível de identidade podem constituir-se como interessantes alvos de pesquisas futuras, tendo em vista que estas proteínas hipotéticas são conservadas em outros genomas de *Bradyrhizobium*, como demonstrado na Figura 8, podendo então, serem proteínas que sofreram processos evolutivos e agora são específicas e codificadas por esse gênero, contudo, análises posteriores e estudos mais aprofundados de filogenia destas proteínas, assim como estudos comparativos estruturais destas proteínas devem ser realizados para inferir com maior confiabilidade esta hipótese. A tabela com os dados compilados da inferência funcional em vários bancos de dados encontra-se no Material Suplementar.

Gene *bjs08216*

De acordo com a predição da localização celular realizada pelos programas *PSORT-B*, *PSLPred*, a proteína hipotética codificada pelo gene *bjs08216* está localizada no citoplasma. Através do programa SecretomeP 2.0, a proteína em questão também não é secretada, o que condiz com a sua localização na citoplasma.

A sequência da proteína BJS08216 não possui nenhum domínio conservado de família de proteínas, como observado pela comparação da sequência desta proteína com as proteínas depositadas no *Pfam*.

Através do alinhamento da sequência da proteína codificada pelo gene *bjs08216* com as depositadas pelo banco de dados Uniprot, observou-se 28,4% de identidade da sequência da proteína BJS08216 com as da chaperona TIR (*Burkholderia* sp. USM B20); 28,0% de identidade com a proteína TIR de *Methylobacterium extorquens*; 29,4% de identidade com a chaperona TIR de *Mesorhizobium* sp. LSJC280B00. O processo biológico no qual as chaperonas TIR atuam é a transdução de sinal, de acordo com o *Gene Ontology*.

Pode-se supor que a proteína hipotética codificada pelo gene *bjs08216* possa ser algum tipo de Chaperona que atua na transdução de sinal, como evidenciado pela sua localização celular no citoplasma e identidade com chaperonas de outras bactérias do solo. No entanto, mais estudos mais aprofundados devem ser realizados a fim de predizer com maior

confiabilidade a função desta proteína, devido, principalmente, ao fato da proteína codificada pelo gene *bjs08216* não possui domínio conservado de numa família proteica.

Gene *bjs08251*

De acordo com a predição da localização celular realizada pelos programas *PSORT-B*, *PSLpred*, a proteína hipotética codificada pelo gene *bjs08251* é uma proteína citoplasmática. O programa *SecretomeP 2.0*, descreveu a proteína como não secretada, o que condiz com a sua localização na citoplasma.

A sequência da proteína BJS08251 possui domínio conservado de Tripsina peptidase. Esse domínio pertence à Família das Tripsinas e ao Clã das Peptidases, como observado pela comparação da sequência desta proteína com as proteínas depositadas no *Pfam*.

Através do alinhamento da sequência da proteína codificada pelo gene *bjs08251* com as depositadas pelo banco de dados *Uniprot*, observou-se 35,4% de identidade com uma serina protease de *Bradyrhizobium* sp. ORS 375; 36.9% de identidade com uma serina protease de *Sorangium cellulosum* So ce56.

O processo biológico no qual essas serinas proteases atuam é atividade catalítica, na degradação de proteínas, peptídeos e glicopeptídeos e a função molecular é na atividade catalítica, atividade de serina endopeptidase, de acordo com o *Gene Ontology*. O número EC dado pelo banco de dados *KEGG* para serinas proteases com domínio para tripsina é EC: 3.4.21.107.

Pode-se supor que a proteína hipotética codificada pelo gene *bjs08251* é uma serina protease com domínio para tripsina, ou seja, uma tripsina, que é uma endopeptidase específica para ligações peptídicas contendo resíduos com cadeia lateral com carga elétrica positiva, em condições fisiológicas, tais como a arginina (Arg) e lisina (Lys), na posição R1.

Gene *bjs08258*

De acordo com a predição da localização celular realizada pelos programas *PSORT-B*, *PSLpred*, a proteína hipotética codificada pelo gene *bjs08258* é uma proteína citoplasmática. O programa *SecretomeP 2.0*, descreveu a proteína como não secretada, o que condiz com a sua localização na citoplasma.

A sequência da proteína BJS08258 possui o domínio conservado da família das Aminotransferases de Classe 1 e 2, como observado pela comparação da sequência desta proteína com as proteínas depositadas no *Pfam*.

Através do alinhamento da sequência da proteína codificada pelo gene *bjs08258* com as depositadas pelo banco de dados *Uniprot*, observou-se 55,6% de identidade com a aspartato aminotransferase de *Mesorhizobium metallidurans*, estirpe STM 2683; 53,3% de identidade com a aspartato aminotransferase de *Rhizobium loti*, estirpe MAFF303099; 49,9% de identidade com a aspartato aminotransferase de *Bradyrhizobium* sp. DOA9.

O processo biológico no qual essas aspartato aminotransferases atuam é na Biossíntese, como parte do metabolismo de substâncias e a função molecular é a atividade de transaminase, de acordo com o *Gene Ontology*. A categoria funcional no qual essa aspartato aminotransferase atua é no metabolismo e transporte de Aminoácidos, de acordo com a anotação feita para essa enzima pelo *Rhizobase*. O número EC dado pelo banco de dados *KEGG* para aspartato aminotransferases é EC: 2.6.1.1.

Pode-se supor que a proteína hipotética codificada pelo gene *bjs08258* é uma aspartato aminotransferase, que atua na degradação de aspartato, ou seja, envolvida na degradação de aminoácidos em vias metabólicas.

Gene *bjs08254*

A predição da localização celular realizada pelo programa *PSORT-B* não foi hábil em prever a localização celular da proteína codificada pelo gene *bjs08254*, já o programa *PSLPred* descreveu a localização da proteína no periplasma. E o programa *SecretomeP 2.0* descreveu que a proteína BJS08254 é do tipo secretada.

A sequência da proteína BJS08254 possui o domínio conservado Tripsina, que pertence à família Tripsina, como observado pela comparação da sequência desta proteína com as proteínas depositadas no *Pfam*.

Através do alinhamento da sequência da proteína codificada pelo gene *bjs08258* com as depositadas pelo banco de dados *Uniprot*, observou-se 39,6% de identidade com a glutamil endopeptidase de *Bacillus licheniformis*, estirpe ATCC 14580.

O processo biológico no qual essa enzima atua é na proteólise e a função molecular é a atividade de serina endopeptidase, de acordo com o *Gene Ontology*. A categoria funcional no qual essa enzima encontra-se é na biossíntese de cofatores, grupos prostéticos e

carregadores, de acordo com a anotação feita para essa enzima pelo *Rhizobase*. O número EC dado pelo banco de dados *KEGG* para esta enzima é EC: 3.4.21.19.

Pode-se supor que a proteína hipotética codificada pelo gene *bjs08254* é uma peptidase, no caso, uma glutamil peptidase, que cliva preferencialmente sítios com ácido glutâmico (Glu) e ácido aspártico (Asp). Em relação a localização desta enzima, a maioria das glutamil peptidase são do tipo secretadas, ou seja, extracelulares, como observado no banco de dados *BRENDA – The Comprehensive Enzyme Information System*. Possivelmente então, essa enzima é secretada e atua na degradação de proteínas do meio extracelular da bactéria, podendo participar então do metabolismo de substâncias.

Gene *bjs08261*

De acordo com a predição da localização celular realizada pelos programas *PSORT-B*, *PSLPred*, a proteína hipotética codificada pelo gene localiza-se no citoplasma. O programa *SecretomeP 2.0*, descreveu a proteína como não secretada, o que condiz com a sua localização na citoplasma.

A sequência da proteína BJS08261 possui o domínio proteico da família dos aldeídos desidrogenases, como observado pela comparação da sequência desta proteína com as proteínas depositadas no *Pfam*. De acordo com o banco de dados *PANTHER*, a subfamília desta proteína seria a dos succinatos-semialdeídos desidrogenases.

Através do alinhamento da sequência da proteína codificada pelo gene *bjs08261* com as depositadas pelo banco de dados *Uniprot*, observou-se 71,2% de identidade com a succinato-semialdeído desidrogenase de *Oligotropha carboxidovorans*- estirpe ATCC 49405; 71.8% de identidade com a Aldeído desidrogenase de *Burkholderia* sp. SJ98.

O processo biológico no qual essa enzima atua é na processo metabólico e a função molecular é a atividade de oxidoreductase que atua no grupo aldeído ou oxo dos doadores, de acordo com o *Gene Ontology*. A categoria funcional seria então o metabolismo de pequenas moléculas, de acordo com a anotação do *Rhizobase*. O número EC dado pelo banco de dados *KEGG* para esta enzima é EC: 1.2.1.24.

Pode-se supor que a proteína hipotética codificada pelo gene *bjs08261* é uma succinato-semialdeído desidrogenase, essa enzima desempenha um papel nas vias metabólicas do glutamato e do butanoato, ao catalisar a oxidação do succinato semialdeído a succinato utilizando NAD⁺ como uma coenzima.

Gene *bjs08267*

De acordo com a predição da localização celular realizada pelos programas PSORT-B, PSLpred, a proteína hipotética codificada pelo gene *bjs08267* localiza-se na membrana citoplasmática. O programa *SecretomeP 2.0*, descreveu a proteína como não secretada, o que condiz com a sua localização na citoplasma.

A sequência da proteína BJS08267 possui o domínio proteico da superfamília dos transportadores facilitadores principais, como observado pela comparação da sequência desta proteína com as proteínas depositadas no *Pfam*. A presença desse domínio de transportadores corrobora com a localização desta proteína na membrana citoplasmática, que é onde os integrantes dessa família se localizam.

Através do alinhamento da sequência da proteína codificada pelo gene *bjs08267* com as depositadas pelo banco de dados *Uniprot*, observou-se 79.9% de identidade com um transportador facilitador principal de *Bradyrhizobium* sp. CCGE-LA001.

O processo biológico no qual essa enzima atua é no transporte e a função molecular é a atividade de transporte transmembrana, de acordo com o *Gene Ontology*. A categoria funcional é o transporte e ligação de proteínas, de acordo com a anotação do *Rhizobase*. O número EC dado pelo banco de dados *KEGG* para esta enzima é EC: TC: 2.A.1.6.

Pode-se supor que a proteína hipotética codificada pelo gene *bjs08267* é um transportador facilitador principal, que são carregadores secundários, que são capazes de transportar apenas pequenos solutos em resposta aos gradientes iônicos.

Gene *bjs08317*

De acordo com a predição da localização celular realizada pelos programas PSORT-B, PSLpred, a proteína hipotética codificada pelo gene *bjs08317* localiza-se no citoplasma. O programa *SecretomeP 2.0*, descreveu a proteína como não secretada, o que condiz com a sua localização na citoplasma.

A sequência da proteína BJS08317 não possui nenhum domínio conservado de família de proteínas, como observado pela comparação da sequência desta proteína com as proteínas depositadas no *Pfam*.

Através do alinhamento da sequência da proteína codificada pelo gene *bjs08317* com as depositadas pelo banco de dados Uniprot, observou-se 31,1% de identidade com a proteína efetora XopP do sistema de secreção do tipo III de *Xanthomonas campestris* pv. *Raphani*, estirpe 756C; 31% de identidade com a proteína efetora HLK1 do sistema de secreção do tipo de *Ralstonia solanacearum*, estirpe GMI1000.

O processo biológico no qual proteínas efetoras do sistema de secreção do tipo III atuam é na virulência de bactérias aos hospedeiros, ou seja, desempenham diversas funções na infecção das células bacterianas a células eucariontes.

As inferências funcionais para a proteína hipotética codificada pelo gene *bjs08317* são mais escassas devido ao fato desta proteína não ter nenhum domínio conservado de nenhuma família proteica e apresentar baixa identidade em comparação com proteínas já descritas como efetoras. No entanto, pode-se supor que essa proteína possa desempenhar funções como efetoras ou desempenham algum papel no processo de infecção do rizóbio a planta hospedeira e que ainda não tenha sido caracterizado.

EXTRAÇÃO DE RNA, DIGESTÃO COM DNase E VERIFICAÇÃO DA QUALIDADE, INTEGRIDADE E PUREZA

O protocolo de extração de RNA com trizol foi hábil na extração de amostras de RNA de alta integridade, o qual pode ser verificado pelo perfil eletroforético apresentado na Figura 5. Além da alta integridade, a qualidade e concentração do RNA total foram avaliados, sendo obtido altos valores de concentração e pureza, como observado na Tabela 3, demonstrando, assim, a eficiência do método de extração. Neste contexto, o perfil eletroforético esperado para amostras de RNA com boa integridade é aquele em que as bandas das subunidades ribossomais (16S e 23S) bacterianas surgem no gel, de uma maneira íntegra e uniforme (ARES, 2012).

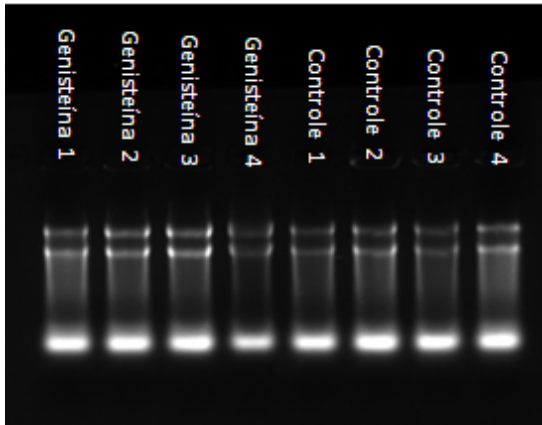


Figura 4. Integridade das amostras de RNA extraídas dos inóculos. A primeira banda corresponde ao 23S rRNA, a segunda banda ao 16S rRNA e o arraste presente entre as bandas e após elas correspondem aos RNA mensageiros de cada amostra. A última e mais forte banda corresponde ao excedente de tampão de amostra utilizado na corrida eletroforética.

Tabela 3 - Quantificação da concentração de RNA extraído via Nanodrop

Amostra	Concentração em ng/ μ L	260/280	260/230
Genisteína1	1189,83	2,07	2,33
Genisteína2	1224,64	2,07	2,26
Genisteína3	1316,51	2,09	2,26
Genisteína4	1106,84	2,08	2,14
Controle1	1041,14	2,06	2,19
Controle2	1140,53	2,06	2,25
Controle3	1066,81	2,07	2,21
Controle4	1095,05	2,05	2,22

As amostras (réplicas biológicas com genisteína 1^a, 2^a e 3^a caneleta e com metanol, 6^a, 7^a e 8^a caneleta) foram as selecionadas para a digestão com a enzima DNase a fim de eliminar resquícios de DNA genômico. Pode-se observar pelo perfil eletroforético que não houve a presença de DNA genômico nas amostras de RNA, mas a fim de eliminar possíveis quantidades mínimas que o gel de agarose não conseguiu detectar procedeu-se à etapa de digestão. Como demonstrado na Figura 6, mais uma vez a integridade foi mantida e não houve a presença de DNA genômico, assim como, as concentrações e relações boas de pureza foram mantidas, como demonstrado na Tabela 4.

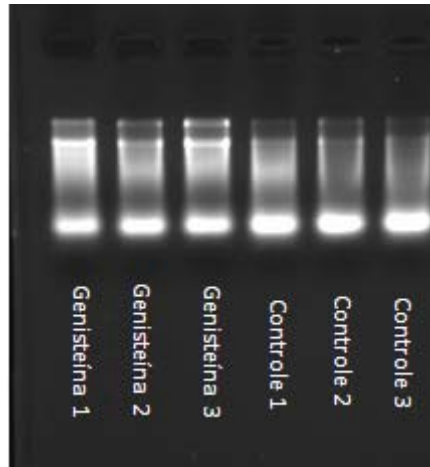


Figura 5 - Gel de agarose com as amostras de RNA tratadas com DNase

Estas etapas de verificação são importantes, pois refletem nas etapas posteriores de análise da expressão gênica. Diversos autores relatam que, para técnicas como RT-qPCR, a amostra de RNA deve estar íntegra e pura, a fim de quantificar com precisão os níveis de RNAm. Dentro destas exigências, o uso de RNA de baixa qualidade compromete os resultados derivados da expressão (BUSTIN, 2002; IMBEAUD et al, 2005). De fato, a diferença observada nos dados de expressão gênica entre RNAs intactos e amostras de RNAs degradadas levou muitos autores a propor que cada vez mais análises do controle de qualidade do RNA são imprescindíveis para a obtenção de resultados confiáveis e precisos por RT-qPCR (COSTA et al., 2004; CHOMCZYNSKI e SACCHI, 2006). Após a obtenção de amostras de RNA de alta qualidade e de boa concentração, procedeu-se a síntese das amostras de cDNA que foram utilizadas nas etapas posteriores de análise da expressão gênica.

Tabela 4 - Quantificação da concentração de RNA tratado com DNase via Nanodrop

Amostra	Concentração em ng/μL	260/280	260/230
Genisteína 1	291,22	2,1	2,24
Genisteína 2	289,07	2,13	2,2
Genisteína 3	269,42	2,11	2,01
Controle 1	284,62	2,1	2,06
Controle 2	275,59	2,1	2,11
Controle 3	281,6	2,09	2,28

EXPRESSÃO RELATIVA DOS GENES HIPOTÉTICOS DA ILHA SIMBIÓTICA

Todos os genes hipotéticos selecionados, quando induzidos com genisteína, resultaram em níveis de expressões significativos, conforme observado na Figura 3. As curvas de eficiência dos *primers* que foram utilizados no cálculo da expressão gênica encontram-se no Material Suplementar. Pode-se concluir que a indução com genisteína foi hábil na indução destes genes, sendo confirmada pela alta indução do controle positivo, o gene *nodC*, um conhecido gene de nodulação que é induzido por este flavonoide. De fato, em outros estudos de análise da expressão gênica com *B. japonicum*, o gene *nodC* e proteínas hipotéticas vêm sendo descritas como induzidas por genisteína. Bortolan et al. (2009) investigaram a expressão de genes em células de *B. japonicum* incubadas por 48h com genisteína (1 μ M) por RT-qPCR e obtiveram uma expressão do gene *nodC* 7,30 vezes maior na presença da genisteína em relação ao controle (metanol). Já Wei et al. (2008) observaram que vários genes dentro da ilha simbiótica de *B. japonicum* foram induzidos por genisteína (5 μ M) por RT-qPCR e obtiveram uma expressão de 25 vezes do gene *nodC* em relação ao controle em um período de indução de 12 horas. Lang et al. (2008) reportaram um nível de expressão do gene *nodC* induzido com genisteína (1 μ M) durante 8 horas via microarranjo de 35,1 vezes em relação ao metanol (controle) em *B. japonicum*.

No presente estudo, o nível de expressão do gene *nodC* foi 29,40 vezes superior ao do controle com metanol (Figura 7). Esse valor de expressão é próximo dos valores encontrados para o gene *nodC* nos trabalhos descrito acima, ressaltando-se que os valores podem ter sido discrepantes devido ao fato dos ensaios anteriores terem sido realizados ou com tempos de indução diferentes e/ou com concentrações diferentes de genisteína, sendo que no presente estudo, foi empregada a concentração de 2,5 μ M.

Os níveis de expressão dos genes codificando para proteínas hipotéticas variaram de 10,15 vezes em relação ao controle para o gene *bjs08254* até 32,22 vezes para o gene *bjs08267*. Isso reflete do fato de que esses genes codificam para diferentes proteínas hipotéticas com diferentes propriedades e funções, e assim, com níveis discrepantes de expressão.

Estudos de expressão gênica em *B. japonicum* estirpe USDA 110 observaram a expressão de genes codificando para proteínas hipotéticas induzidas por genisteína. Lang et al. (2008) obtiveram desde níveis de expressão de genes para proteínas

hipotéticas de 11,8 vezes para o gene *bsr1863* (codifica para uma proteína de função desconhecida), e de 2,3 vezes para o gene *blr4709* (codifica para uma proteína de função desconhecida). Ao analisarem o perfil de expressão de genes de *B. japonicum* USDA 110 induzidos por genisteína (5 μ M) a 30°C e 15°C via microarranjo Wei et al. (2010), observaram a expressão de diversos genes codificando para proteínas hipotéticas. Por exemplo, o gene *bsl2014* (codifica para uma proteína de função desconhecida) obteve um nível de expressão de 38 vezes induzido com genisteína em relação ao controle (metanol) num tempo de indução de 12 horas a 30°C. Já o gene *bl1810*, nas mesmas condições, mostrou um nível de expressão de 2,4 vezes em relação ao controle. Gomes et al. (2014) analisaram o nível de expressão de genes codificando para proteínas hipotéticas de *B. diazoefficiens* estirpe CPA 7 sob indução de genisteína (5 μ M) por 48 horas e obtiveram, por exemplo, níveis de expressão para o gene *bl10565* (codifica para uma proteína conservada de função desconhecida) de 1,82 vezes em relação ao controle (metanol) e de 1,68 vezes para o gene *blr2961* (codifica para uma proteína de função desconhecida). Demonstrando assim, como os níveis de expressão de diferentes genes para proteínas hipotéticas podem variar em função do próprio gene, espécie bacteriana, método e tempo de indução.



Figura 6 - Níveis de expressão de genes codificando proteínas hipotéticas de *B. japonicum* Estirpe CPAC 15 após crescimento até fase exponencial em presença de genisteína. Dados de expressão mostrados para cada gene representam a média das três réplicas biológicas. Dados foram normalizados em relação ao controle endógeno (16S rRNA). Todos os genes obtiveram uma expressão estaticamente significativa ao nível de 5% determinada pelo software *REST2009*.

CONSERVAÇÃO DAS PROTEÍNAS HIPOTÉTICAS EM OUTROS GENOMAS DE ESTIRPES DE *Bradyrhizobium*

A partir da busca por similaridade utilizando a ferramenta BLASTp no banco de dados NCBI foi possível compilar o nível de distribuição de proteínas com similaridade e identidade de, no mínimo, 60% entre as proteínas hipotéticas e as proteínas encontrados no banco de dados de outras espécies do gênero *Bradyrhizobium* (Figura 8). Esse nível de identidade, principalmente, foi empregado de acordo com o utilizado por Tian e Skolnick et al. (2003) que sugere que a partir de 40% de identidade já é possível apontar a função entre uma proteína conhecida e outra hipotética, mas que a transferência de função por homologia com mais acurácia só deve ser feita com um grau de identidade, no mínimo, de 60%. Resultados obtidos por Espadaler et al. (2009) também reforçam que esse nível de 60% já é hábil para identificar proteínas com a mesma função. Tendo em vista, que essas proteínas compartilham alto grau de conservação entre estirpes do mesmo gênero e espécie, pode-se supor que estas possam ser a mesma proteína, contudo, estudos mais robustos devem ser feitos para confirmar com maior confiabilidade esta afirmação.

O panorama apresentado pelas Figura 8 é útil para avaliar o nível de conservação que estas proteínas apresentam ao longo das diferentes espécies do gênero *Bradyrhizobium* e pode fornecer indícios da essencialidade da função dessas proteínas para as essas bactérias, sendo úteis para tentar correlacionar os níveis de expressão com a distribuição destas proteínas nos genomas. Ao observar a Figuras 8 e ao correlacionar com os níveis de expressão que os genes codificando essas proteínas hipotéticas obtiveram, pode-se notar que os genes com menores níveis de expressão, *bjs08254* (10,15 vezes mais expresso que o controle), *bjs08251* (15,58 vezes mais expresso que o controle), *bjs08216* (21,51 vezes mais expresso que o controle) foram os mesmos que apresentaram os menores níveis de distribuição das suas proteínas correspondentes, Figura 8. As proteínas hipotéticas dos genes *bjs08254*, *bjs08251* e *bjs08216* estavam presentes em apenas 7 diferentes espécies do gênero *Bradyrhizobium*. Mais apontamentos sobre a essencialidade, função e níveis de expressão destes genes na capacidade saprofítica e na FBN da estirpe CPAC 15 serão discutidas em tópicos posteriores.

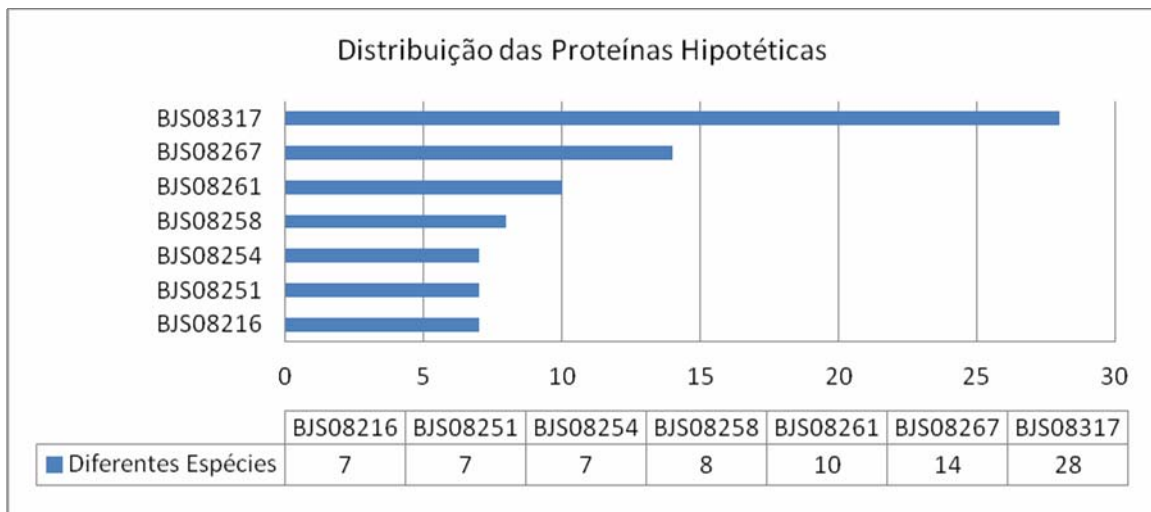


Figura 7 - Distribuição das proteínas hipotéticas em diferentes espécies do gênero *Bradyrhizobium*. Os dados separados por gene e utilizados para a criação dos gráficos encontram-se no Material Suplementar.

É importante ressaltar que a proteína hipotética codificada pelo gene *bjs08317* apresentou a maior distribuição nos genomas de *Bradyrhizobium*, o que corrobora com o fato do gene que codifica esta proteína ser o único no qual a sua sequência nucleotídica estava presente em outros genomas de *Bradyrhizobium*. Isto é, quando a sequência nucleotídica do gene *bjs08317* foi confrontada com o banco de dados de nucleotídeos do NCBI via Blastn, a sequência do gene foi encontrada em outros genomas. Em contrapartida, todos os outros 6 genes codificando para proteínas hipotéticas avaliados não possuem similaridade com genes nos genomas de *Bradyrhizobium*, ou seja, quando a sequências destes genes foram confrontadas com o banco de dados de nucleotídeos do NCBI via Blastn, a única sequência que retornou foi a própria sequência do gene, sendo portando, a sequência destes genes exclusivas a estirpe CPAC 15. Logicamente, então, era previsível que a distribuição da proteína codificada pelo gene *bjs08317* fosse maior do que das outras proteínas.

DISCUSSÃO

Através da anotação funcional da proteína hipotética codificada pelo gene *bjs08267*, foi inferido que possivelmente, a proteína hipotética em questão é uma proteína de

membrana pertencente à Superfamília dos Facilitadores Principais – *Major Facilitator Superfamily* (MFS). Entre as diferentes famílias de transportadores, apenas duas ocorrem onipresentes em todas as classificações de organismos, a família de transportadores ABC (ABC) e a superfamília de facilitadores Principais (MFS). Os transportadores MFS são carregadores secundários, isto é, o gradiente eletroquímico transmembranar de prótons (gerado pela H^+ -ATPase) constitui a força motora do transporte ativo dos solutos da membrana plasmática. Estes transportadores são capazes de transportar pequenos solutos, tais como açúcares, aminoácidos, drogas, entre outros, em resposta a gradientes iônicos (KUMAR e SCHWEIZER, 2005).

Em um estudo recente desenvolvido por Araújo e colaboradores (2015) foi demonstrado que genes relacionados aos transportadores de membrana e bombas de efluxo são induzíveis por flavonoides. Esses autores, ao estudarem o perfil transcricional da bactéria *Methylobacterium mesophilicum* (bactéria promotora do crescimento vegetal que coloniza raízes de soja) em presença dos exsudatos de raízes de soja, observaram a expressão de muitos genes envolvidos no transporte de membrana e, entre eles, genes relacionados aos sistemas de efluxo de resistência a múltiplas drogas (MDR).

Funções relacionadas à FBN para transportadores MFS vêm sendo descritas na literatura, demonstrando que genes codificando para estes transportadores são induzidos por flavonoides e participam de várias atividades biológicas nos rizóbios. González-Pasayo e Martínez-Romero (2000) identificaram os genes *rmrA* e *rmrB* em *Rhizobium etli* que são induzidos pelos flavonoides liberados pelas raízes do feijão (*Phaseolus vulgaris*). Os produtos preditos destes genes compartilham significativa homologia com proteínas de fusão de membrana (*rmrA*) e proteínas de facilitadores principais - MFS (*rmrB*). Mutantes de *rmrA* formaram, em média, 40% menos nódulos em feijão, enquanto os mutantes para o gene *rmrB* apresentaram um pequeno decréscimo na nodulação. Observou-se que não houve crescimento das estirpes mutantes na presença de 735 μ M de naringenina, 1,2 mM de ácido cumárico ou 725 μ M de ácido salicílico. Em contrapartida, a estirpe de tipo selvagem cresceu mesmo em elevadas concentrações destes compostos (naringenina 1,1 mM; 1,8 mM de ácido cumárico ou 3,6 mM de ácido salicílico). Desta forma, os autores propõem que estes genes codificam para um sistema de exportação induzível que evita a acumulação de compostos tóxicos de plantas dentro da célula bacteriana, conferindo, assim, uma vantagem para *R. etli* nodular feijão. Em contrapartida, Rossbach et al. (2014) observaram que a expressão do gene *emrA* (que é um ortólogo do gene *rmrA* de *R. etli*) é induzida por flavonoides, incluindo luteolina e

apigenina. Esse gene, juntamente com o *emrB* devem codificar um sistema de efluxo da superfamília de facilitadores principais. Mutações em *emrA* não tiveram efeito sobre a simbiose de *S. meliloti* em *Medicago sativa*.

Eda et al. (2011) identificaram todas as bombas de sistema de efluxo no genoma de *S. meliloti*, compostas de uma família dos transportadores ABC, três da superfamília (MFS) e 10 das famílias de resistência, nodulação e divisão celular (RND), assim como outras três proteínas de membrana, incluindo a proteína TolC. Apenas o mutante no gene *TolC* exibiu um defeito simbiótico significativo, pois observou-se que houve redução no número de nódulos fixando nitrogênio.

Wei et al. (2008) observaram, através de estudos de microarranjo, a indução de um *locus* contendo cinco genes que codificam para proteínas de resistência a multidrogas e que são induzidos por extratos de sementes de soja e por genisteína. Estes genes apresentaram semelhança com o gene *blr1629*, que também foi identificado no genoma da estirpe CPAC 15, codificando um transportador RDN da bomba de efluxo multidroga, que, inclusive, está localizado dentro da ilha simbiótica. Interessante, outros genes identificados como podendo atuar como sistemas de efluxo e que são induzidos por genisteína e extrato de sementes de soja não foram detectados no genoma da CPAC 15. O gene *bll4320* que codifica para uma provável proteína de efluxo do tipo RDN foi fortemente induzido por SSE e genisteína, sendo os níveis de expressão verificados por microarranjo e, posteriormente, validados por RT-qPCR (WEI et al., 2008). Os níveis de expressão alcançados por esse gene por RT-qPCR, em presença de genisteína, por 6 horas e 12 horas, foram de aproximadamente 28 vezes e 22 vezes, respectivamente. Por fim, os autores propõem que estes genes de resistência a multidrogas e que são induzidos por flavonoides, podem desempenhar um importante papel na competitividade de *B. japonicum* na rizosfera de soja.

Genes para sistemas de bombas de efluxo e outras classes de transportadores foram induzidos durante estresse oxidativo em *B. japonicum*, como observado por Donati et al., (2011) e podem desempenhar um papel importante na tolerância a este tipo de estresse. As plantas, quando infectadas por bactérias, mesmo rizóbios, liberam uma *burst* oxidativo sendo importante, então, que as bactérias tenham mecanismos eficientes para tolerar esse tipo de estresse e consigam colonizar eficientemente as células vegetais.

De modo geral, transportadores facilitadores principais (MFS), juntamente com transportadores do tipo ABC são as duas classes de transportadores amplamente distribuídos em procariotos, sendo já relatada na literatura, que principalmente,

transportadores do tipo ABC possam ter papéis fundamentais na tolerância a diferentes estresses e em características superiores de nodulação e competitividade em determinadas estirpes de *Bradyrhizobium* (SIQUEIRA et al., 2014; LIU et al., 2015). No presente estudo observou-se um alto nível de indução do gene *bjs08367* em presença de genisteína (cerca de 33 vezes em relação ao controle), o que corrobora com as evidências citadas anteriormente (GONZÁLEZ-PASAYO e MARTÍNEZ-ROMERO et al., 2000; ROSSBACH et al., 2014; ARAÚJO et al., 2015) de que genes para transportadores do tipo MFS podem ser induzidos por flavonoides. Estes transportadores, principalmente os que compõem os sistemas de efluxo de resistência a multidrogas, parecem ser fundamentais, tanto para a tolerância a diversos compostos liberados pelas plantas, quanto para a tolerância a antibióticos e outros compostos presentes na rizosfera. Por consequente, podem desempenhar papéis fundamentais nas características de competitividade e atividade saprofítica de estirpes de *B. japonicum* na rizosfera e em simbiose com a soja.

O gene *bjs08261* codifica para uma possível succinato-semialdeído desidrogenase, pertencente à família dos aldeídos desidrogenases, que catalisam a oxidação de uma grande variedade de aldeídos alifáticos e aromáticos usando NADP^+ como co-factor (SKINNER e COOPER, 1982). Exemplos desta enzima e de outras enzimas que possam fazer parte da mesma via são demonstrados no decorrer do texto e demonstram como essas enzimas são correlacionadas e abundantes nos genomas bacterianos. Essa redundância pode ocorrer devido a cada enzima participar de uma via bioquímica diferente e/ou como mecanismos de plasticidade destes micro-organismos, onde mutações em um dos genes não afetariam a atividade biológica do organismo, pois outros genes no genoma seriam ativados para suprir o gene mutado.

O gene succinato semialdeído desidrogenase (*gabD*) de *Escherichia coli* cataliza a oxidação dependente de NADP^+ de succinato semialdeído a succinato, sendo tal enzima envolvida na via de degradação do 4-aminobutanoato, que faz parte das vias de degradação de aminoácidos. O gene *gabD* é induzido por RpoS (um fator sigma especializado) em resposta a múltiplas condições de estresse, como mudanças no pH, limitação de nitrogênio, alta osmolaridade, sendo também induzido por gama-aminobutirato (GABA) (DONNELLY e COOPER, 1981; METZNER et al., 2004; LANGENDORF et al., 2010). GABA atua como um sinalizador intercelular em eucariotos e como um sinal interespecie em interações de micro-organismos com hospedeiros, Planamente et al. (2012) identificaram uma proteína seletiva de ligação ao GABA em *Agrobacterium tumefaciens*. Os

mutantes para essa proteína aumentaram a agressividade da doença em duas plantas hospedeiras, o tomate (*Solanum lycopersicum*) e o tabaco (*Nicotiana tabacum*). Além disso, por análises genômicas, os autores ainda encontraram ortólogos para essa proteína nos genomas de proteobacterias diazotróficas, como *Rhizobium*, *Azospirillum*, *Burkholderiae* e *Pseudomonas*.

Tanto a enzima succinato semialdeído desidrogenase quanto a alfa-cetoglutarato desidrogenase participam da via do ácido cítrico, e mutantes nestes genes têm demonstrado diferentes fenótipos. Por exemplo, Green e Emerich, (1997) observaram, em mutantes de *B. japonicum* para o gene da alfa-cetoglutarato desidrogenase, que um ciclo de ácido cítrico clássico não era necessário para fornecer energia ou elétrons para a redução efetiva do nitrogênio atmosférico, sendo a atividade de fixação simbiótica de nitrogênio deste mutante normal. No entanto, o mutante demonstrou um fenótipo de colonização de nódulo incomum, sugerindo que um ciclo do ácido cítrico clássico era necessário para a infecção e colonização adequada.

Em um estudo posterior, Green et al. (2000) encontraram, em *B. japonicum*, uma alfa-cetoglutarato descarboxilase independente de CoA (acetilcoenzima) que é capaz de formar succinato semialdeído diretamente de alfa-cetoglutarato. A atividade requer tiamina pirofosfato, mas não CoA ou nucleótidos de pirimidina e, desta forma, é distinta da alfa-cetoglutarato desidrogenase. Os resultados deste trabalho sugerem que a estirpe selvagem de *B. japonicum* pode descarboxilar alfa-cetoglutarato via duas enzimas diferentes, sendo uma delas, a já conhecida alfa-cetoglutarato desidrogenase e outra sendo uma alfa-cetoglutarato desidrogenase com atividade independente de NAD^+ e CoA. Os autores propõem que a alfa-cetoglutarato desidrogenase independente de NAD^+ e CoA e a succinato semialdeído desidrogenase (também observada no estudo a existência de uma enzima semialdeído desidrogenase alternativa) representam um desvio funcional do ciclo do ácido cítrico (TCA), sendo que estas duas vias para o catabolismo do alfa-cetoglutarato podem permitir uma maior flexibilidade do TCA em resposta a mudanças das demandas metabólicas. Os autores ainda demonstram que tanto a atividade da alfa-cetoglutarato desidrogenase independente de CoA, quanto da succinato semialdeído desidrogenase estão presentes em *Mesorhizobium loti* e *Rhizobium leguminosarum*, demonstrando que este desvio funcional pode ser amplamente distribuído entre os rizóbios. Os autores ainda exemplificam que, em estudo desenvolvido por Fitzmaurice e O’Gara (1993), a atividade da succinato semialdeído desidrogenase foi necessária para uma fixação biológica ideal em *Sinorhizobium meliloti*. Há relatos de genes

que codificam para succinato semialdeído desidrogenase no plasmídeo simbiótico de *Sinorhizobium fredii* NGR234 e na ilha simbiótica de *M. loti*, podendo, assim, desempenhar funções na simbiose.

Ao longo do genoma da estirpe de *B. japonicum* CPAC 15, outros genes codificando para succinato semialdeído desidrogenase foram encontrados (*BS02395*, *BS05026*, *BS05346*). Prell et al. (2009), ao estudarem o metabolismo de GABA em *Rhizobium leguminosarum* bv. *viciae* 3841, demonstraram que o semialdeído que é liberado do GABA por transaminação é oxidado pela enzima succinato semialdeído desidrogenase (GabD). Observou-se, que 5, das 6 potenciais GabD de *R. leguminosarum* (GabD1, D2, D3, D4 e D5) presentes no genoma desta estirpe mostraram ter a atividade de oxidar o semialdeído. No entanto, apenas mutantes em *GabD1*, *GabD2* e *GabD4* foram necessários para impedir a utilização de GABA como única fonte de nitrogênio na cultura. Posteriormente, os níveis de expressão destes genes foram avaliados por RT-qPCR e demonstrou-se que o gene *gabD2* foi altamente induzido em bacteróides em relação aos outros genes *gab*. Contudo, os três mutantes *Gab* continuam fixando nitrogênio em nódulos de ervilha. Os autores supõem que, apesar do catabolismo de GABA não ser essencial para a fixação de nitrogênio em bacteróides, os genes possam desempenhar funções na geração de energia e em desvios funcionais de descarboxilação no TCA. Em uma recente estudo genômico e proteômico na estirpe CPAC 7 de *B. diazoefficiens* (GOMES et al., 2014), uma proteína identificada como succinato semialdeído desidrogenase estava presente, identificada por homologia como a proteína *GabD1* (gene *bll3998*) da estirpe USDA 110 de *B. diazoefficiens*. Demonstrando que, talvez, dependendo da condição, há uma seleção para a indução de um dos genes codificantes para succinato semialdeído desidrogenase.

Pode-se supor então, que este gene possa desempenhar funções em vias de degradação de aminoácidos, e desta forma, confira características superiores à estirpe de CPAC 15, principalmente em relação à competitividade e atividade saprofítica.

O terceiro gene induzido por genisteína é o *bjs08258*, que possivelmente codifica para uma aspartato aminotransferase. Esta proteína pertence à família das aminotransferases classe I e II. As aminotransferases compartilham certas características mecânicas com outras enzimas dependentes de piridoxal-fosfato, tais como a ligação covalente do grupo piridoxal-fosfato em um resíduo de lisina. Para cada aminoácido existe uma aminotransferase correspondente (NAKAI et al., 1999). Essas enzimas são essenciais e estão envolvidas no metabolismo central de todos os organismos. O primeiro passo no

catabolismo é a remoção do grupo alfa-amino através de reações de transaminação, onde o grupo alfa-amino é transferido para o átomo de carbono alfa do alfa-cetoglutarato, produzindo o respectivo alfa-cetoácido análogo do aminoácido. O objetivo das reações de transaminação é coletar os grupos amino de muitos diferentes aminoácidos na forma de apenas um, o L-glutamato, que funciona como doador de grupos amino para as vias biossintéticas e excreção de compostos nitrogenados (NELSON e COX, 2002).

Rastogi e Watson (1991) isolaram um mutante de *S. meliloti* que não era capaz de crescer em meio contendo aspartato como fonte de nitrogênio e carbono, esse defeito era especificamente na utilização de aspartato, e foi correlacionado com a incapacidade em fixar nitrogênio dentro dos nódulos de alfafa. Através de análises de regiões genômicas, observou-se que o mutante não possuía a região codificadora para uma aspartato aminotransferase, outras análises foram realizadas, com outras aminotransferases e observou-se que o mutante contendo um cosmídeo codificando para uma aminotransferase aromática, conseguia utilizar aspartato parcialmente. Esses resultados demonstraram que a enzima aspartato aminotransferase é essencial para a fixação biológica de nitrogênio e que aspartato é um substrato importante nos nódulos.

Outro exemplo de vias nas quais o produto deste gene possa atuar foi demonstrado por Shao et al., (2015), que estudaram e validaram genes que possivelmente estão relacionados com a síntese de Ácido Indolacético (AIA) dependente de triptofano. Neste estudo, eles predizeram a função para o gene *patB*, que codifica para uma proteína hipotética conservada predita como uma aminotransferase da família do aspartato aminotransferase. Apesar do mutante para o gene *patB* não apresentar baixa produção de AIA, a superexpressão deste gene confirmou que ele estava envolvido na via de síntese de AIA, pois a superexpressão aumentou os níveis de produção do fitohormônio. Os autores ainda citaram um estudo desenvolvido por Kittel e colaboradores (1989), no qual foram construídos mutantes para duas aspartatos aminotransferases em *S. meliloti*, nos quais, nenhum dos genes foi essencial para a síntese de AIA na ausência de triptofano endógeno, mas ambos contribuíram para a síntese de AIA quando uma alta concentração de triptofano exógeno estava presente.

Em outro estudo foi constatado que os mutantes para o gene *attA* (aspartato aminotransferase A) em *R. leguminosarum* bv. *viciae* foram capazes de nodular (Nod⁺) ervilha, porém, os nódulos não eram hábeis em fixar nitrogênio (Fix⁻) (LODWIG et al., 2003),

sendo proposto que os mutantes foram metabolicamente interrompidos pela acumulação do glutamato, que não pode ser convertido em aspartato.

Como discutido na revisão de Dunn (2014) sobre os papéis-chave do metabolismo de aminoácidos microsimbiontes nas interações rizóbios-legumes, o aspartato é um precursor de muitos metabólitos, incluindo purinas e pirimidinas essenciais simbioticamente. Já a enzima aspartato aminotransferase é altamente ativa em bacteroides de vários rizóbios, no entanto, seu principal papel simbiótico é, provavelmente, no catabolismo de aspartato e não na síntese. Geralmente, rizóbios contêm duas cópias de genes codificando para aspartato aminotransferase, *aatA* e *aatB*. Uma função no catabolismo de aspartato foi demonstrada, uma vez que mutantes para o gene *aatA* de *S. meliloti* não crescem em meio de cultura com aspartato como única fonte de carbono. O papel desta enzima em bacteroides foi demonstrado pelo fato de mutantes *S. meliloti* JJ1c10 formarem nódulos que foram essencialmente Fix⁻ em alfafa (*Medicago sativa*), sendo a atividade residual da enzima AatA atribuída para uma aminotransferase aromática capaz de usar parcialmente aspartato como substrato e pela AatB, ressaltando-se que um mutante para o gene *aatB* em *S. meliloti* foi capaz de nodular e fixar nitrogênio. Consistente com a função simbiótica de *aatA*, análises proteômicas mostram que o gene *aatA* é encontrado em bacteroides de *S. meliloti*, enquanto *aatB* é apenas encontrado em bactérias em vitro (DJORDJEVIC, 2004). Em análises proteômicas da estirpe *B. japonicum* CPAC 15, a proteína codificada pelo *aatA* foi encontrada sendo induzida por genisteína (DA SILVA BATISTA e HUNGRIA, 2012). Em geral, esses resultados demonstram que aspartato é possivelmente usado como fonte de carbono durante a nodulação e pelos bacteróides.

O gene *bjs08258* que codifica para uma possível aspartato aminotransferase pode ser denominada como o gene *aatC*, tendo em vista, que no genoma da CPAC 15 já existe dois genes codificando para aspartato aminotransferase, os genes *aatA* e *aatB*. A indução deste terceiro gene é interessante do ponto de vista que ela encontra-se na ilha simbiótica e pode desempenhar um papel interessante para a estirpe de *B. japonicum*. Em um estudo desenvolvido por Gomes e colaboradores (2012), realizou-se a análise proteômica de *Rhizobium tropici* crescida sobre estresse térmico e encontrou-se a proteína codificada pelo gene *aatB*, o que leva a supor que este gene pode desempenhar papéis na tolerância a estresses e não somente desempenha papéis na simbiose.

Em especial, o gene *aatC* é induzido diferencialmente e expresso em altos níveis em presença de genisteína (29,18 vezes em relação ao controle) fortalecendo, assim,

que este pode desempenhar funções importantes para a estirpe de *B. japonicum*. Por fim, todos os resultados apresentados anteriormente sugerem que a presença de três genes para aspartato aminotransferase no genoma de *B. japonicum* CPAC 15 possivelmente confere uma grande plasticidade ecológica para esta estirpe, uma hipótese seria que as diferentes enzimas são induzidas em diferentes condições nutricionais, de estresse, tanto em bacteroides quanto em vida livre desempenhando, assim, funções tanto para a atividade saprofítica no solo, através de degradação de diferentes compostos para utilização como fonte de carbono e nitrogênio quanto para a fixação biológica de nitrogênio em si, tendo em vista que esta enzima é essencial para a funcionalidade simbiótica nos bacteróides.

O gene *bjs08251* codifica para uma proteína que possui 40% de identidade com serinas proteases, além de possuir o domínio para uma tripsina peptidase, podendo então, desempenhar uma função de tripsina. Esta enzima é específica para ligações peptídicas, contendo resíduos com cadeia lateral com carga elétrica positiva, em condições fisiológicas, tais como os aminoácidos arginina e lisina (READ e JAMES, 1988).

Em estudos proteômicos, Mollahosseini et al. (2015) observaram as proteínas que estavam singularmente presentes na germinação do cisto tanto de *Phytophthora pisi* quanto de *Phytophthora sojae* (fitopatógenos de ervilha e soja, respectivamente) e que não possuíam ortólogos interespecie. Com isto, os autores consideraram estas proteínas como candidatas para possíveis fatores patológicos espécie específico envolvidos na especificidade de infecção ao hospedeiro. Dentre as proteínas selecionadas em *P. pisi*, estavam serinas proteases, que poderiam estar envolvidas na degradação de proteínas do hospedeiro durante a infecção. Outras proteínas identificadas foram transportadores do tipo ABC e MFS, os quais já vêm sendo associados com a desintoxicação, resistência a multidrogas e transporte de metabólicos anteriormente. Essas proteínas poderiam conferir proteção a *P. pisi* contra compostos tóxicos liberados pelo hospedeiro vegetal. Em relação às proteínas específicas em cistos germinados de *P. sojae* também foi identificada uma serina protease.

Outras funções têm sido atribuídas às serinas proteases, como a tolerância a diferentes estresses. Por exemplo, genes bacterianos de bactérias extremófilas vêm sendo relacionados à tolerância a estresse salino. Dentre esses genes, destacam-se os que codificam para diferentes proteases que possuem condições ótimas de atividade e estabilidade em uma ampla faixa de salinidade. Dentre estas, as serinas proteases têm sido identificadas, como as isoladas de espécies de *Bacillus* (DAS et al., 2015). Em *B. japonicum* uma serina protease codificada pelo gene *degP* (= *htrA*) é ativada após choque térmico e tem como função ajudar

no *foldings* de proteínas desnaturadas que acumulam-se durante o estresse, porém, mutantes neste gene não afetaram a simbiose (NARBERHAUS et al., 1998). Em contra partida, este gene também responde diferencialmente em outras espécies de bactérias, como em *E. coli*, na qual ele é necessário para o crescimento em altas temperaturas (LIPINSKA et al., 1988), assim como observado em mutantes termosensíveis a altas temperaturas de *Acetobacter tropicalis*, no qual a mutação levou a um decréscimo no crescimento em meio sólido e à total inibição de crescimento em meio líquido (SOEMPHOL et al., 2011). Já LI et al. (1996) observaram que uma mutação no gene *degP* em *Yersinia enterocolitica*, que também codifica para uma tripsina serina protease, levou a um acréscimo na sensibilidade ao estresse oxidativo (H_2O_2 , O_2^-) e térmico ($50^\circ C$), além do mutante ter a sua virulência atenuada. *R. tropici* CIAT 899 e *Rhizobium freirei* PRF 81 podem crescer em temperaturas superiores a $40^\circ C$ e ao longo dos genomas destas estirpes, quatro homólogos para o gene *htrA* foram encontrados (ORMEÑO-ORRILLO et al., 2012). Entre as proteínas secretadas apenas na fase exponencial de *R. etli* e não na fase estacionária, foi identificada uma tripsina serina protease (DegPch2) (MENESES et al., 2010). Com isto, os autores levantam mais uma vez a suposição de que essas proteínas, além de desempenharem um papel na tolerância a estresses, desempenham funções na degradação de compostos tóxicos liberados pela planta.

No perfil proteômico de *R. freirei* PRF 81 sob choque térmico Gomes et al. (2012) observaram a expressão de diferentes proteínas relacionadas à tolerância ao estresse térmico e até ao estresse oxidativo, contudo, nenhuma das proteínas referentes aos quatro homólogos de *htrA* do genoma de *R. tropici* foram identificadas, o que levanta a questão de que a mesma pode desempenhar outros papéis nesta estirpe que não a tolerância ao estresse térmico, ou mais condizente, devido ao fato do estudo ter sido realizado por eletroforese bidimensional, que apesar de eficiente, não consegue capturar todas as proteínas secretadas.

Em outro estudo recente realizado por López-Leal et al., (2014), os autores identificaram genes diferencialmente expressos sob choque térmico e salino através da técnica de RNA-seq em *R. Etli*. Neste estudo, os autores evidenciaram que genes codificando para chaperonas foram exclusivamente *up*-regulados durante o choque térmico, enquanto dois genes (*degPch2* e *degPch1*) codificando para tripsina-serina proteases foram induzidos apenas no estresse salino. Neste contexto, em mutantes de *S. meliloti* e *Brucella abortus* em apenas um gene parálogo do *htrA* levou a um pequeno efeito sobre o crescimento em altas temperaturas, o que pode ter ocorrido devido à redundância de cópias deste gene no genoma

destas bactérias ou porque esse gene em rizóbios não seja responsável pela tolerância ao estresse térmico (GLAZEBROOK et al., 1996; PHILLIP e ROOP, 2001).

O gene *bjs08254* possivelmente codifica para uma glutamil endopeptidase, da família das serinas-peptidase, assim como o gene *bjs08251* descrito anteriormente. Interessante, os dois genes tiveram níveis de expressão diferencial semelhantes (15,58 vezes para o *bjs08251* e 10,15 vezes para o *bjs08254*). Assim como uma mutação no gene *degP* (uma serina protease) levou a uma atenuação da virulência em mutantes de *Y. enterocolitica* (LI et al., 1996), mutações no gene que codifica para glutamil endopeptidase (*sspA*) em *Staphylococcus aureus* também resultaram em uma significativa atenuação da virulência (SHAW et al., 2004, NICKERSON et al., 2007).

Ao que tudo indica, serinas proteases desempenham várias funções nos micro-organismos, desde a degradação de compostos tóxicos liberados pela planta até a tolerância a diferentes estresses abióticos. Provavelmente estas enzimas podem desempenhar papéis fundamentais na atividade saprofítica do *B. japonicum* no solo e para o processo inicial de interação da bactéria com o hospedeiro. Levanta-se a hipótese de que a bactéria, ao perceber os compostos liberados pelas raízes da soja, em especial a genisteína, aciona a síntese destas serinas proteases, para que estas atuem no processo de infecção do rizóbio nas raízes através da degradação de compostos vegetais liberados pelos mecanismos de defesa da planta. Vale ressaltar que a indução destes genes para serinas proteases foi aproximadamente a metade dos genes *bjs08267*, *bjs08261* e *bjs08258*, o que poderia ser atribuído ao fato de que os três genes já descritos anteriormente possam desempenhar papéis mais críticos e essenciais tanto para a atividade saprofítica, quanto para a fixação biológica de nitrogênio em rizóbios, sendo mais fortemente induzidos pelo flavonoide.

O gene *bjs08216* apresentou similaridade com chaperonas Tir em diferentes bactérias. Essas chaperonas fazem parte da família CesT. Esta família consiste de um grande número de proteínas que se assemelham à proteína chaperona Tir em *E. coli*. Em muitas bactérias Gram-negativas, o potencial patológico e de infecção é realizado pelo sistema de secreção do tipo III (T3SS), sendo que as chaperonas Tir fazem parte deste sistema, auxiliando a manter os substratos secretados em um estado-secreção competente (DELAHAY et al., 2002). Esta família ainda contém chaperona SicP e muitas DspF de sequências de muitas bactérias fitopatógenas. No geral, estas chaperonas desempenham funções no processo de virulência e infecção de bactérias em diferentes hospedeiros (BOGDANOVE et al., 1998). Estudos demonstram que o T3SS desempenha funções na simbiose, como em *Bradyrhizobium*

elkanii, onde foi constatado que este sistema ativa a sinalização de nodulação no hospedeiro na ausência dos fatores Nod (OKAZAKI et al., 2013). Em *B. japonicum* mutações no *cluster* para o *T3SS* também afetaram a capacidade de nodulação dependendo do hospedeiro (KRAUSE et al., 2002). Por fim, a presença de um gene codificando para uma possível chaperona Tir ainda não caracterizada dentro da ilha simbiótica leva à suposição que a estirpe CPAC 15 possui mecanismos diferenciais para a sua competitividade de infecção da planta.

O gene *bjs08317* encontra-se um pouco mais distante dos genes descritos anteriormente (*bjs08267*, *bjs08261*, *bjs08258*, *bjs08251*, *bjs08254*, *bjs08216*) e diferentemente desses genes, e como já mencionado anteriormente, o gene *bjs08317* possui sequência nucleotídica encontrada em outras estirpes de *Bradyrhizobium*, isso é, a sequência do gene não é exclusiva da estirpe CPAC 15. Ao realizar a análise de BLASTn (Material Suplementar), as sequências que apresentam 100% de identidade e 100% de similaridade são de *B. japonicum* estirpe E109 (umas das rizobactérias mais utilizadas em inoculantes para soja na Argentina desde 1970) (TORRES et al., 2015). Em *B. japonicum* estirpe USDA 6, em *B. diazoefficiens* USDA 110 e sendo parte do *cluster* do sistema de secreção do tipo III em *B. japonicum* USDA 122 (TSUKUI et al., 2013). E, por último, esse gene também é encontrado no genoma de *B. diazoefficiens* estirpe NK6 (estirpe contendo várias regiões de inserções em relação à USDA 110 e que possui quatro pequenos plasmídeos, além de um grande cromossomo) (IIDA et al., 2015).

O gene *bjs08317* pode codificar para uma possível proteína efetora do sistema de secreção do tipo III (TT3S). A proteína codificada por este gene apresentou 29% de identidade e 51% de similaridade com a proteína efetora XopP de *Xanthomonas fragariae* LMG 25863, 31% de identidade e 47% de similaridade com a proteína efetora XopP em *Xanthomonas campestris* e 34% de identidade e 50% de similaridade com a proteína efetora HIK1 de *Ralstonia solanacearum*. (Blastp – NCBI) (Material Suplementar).

O gênero *Xanthomonas* é um gênero de *Gammaproteobacteria* que inclui inúmeras espécies de fitopatógenos, Como um todo, os membros desse gênero são capazes de infectar uma ampla gama de plantas, distribuídas entre 124 espécies de monocotiledôneas e 268 espécies de dicotiledôneas. *X. fragariae* é agente causador da mancha angular em morangos. Já *X. campestris* é o principal fitopatógeno em crucíferas e causam a podridão negra das raízes (WILLIAMS, 1980). A pústula bacteriana em soja é causada pela *X. campestris* bv. *glycines* (RUKAYADI et al., 2000). Já em relação ao fitopatógeno *Ralstonia solanacearum*, este se caracteriza por ser um importante fitopatógeno causador da murcha

bacteriana, possuindo características únicas e significantes, como a sobrevivência a solos com altas temperaturas, degradação de diferentes compostos orgânicos, assim como a ampla faixa de hospedeiros, cerca de 200 espécies pertencentes a mais de 50 famílias botânicas. Sabe-se que o T3SS é importante na habilidade saprofítica, interação simbiótica e formação de biofilme, entre outras características importantes aos rizóbios e outras bactérias. Em específico, ao que tudo indica, a estirpe de *B. japonicum* CPAC 15 também possui genes efetores do T3SS de fitopatógenos que têm como característica a infecção em diversas espécies vegetais.

O gene *bjs08317* apresentou um alto nível de expressão (26,72 vezes em relação ao controle), sendo amplamente distribuído em diferentes estirpes e espécies de *Bradyrhizobium*, e ainda tendo similaridade com proteínas efetoras de bactérias que possuem uma ampla gama de hospedeiros. Todas essas dados dão robustez à hipótese de que o gene *bjs08317* pode desempenhar um importante papel no processo de infecção de *Bradyrhizobium* a soja e a outros hospedeiros vegetais.

A grande maioria dos genomas de *Xanthomonas* sequenciados apresenta um core central de nove genes codificando para proteínas efetoras do T3SS, e entre estes, está presente o gene *xopP*. Mutações em genes codificando para efetores centrais geralmente levam a redução da virulência e viabilidade do patógeno (RYAN et al., 2011). Interessante, mesmo o gene *XopP* sendo conservado em grande parte dos genomas de *Xanthomonas*, mutações neste gene não afetaram a virulência de *Xanthomonas oryzae* pv. *oryzae* em arroz (SONG e YANG, 2010), nem na virulência de *X. campestris* pv. *vesicatoria* em pimenta (*Piper nigrum*) e tomate (RODEN et al., 2004). Provavelmente, esse gene desempenha funções ainda desconhecidas nestas bactérias, sendo que evidências de eventos de transferência deste gene entre estirpes de *Xanthomonas* vêm sendo supostas (MHEDBI-HAJRI et al., 2013). A suposição de que este gene é transferido entre estirpes de *Xanthomonas* e até entre diferentes espécies de bactérias ganha reforço ao analisar o contexto genômico do gene *bjs08317*, pois ao lado do mesmo, há um gene codificando para transposase, o gene *bs08315*. Outro gene localizado próximo do *bs08317* que recentemente teve a sua função elucidada foi o *bjs08325* (proteína Y4yS), mutações neste gene favorecem a simbiose em *M. loti* e afetam negativamente a expressão de proteínas do sistema de secreção do tipo III (MERCANTE et al., 2015). Outro gene localizado próximo, é o *bjs08322* que codifica para um homólogo da proteína VirK, que é encontrada em *Agrobacterium tumefaciens*, *Ralstonia solanacearum* e *Xanthomonas axonopodis* pv. *citrie* e desempenham papéis na virulência

destas estirpes. Levanta-se mais uma vez a hipótese de que esta região de genes foi obtida através de transferências de genes de bactérias fitopatógenas e constitui-se como a região do T3SS. Por fim, estes genes, ao que tudo indica, podem desempenham funções relacionados com o processo de virulência e infecção do rizóbio aos hospedeiros.

CONSIDERAÇÕES FINAIS

Em geral, todos os genes codificando para proteínas hipotéticas foram induzidos em presença de genisteína, obtiveram valores de expressões elevados e podem possuir funções importantes tanto para a sobrevivência da estirpe de *B. japonicum* no solo quanto para um processo eficiente de infecção, colonização e estabelecimento da simbiose em soja. Há uma tendência, a partir das discussões apresentadas de que a estirpe CPAC 15 de *B. japonicum* possui um genoma extremamente versátil, com a redundância de vários genes que podem possuir tanto a mesma função quanto funções distintas e, assim, permitindo o sucesso desta bactéria no solo e na simbiose, isso ganha força no fato de que de genes que estão localizados fora da ilha simbiótica e que desempenham papéis vitais de sobrevivência, possuem genes codificando os mesmos produtos e localizados dentro da ilha simbiótica, o que pode ser atribuído, talvez, há um mecanismo evolutivo para a sobrevivência desta bactéria e competitividade a adversas condições ambientais.

Por fim, a inferência funcional apresentada caracteriza-se como uma etapa inicial na caracterização e validação funcional de proteínas hipotéticas e não deve ser tomada como a única análise necessária. Conforme demonstrado ao longo da discussão, etapas como mutações devem ser empregadas a fim de observar como a bactéria se comporta frente à inativação dos genes em questão. Além disso, análises proteômicas mais refinadas, como clonagem em vetor de expressão da proteína em questão, identificação da estrutura terciária por métodos como cristalografia de raios x e sequenciamento da proteína, além de análises enzimáticas, são métodos que devem ser empregados a fim de obter uma compreensão mais específica da proteína e assim, validar a existência da mesma e principalmente, elucidar com precisão as funções moleculares destas.

REFERÊNCIAS BIBLIOGRÁFICAS

ALTSCHUL, S. F.; GISH, W.; MILLER, W.; MYERS, E. W.; LIPMAN, D. J. Basic Local Alignment Search Tool. **The Journal of Molecular Biology**, v. 215, 1990

ALVES, B. J. R.; BODDEY, R. M.; URQUIAGA, S. The success of BNF in soybean in Brazil. **Plant and Soil**, v.252, p.1–9, 2003.

ARAÚJO, F. F.; HUNGRIA, M. Nodulação e rendimento de soja co-infectada com *Bacillus subtilis* e *Bradyrhizobium japonicum/Bradyrhizobium elkanii*. **Pesquisa Agropecuária Brasileira**, Brasília, v.34, p.1633-1643, 1999.

ARAÚJO, W. L.; SANTOS, D. S.; DINI-ANDREOTE, F.; et al. Genes related to antioxidant metabolism are involved in *Methylobacterium mesophilicum*-soybean interaction. **Antonie van Leeuwenhoek**, v. 108, n. 4, p. 951–963, 2015.

ARES, M. Bacterial RNA isolation. **Cold Spring Harbor Protocols**, v. 9, 2012.

BAHAR, M.; MAJNIK, J. DE; WEXLER, M.; et al. A model for the catabolism of rhizopine in *Rhizobium leguminosarum* involves a ferredoxin oxygenase complex and the inositol degradative pathway. **Molecular Plant-Microbe Interactions: MPMI**, v. 11, n. 11, p. 1057–1068, 1998.

BATISTA, J. S. D. S.; TORRES, A. R.; HUNGRIA, M. Towards a two-dimensional proteomic reference map of *Bradyrhizobium japonicum* CPAC 15: Spotlighting “hypothetical proteins.” **Proteomics**, v. 10, n. 17, p. 3176–3189, 2010.

BOGDANOVA, A. J.; KIM, J. F.; WEI, Z.; et al. Homology and functional similarity of an hrp-linked pathogenicity locus, *dspEF*, of *Erwinia amylovora* and the avirulence locus *avrE* of *Pseudomonas syringae* pathovar tomato. **Proceedings of the National Academy of Sciences of the United States of America**, v. 95, n. 3, p. 1325–1330, 1998.

BUSTIN, S. A. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. **Journal of Molecular Endocrinology**, v. 29, p. 23–29, 2002.

CHAUDHRY, V.; BHATIA, A.; BHARTI, S. K.; et al. Metabolite profiling reveals abiotic stress tolerance in Tn5 mutant of *Pseudomonas putida*. **Plos One**, v. 10, n. 1, p. e0113487, 2015.

CHOMCZYNSKI, P., SACCHI, N. The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years. **Nature Protocols**, v. 1, n. 2, p. 581–585, 2006.

COSTA, R., GOMES, N.C.M., MILLING, A., SMALLA, K. An optimized protocol for simultaneous extraction of DNA and RNA from soils. **Brazilian Journal of Microbiology**, v. 35, n.3, p. 230–234, 2004

D. W. Mount. Using the Basic Local Alignment Search Tool (BLAST). *Cold Spring Harbor Protocols*, v. 7, 2007.

D'UTRA, G. Cultura do feijão chinês. Soja. **Jornal do Agricultor**, v.VII, p.185-188, 1882.
 DELAHAY, R. M.; SHAW, R. K.; ELLIOTT, S. J.; et al. Functional analysis of the enteropathogenic *Escherichia coli* type III secretion system chaperone CesT identifies domains that mediate substrate interactions. **Molecular Microbiology**, v. 43, n. 1, p. 61–73, 2002.

DJORDJEVIC, M. A. *Sinorhizobium meliloti* metabolism in the root nodule: A proteomic perspective. **Proteomics**, v. 4, n. 7, p. 1859–1872, 2004.

DÖBEREINER, J.; FRANCO, A.A.; GUZMÁN, I. Estirpes de *Rhizobium japonicum* de excepcional eficiência. **Pesquisa Agropecuária Brasileira**, Série Agronomia, v.5, p.155-161, 1970.

DONATI, A. J.; JEON, J.-M.; SANGURDEKAR, D.; SO, J.-S.; CHANG, W.-S. Genome-wide transcriptional and physiological responses of *Bradyrhizobium japonicum* to paraquat-mediated oxidative stress. **Applied and Environmental Microbiology**, v. 77, n. 11, p. 3633–3643, 2011

DONNELLY, M. I.; COOPER, R. A. Succinic semialdehyde dehydrogenases of *Escherichia coli* and their role in the degradation of p-hydroxyphenylacetate and γ -aminobutyrate. **Enzyme**, v. 561, n. 113, p. 555–561, 1981.

DUNN, M. F. Key roles of microsymbiont amino acid metabolism in rhizobia-legume interactions. **Critical Reviews in Microbiology**, v. 7828, p. 1–41, 2014.

EDA, S.; MITSUI, H.; MINAMISAWA, K. Involvement of the SmeAB multidrug efflux pump in resistance to plant antimicrobials and contribution to nodulation competitiveness in *Sinorhizobium meliloti*. **Applied and Environmental Microbiology**, v. 77, n. 9, p. 2855–2862, 2011.

EMBRAPA, 2010. **Histórico da soja no Brasil**. Disponível em:
 <http://www.cnpso.embrapa.br/index.php?op_page=113&cod_pai=35> Acesso em:
 20/10/2015

ESPADALER, J.; ESWAR, N.; QUEROL, E.; et al. Prediction of enzyme function by combining sequence similarity and protein interactions. **BMC Bioinformatics**, v. 9, n. 249, 2008.

FAGAN, E. B.; MEDEIROS, S.L. P.; MANFRON, P. A.; CASAROLIL, D.; SIMON, J.; NETO, D. D.; LIER, Q. J.; SANTOS, O. S.; MÜLLER, L. Fisiologia da fixação biológica do nitrogênio em soja – revisão. **Revista da Faculdade Zootecnia, Veterinária e Agronomia, Uruguaiana**, v.14, n.1, p. 89-106, 2007.

FERRAIOLI, S.; TATÈ, R.; CERMOLA, M.; et al. Auxotrophic mutant strains of *Rhizobium etli* reveal new nodule development phenotypes. **Molecular plant-microbe interactions: MPMI**, v. 15, n. 5, p. 501–510, 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 & development**, v. 2, n. 3, p. 282–93, 1988.
- FITZMAURICE, A.; M. O'GARA, F. A *Rhizobium meliloti* mutant, lacking a functional gamma-aminobutyrate (GABA) bypass, is defective in glutamate catabolism and symbiotic nitrogen fixation. **FEMS Microbiology Letters**, v.109, p. 195-202, 1993.
- FREIBERG, C.; FELLAY, R.; BAIROCH, A.; BROUGHTON, W. J.; ROSENTHAL, A.; PERRET, X. Molecular basis of symbiosis between *Rhizobium* and legumes. **Nature**, n. 387, n. 6631, p. 394-401, 1997.
- GARCIA, M.; DUNLAP, J.; LOH, J.; STACEY, G. Phenotypic characterization and regulation of the *nolA* gene of *Bradyrhizobium japonicum*. **Molecular Plant-Microbe Interact**, 1996.
- GARCIA-FRAILE, P.; SEAMAN, J. C.; KARUNAKARAN, R.; et al. Arabinose and protocatechuate catabolism genes are important for growth of *Rhizobium leguminosarum* biovar viciae in the pea rhizosphere. **Plant and Soil**, v. 390, n. 1-2, p. 251–264, 2015.
- GLAZEBROOK, J.; ICHIGE, A; WALKER, G. C. Genetic analysis of *Rhizobium meliloti* *bacA-phoA* fusion results in identification of *degP*: two loci required for symbiosis are closely linked to *degP*. **Journal of Bacteriology**, v. 178, n. 3, p. 745–52, 1996.
- GOMES, D. F.; BATISTA, J. S. D. S.; SCHIAVON, A. L.; ANDRADE, D. S.; HUNGRIA, M. Proteomic profiling of *Rhizobium tropici* PRF 81: identification of conserved and specific responses to heat stress. **BMC Microbiology**, v. 12, p. 84, 2012.
- GONZÁLEZ-PASAYO, R.; MARTÍNEZ-ROMERO, E. Multiresistance genes of *Rhizobium etli* CFN42. **Molecular Plant-Microbe Interactions: MPMI**, v. 13, n. 5, p. 572–7, 2000.
- GOURION, B.; SULSER, S.; FRUNZKE, J.; et al. The PhyR- σ EcfG signalling cascade is involved in stress response and symbiotic Efficiency in *Bradyrhizobium japonicum*. **Molecular Microbiology**, v. 73, n. 2, p. 291–305, 2009.
- GREEN, L. S.; EMERICH, D. W. *Bradyrhizobium japonicum* does not require alpha-ketoglutarate dehydrogenase for growth on succinate or malate. **Journal of Bacteriology**, v. 179, n. 1, p. 194–201, 1997.
- GREEN, L. S.; LI, Y.; EMERICH, D. W.; FRASER, J.; DAY, D. A. Catabolism of alpha-ketoglutarate by a *sucA* mutant of *Bradyrhizobium japonicum*: evidence for an alternative tricarboxylic acid cycle. **Journal of Bacteriology**, v. 182, n. 10, p. 2838–2844, 2000.
- GUO GAO, T.; YUAN XU, Y.; JIANG, F.; et al. Nodulation characterization and proteomic profiling of *Bradyrhizobium liaoningense* CCBAU05525 in response to water-soluble humic materials. **Scientific Reports**, v. 5, n. April, p. 10836, 2015.
- GUO, W.; CAI, L. L.; ZOU, H. S.; et al. Ketoglutarate transport protein KgtP is secreted through the Type III Secretion System and contributes to virulence in *Xanthomonas oryzae* pv. *oryzae*. **Applied and Environmental Microbiology**, v. 78, n. 16, p. 5672–5681, 2012.

HENIKOFF, S.; HAUGHN, G. W.; CALVO, J. M.; WALLACE, J. C.; A large family of bacterial activator proteins. **PNAS, Proceedings of the National Academy of Sciences**, v. 85, p. 6602–6606, 1988.

HONTELEZ, J. G.; LANKHORST, R. K.; KATINAKIS, P.; VAN DEN BOS, R. C.; VAN KAMMEN, A. Characterization and nucleotide sequence of a novel gene *fixW* upstream of the *fixABC* operon in *Rhizobium leguminosarum*. **Molecular Genetics and Genomics**, v. 218, n. 3, p. 536-44, 1989.

HUNGRIA, M., MENDES, I.C. Nitrogen fixation with soybean: the perfect symbiosis? In: DE BRUIJN, F., (Ed.). **Biological nitrogen fixation**. New Jersey, Hoboken: Wiley-Blackwell, Wiley Publisher, 2015.

HUNGRIA, M.; CAMPO, J. R.; MENDES, I. C. **Fixação biológica do nitrogênio na cultura da soja**. Londrina: EMBRAPA-CNPSo, 2001. 48p.

HUNGRIA, M.; CAMPO, R.J.; MENDES, I.C.; GRAHAM, P.H. Contribution of biological nitrogen fixation to the N nutrition of grain crops in the tropics: the success of soybean (*Glycine max* L. Merr.) in South America. In: SINGH, R.P.; SHANKAR, N.; JAIWAL, P.K., eds. **Nitrogen nutrition and sustainable plant productivity**. Houston, Texas: Studium Press, LLC, p.43-93, 2006a

HUNGRIA, M.; STACEY, G. Molecular signals exchanged between host plants and rhizobia: basic aspects and potential application in agriculture. **Soil Biology Biochemistry**, v. 29, n. 5/6, p. 819-830, 1997.

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

IIDA, T.; ITAKURA, M.; ANDA, M.; et al. Symbiosis island shuffling with abundant insertion sequences in the genomes of extra-slow-growing strains of soybean Bradyrhizobia. **Applied and Environmental Microbiology**, v. 81, n. 12, p. 4143–4154, 2015.

IMBEAUD, S.; GRAUDENS, E.; BOULANGER, V.; BARLET, X.; ZABORSKI, P.; EVENO, E.; MUELLER, O.; SCHROEDER, A.; AUFRAY, C.; Towards standardization of RNA quality assessment using user-independent classifiers of microcapillary electrophoresis traces. **Nucleic Acids Research**, v. 33, n. 6, 2005.

ITO, N.; ITAKURA, M.; EDA, S.; et al. Global gene expression in *Bradyrhizobium japonicum* cultured with vanillin, vanillate, 4-hydroxybenzoate and protocatechuate. **Microbes and Environments**, v. 21, n. 4, p. 240–250, 2006.

JARDIM FREIRE, J. R.; VERNETTI, F. J. A pesquisa com soja, a seleção de rizóbio e produção de inoculantes no Brasil. **Pesquisa Agropecuária Gaúcha**, v. 5, p. 117-126, 1999.
KIMBREL, J. A.; THOMAS, W. J.; JIANG, Y.; CREASON, A. L.; THIREAULT, C. A.; SACHS, J. L.; CHANG, J. H. Mutualistic co-evolution of type III effector genes in *Sinorhizobium fredii* and *Bradyrhizobium japonicum*. **PLOS Pathogens**, n. 9, n. 2, 2013.

- KITTELL, B. L.; HELINSKI, D. R.; DITTA, G. S. Aromatic aminotransferase activity and indoleacetic acid production in *Rhizobium meliloti*. **Journal of Bacteriology**, v. 171, n. 10, p. 5458–5466, 1989.
- KRAUSE, A.; DOERFEL, A.; GÖTTTFERT, M. Mutational and transcriptional analysis of the type III secretion system of *Bradyrhizobium japonicum*. *Molecular plant-microbe interactions* : MPMI, v. 15, n. 12, p. 1228–1235, 2002.
- KUMAR, A; SCHWEIZER, H. Bacterial resistance to antibiotics: Active efflux and reduced uptake. **Advanced Drug Delivery Reviews**, v. 57, n. 10, p. 1486–1513, 2005.
- LANGENDORF, C. G.; KEY, T. L. G.; FENALTI, G.; et al. The X-Ray crystal structure of *Escherichia coli* succinic semialdehyde dehydrogenase; structural insights into NADP+/Enzyme interactions. **PLoS ONE**, v. 5, n. 2, p. e9280, 2010
- LI, S. R.; DORRELL, N.; EVEREST, P. H.; DOUGAN, G.; WREN, B. W. Construction and characterization of a *Yersinia enterocolitica* O:8 high-temperature requirement (*htrA*) isogenic mutant. **Infection and immunity**, v. 64, n. 6, p. 2088–94, 1996.
- LIPINSKA, B.; KING, J.; ANG, D.; GEORGOPOULOS, C. Sequence analysis and transcriptional regulation of the *Escherichia coli* *grpE* gene, encoding a heat shock protein. **Nucleic Acids Research**, v. 16, n. 15, p. 7545–7562, 1988.
- LIU, Y.; GUAN, D.; JIANG, X.; et al. Proteins involved in nodulation competitiveness of two *Bradyrhizobium diazoefficiens* strains induced by soybean root exudates. **Biology and Fertility of Soils**, v. 51, n. 2, p. 251–260, 2015.
- LODWIG, E. M.; HOSIE, A. H. F.; BOURDÈS, A; et al. Amino-acid cycling drives nitrogen fixation in the legume-*Rhizobium* symbiosis. **Nature**, v. 422, n. 6933, p. 722–726, 2003.
- LÓPEZ-LEAL, G.; TABCHE, M.; CASTILLO-RAMÍREZ, S.; et al. RNA-Seq analysis of the multipartite genome of *Rhizobium etli* CE3 shows different replicon contributions under heat and saline shock. **BMC Genomics**, v. 15, n. 1, p. 770, 2014.
- MADDOCKS, S. E.; OYSTON, P. C. F. Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. **Microbiology**, v. 154, n. 12, p. 3609–3623, 2008.
- MENESES, N.; MENDOZA-HERNÁNDEZ, G.; ENCARNACIÓN, S. The extracellular proteome of *Rhizobium etli* CE3 in exponential and stationary growth phase. **Proteome science**, v. 8, p. 51, 2010.
- MERCANTE, F. M. Uso de inoculante garante economia de 3 bilhões de dólares na cultura soja no país. 2005. Disponível em: <http://www.embrapa.br/noticias/artigos/folder.20050202.1550581232/artigo.2005-12-05.0506770395/mostra_artigo>. Acesso em: 20/09/2015

MERCANTE, V.; DUARTE, C. M.; SÁNCHEZ, C. M.; et al. The absence of protein Y4yS affects negatively the abundance of T3SS *Mesorhizobium loti* secretin, RhcC2, in bacterial membranes. **Frontiers in Plant Science**, v. 6, n. January, p. 1–12, 2015.

METZNER, M.; GERMER, J.; HENGGE, R. Multiple stress signal integration in the regulation of the complex σ S-dependent *csiD-ygaF-gabDTP* operon in *Escherichia coli*. **Molecular Microbiology**, v. 51, n. 3, p. 799–811, 2004.

MHEDBI-HAJRI, N.; HAJRI, A.; BOUREAU, T.; et al. Evolutionary History of the Plant Pathogenic Bacterium *Xanthomonas axonopodis*. **PLoS ONE**, v. 8, n. 3, p. e58474, 2013.

MOLLAHOSSEIN, S. H. **Host-Pathogen Interactions in Root Infecting Oomycete Species**. 2015.

MORAES, B. W.; MARTINS, F. S.; GARCIA, G. O.; CAETANO, S. P.; BUCKER, M. W.; CARRARA, C. F. Avaliação da fixação biológica do nitrogênio em genótipos de feijoeiros tolerantes a seca. **Idesia**, v. 28, n. 1, p. 61-68, 2010.

MOREIRA, W. M. Q. **Metabolismo respiratório de Bradirrizóbios durante processos in vitro e simbiótico analisado por PCR quantitativo em tempo real**. 97f. Mestrado (Microbiologia Agrícola). Faculdade de Ciências Agrárias e Veterinárias, Câmpus de Jaboticabal, 2009.

MURPHY, P. J.; WEXLER, W.; GRZEMSKI, W.; RAO, J. P.; GORDON, D. Rhizopines—Their role in symbiosis and competition. **Soil Biology and Biochemistry**, v. 27, n. 4-5, p. 525–529, 1995.

NAKAI, T.; OKADA, K.; AKUTSU, S.; et al. Structure of *Thermus thermophilus* HB8 aspartate aminotransferase and its complex with maleate. **Biochemistry**, v. 38, n. 8, p. 2413–2424, 1999.

NAKANISHI, K.; NUREKI, O. Recent progress of structural biology of tRNA processing and modification. **Molecules and Cells**, v. 19, p. 157-166, 2005.

NARBERHAUS, F.; WEIGLHOFER, W.; FISCHER, H. M.; HENNECKE, H. Identification of the *Bradyrhizobium japonicum degP* gene as part of an operon containing small heat-shock protein genes. **Archives of Microbiology**, v. 169, n. 2, p. 89-97, 1998.

NELSON, D. L. & COX, M.M. 2002. **Lehninger: Princípios de Bioquímica**. 3ª edição. Editora Sarvier, São Paulo, SP, Brasil. ISBN: 85-7378-125-4.

NEVES, M.C.P.; DIDONET, A.D.; DUQUE, F.F.; DÖBEREINER, J. *Rhizobium* strain effects on nitrogen transport and distribution in soybeans. **Journal of Experimental Botany**, v. 36, p. 1179-1192, 1985.

NICKERSON, N. N.; PRASAD, L.; JACOB, L.; DELBAERE, L. T.; MCGAVIN, M. J. Activation of the *SspA* serine protease zymogen of *Staphylococcus aureus* proceeds through unique variations of a trypsinogen-like mechanism and is dependent on both autocatalytic and metalloprotease-specific processing. **Journal of Biological Chemistry**, v. 282, n. 47, p. 34129–34138, 2007.

OKAZAKI, S.; KANEKO, T.; SATO, S.; SAEKI, K. Hijacking of *leguminous* nodulation signaling by the rhizobial type III secretion system. **Proceedings of the National Academy of Sciences of the United States of America**, v. 110, n. 42, p. 17131–6, 2013.

OLIVEIRA, L. R. **Análise da Expressão dos genes *nodC* e *nodG* de *Rhizobium tropici* sob indução com flavonóides pela técnica de PCR quantitativo**. 93f. Dissertação (Mestrado em Microbiologia) – Departamento de Microbiologia, Universidade Estadual de Londrina, Londrina, 2009.

ORMEÑO-ORRILLO, E.; MENNA, P.; ALMEIDA, L. G. P.; et al. Genomic basis of broad host range and environmental adaptability of *Rhizobium tropici* CIAT 899 and *Rhizobium* sp. PRF 81 which are used in inoculants for common bean (*Phaseolus vulgaris* L.). **BMC Genomics**, v. 13, n. 1, p. 735, 2012.

PECK, M. C.; FISHER, R. F.; BLISS, R.; LONG, S. R. Isolation and characterization of mutant *Sinorhizobium meliloti* NodD1 Proteins with altered responses to luteolin. **Journal of Bacteriology**, v. 195, n. 16, p. 3714–3723, 2013.

PERES, J.R.R. **Seleção de estirpes de *Rhizobium japonicum* e competitividade por sítios de infecção nodular em cultivares de soja (*Glycine max* (L.) Merrill)**. 81f. Dissertação (Mestrado em Agronomia) – Departamento de Solos, Universidade Federal do Rio Grande do Sul, Porto Alegre. 1979.

PÉREZ-MONTAÑO, F.; JIMÉNEZ-GUERRERO, I.; CERRO, P. DEL; et al. The symbiotic biofilm of *Sinorhizobium fredii* SMH12, necessary for successful colonization and symbiosis of *Glycine max* cv osumi, is regulated by quorum sensing systems and inducing flavonoids via NodD1. **PLoS ONE**, v. 9, n. 8, p. e105901, 2014.

PFAFFL, M. W.; HORGAN, G. W.; DEMPFLER, L. Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. **Nucleic acids research**, v. 30, n. 9, p. e36, 2002.

PHILLIPS, R. W.; ROOP, R. M. The *Brucella abortus* Lon functions as a generalized stress response protease and is required for wild-type virulence in BALB/c mice. **Infection and immunity**, v. 69, n. 9, p. 5911–5913, 2001.

PLANAMENTE, S.; MONDY, S.; HOMMAIS, F.; et al. Structural basis for selective GABA binding in bacterial pathogens. **Molecular Microbiology**, v. 86, n. 5, p. 1085–1099, 2012.

PORRÚA O.; GARCÍA-JARAMILLO M.; SANTERO E.; GOVANTES F.; Complex interplay between the LysR-type regulator AtzR and its binding site mediates atzDEF activation in response to two distinct signals. **Molecular Microbiology**, v. 76, p. 331-347, 2010.

PRELL, J.; BOURDES, A.; KARUNAKARAN, R.; LOPEZ-GOMEZ, M.; POOLE, P. Pathway of γ -Aminobutyrate metabolism in *Rhizobium leguminosarum* 3841 and its role in symbiosis. **Journal of Bacteriology**, v. 191, n. 7, p. 2177–2186, 2009.

RASTOGI, V. K.; WATSON, R. J. Aspartate aminotransferase activity is required for aspartate catabolism and symbiotic nitrogen fixation in *Rhizobium meliloti*. **Journal of Bacteriology**, v. 173, n. 9, p. 2879–87, 1991.

READ, R. J.; JAMES, M. N. Refined crystal structure of *Streptomyces griseus* trypsin at 1.7 Å resolution. **Journal Molecular Biology**, v. 200, n. 3, p. 523–551, 1988.

REEN, F. J.; BARRET, M.; FARGIER, E.; O’MUIÑNEACHÁIN, M.; O’GARA, F. Molecular evolution of LysR-type transcriptional regulation in *Pseudomonas aeruginosa*. **Molecular Phylogenetics and Evolution**, v. 66, n. 3, p. 1041–9, 2013.

RODEN, J. A.; BELT, B.; ROSS, J. B.; et al. A genetic screen to isolate type III effectors translocated into pepper cells during *Xanthomonas* infection. **Proceedings of the National Academy of Sciences of the United States of America**, v. 101, p. 16624–16629, 2004.

ROSSBACH, S.; KUNZE, K.; ALBERT, S.; ZEHNER, S.; GÖTTFERT, M. The *Sinorhizobium meliloti* EmrAB efflux system is regulated by flavonoids through a TetR-like regulator (*EmrR*). **Molecular Plant-Microbe Interactions: MPMI**, v. 27, n. 4, p. 379–87, 2014.

RUKAYADI, Y.; SUWANTO, A; TIAHJONO, B.; HARLING, R. Survival and epiphytic fitness of a nonpathogenic mutant of *Xanthomonas campestris* sp. *glycines*. **Applied and Environmental Microbiology**, v. 66, n. 3, p. 1183–9, 2000.

RYAN, R. P.; VORHÖLTER, F.-J.; POTNIS, N.; et al. Pathogenomics of *Xanthomonas*: understanding bacterium–plant interactions. **Nature Reviews Microbiology**, v. 9, n. 5, p. 344–355, 2011.

SANTANA, B. G. DE. Diversidade de isolados brasileiros de *Rasltonia solanacearum* da biovar 2. 110f. Dissertação (Mestrado em Ciências Genômicas e Biotecnologia). Universidade Católica de Brasília, Brasília, 2009.

SCHELL, M. A. Molecular biology of the LysR family of transcriptional regulators. **Annual Reviews**, v. 47, p. 597–626, 1993.

SCOTTI, M. R. M. M. L.; SÁ, N. M. H.; VARGAS, M. A. T.; DÖBEREINER, J. Streptomycin resistance of *Rhizobium* isolates from brazilian cerrados. **Anais da Academia Brasileira de Ciências**, Rio de Janeiro, v.54, p.733-738, 1982.

SHAO, J.; LI, S.; ZHANG, N.; et al. Analysis and cloning of the synthetic pathway of the phytohormone indole-3-acetic acid in the plant-beneficial *Bacillus amyloliquefaciens* SQR9. **Microbial Cell Factories**, v. 14, n. 1, p. 130, 2015.

SHAW, L. The role and regulation of the extracellular proteases of *Staphylococcus aureus*. **Microbiology**, v. 150, n. 1, p. 217–228, 2004.

SHURTLEFF, W.; AOYAGI, A. History of edamame, green vegetable soybeans and vegetable-type soybeans (1275 - 2009): **Extensively annotated bibliography and sourcebook**. Soyinfo Center, USA, 2009.

- SINGER, G. A. C.; HICKEY, D. A. Nucleotide bias causes a genomewide bias in the amino acid composition of proteins. **Molecular Biology and Evolution**, v. 17, n. 11, p. 1581–8, 2000.
- SKINNER, M. A.; COOPER, R. A. An *Escherichia coli* mutant defective in the NAD-dependent succinate semialdehyde dehydrogenase. **Archives of Microbiology**, v. 132, p. 270–275, 1984.
- SOEMPHOL, W.; DEERAKSA, A.; MATSUTANI, M.; et al. Global analysis of the genes involved in the thermotolerance mechanism of thermotolerant *Acetobacter tropicalis* SKU1100. **Bioscience, Biotechnology and Biochemistry**, v. 75, n. 10, p. 1921–1928, 2011.
- SONG, C.; YANG, B. Mutagenesis of 18 type III effectors reveals virulence function of XopZ (PXO99) in *Xanthomonas oryzae* pv. *oryzae*. **Molecular Plant-Microbe Interactions: MPMI**, v. 23, n. 7, p. 893–902, 2010.
- TAKEHARA, M.; LING, F.; IZAWA, S.; INQUE, Y.; KIMURA, A. Molecular cloning and nucleotide sequence of purine nucleoside phosphorylase and uridine phosphorylase genes from *Klebsiella* sp. **Bioscience, Biotechnology, and Biochemistry**, v. 59, n. 10, p. 1987–1990, 1995.
- TIAN, W.; SKOLNICK, J. How well is enzyme function conserved as a function of pairwise sequence identity? **Journal of Molecular Biology**, v. 333, n. 4, p. 863–82, 2003.
- TORRES, D.; REVALE, S.; OBANDO, M.; et al. Genome sequence of *Bradyrhizobium japonicum* E109, one of the most agronomically used nitrogen-fixing rhizobacteria in Argentina. **Genome Announcements**, v. 3, n. 1, p. 1–2, 2015.
- TRAMONTANO, A. The ten most wanted solutions in protein bioinformatics. **Chapman & Hall/CRC mathematical biology and medicine series**, 2005.
- TREVASKIS, B.; COLEBATCH, G.; DESBROSSES, G.; WANDREY, M. WIENKOOP, S.; SAALBACH, G. UDVARDI, M. Differentiation of plant cells during symbiotic nitrogen fixation. **Comparative and Functional Genomics**, v. 3, p. 151–157, 2002.
- TSUKUI, T.; EDA, S.; KANEKO, T.; et al. The Type III Secretion System of *Bradyrhizobium japonicum* USDA122 mediates symbiotic incompatibility with Rj2 soybean plants. **Applied and Environmental Microbiology**, v. 79, n. 3, p. 1048–1051, 2013.
- UDIKOVIĆ-KOLIĆ, N., SCOTT, C., MARTIN-LAURENT, F., Evolution of atrazine-degrading capabilities in the environment. **Applied Microbiology and Biotechnology**, v. 96, p. 1175–1189, 2012.
- VARGAS, M. A. T.; PERES, J. R. R.; SUHET, A. R. Reinoculação da soja em função dos sorogrupos de *Rhizobium japonicum* predominantes em solos de Cerrados. In: SEMINÁRIO NACIONAL DE PESQUISA DE SOJA, 2., 1980, Brasília: **Anais do Seminário Nacional de Pesquisa de Soja**, Embrapa-CPAC, p.715-723, 1981.
- VARGAS, M. A. T.; SUHET, A. R. Efeito de tipos e níveis de inoculantes na soja cultivada em um solo de Cerrados. **Pesquisa Agropecuária Brasileira**, v. 15, p. 343–347, 1980.

WANG, L.; STEGEMANN, J. P. Extraction of high quality RNA from polysaccharide matrices using cetyltrimethylammonium bromide. **Biomaterials**, v. 31, n. 7, p. 1612-1618, 2010.

WEI, M.; TAKESHIMA, K.; YOKOYAMA, T.; et al. Temperature-dependent expression of type III secretion system genes and its regulation in *Bradyrhizobium japonicum*. **Molecular Plant-Microbe Interactions : MPMI**, v. 23, n. 5, p. 628–637, 2010.

WENZEL M.; LANG K.; GÜNTHER T.; BHANDARI A.; WEISS A.; LULCHEV P.; SZENTGYÖRGYI E.; KRANZUSCH B.; GÖTTFERT M.;. Characterization of the flavonoid-responsive regulator FrrA and its binding sites. **Journal of Bacteriology**, v. 194, p. 2363–2370, 2012.

WILLIAMS P. H.; Black rot: a continuing threat to world crucifers. **Plant Disease**, v. 64, p.736 742, 1980.

WILLEMS, A. The taxonomy of rhizobia: an overview. **Plant and Soil**, v. 287, p. 3-14, 2006.

ZILLI, J. E. ; CAMPO, MARSON, L. C.; CAMPO, R. J.; GIANLUPPI, V.; HUNGRIA, M. Avaliação da fixação biológica do nitrogênio na soja em áreas de primeiro cultivo no cerrado de Roraima. Boa vista: Embrapa Roraima, 2006. 9p. (**Comunicado Técnico** nº 20).

7. MATERIAL SUPLEMENTAR

Tabela 5. Presença de proteínas hipotéticas similares a proteína BS08216 em outros genomas de estirpes de *Bradyrhizobium*

Fonte	Proteína	Nome	Organismo	Estirpe
RefSeq	WP_035711403.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobium</i> sp. CCBAU 43298	CCBAU 43298
RefSeq	WP_028143522.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobium</i> sp. CCBAU 15635	CCBAU 15635
RefSeq	WP_028143522.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobiumjaponicum</i> USDA 4	USDA 4
RefSeq	WP_028160545.1	hypothetical protein [<i>Bradyrhizobiumjaponicum</i>]	<i>Bradyrhizobiumjaponicum</i> USDA 38	USDA 38
RefSeq	WP_038974989.1	hypothetical protein [<i>Bradyrhizobium</i> sp. CCBAU 15615]	<i>Bradyrhizobium</i> sp. CCBAU 15615	CCBAU 15615
RefSeq	WP_018269858.1	hypothetical protein [<i>Bradyrhizobium elkanii</i>]	<i>Bradyrhizobium elkanii</i> USDA 76	USDA 76
RefSeq	WP_049820113.1	hypothetical protein [<i>Bradyrhizobium japonicum</i>]	<i>Bradyrhizobium japonicum</i> 22	22
RefSeq	WP_051378392.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobium</i> sp. USDA 3384	USDA 3384
RefSeq	WP_051378392.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobium</i> genosp. SA-4 str. CB756	CB756

Tabela 6. Presença de proteínas hipotéticas similares a proteína BS08240 em outros genomas de estirpes de *Bradyrhizobium*

Fonte	Proteína	Nome	Organismo	Estirpe
RefSeq	WP_035711445.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobiumjaponicum</i> CCBAU 25435	CCBAU 25435
RefSeq	WP_035711445.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobium</i> sp. CCBAU 43298	CCBAU 43298
RefSeq	WP_038950847.1	hypothetical protein [<i>Bradyrhizobium</i> sp. CCBAU 15544]	<i>Bradyrhizobium</i> sp. CCBAU 15544	CCBAU 15544
RefSeq	WP_041960905.1	hypothetical protein [<i>Bradyrhizobiumjaponicum</i>]	<i>Bradyrhizobiumjaponicum</i>	Is-34
RefSeq	WP_028143540.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobium</i> sp. CCBAU 15615	CCBAU 15615
RefSeq	WP_028143540.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobiumjaponicum</i> USDA 4	USDA 4
RefSeq	WP_038973109.1	hypothetical protein [<i>Bradyrhizobium</i> sp. CCBAU 15635]	<i>Bradyrhizobium</i> sp. CCBAU 15635	CCBAU 15635
RefSeq	WP_028160484.1	hypothetical protein [<i>Bradyrhizobiumjaponicum</i>]	<i>Bradyrhizobiumjaponicum</i> USDA 38	USDA 38

Tabela 7. Presença de proteínas hipotéticas similares a proteína BS08251 em outros genomas de estirpes de *Bradyrhizobium*

Fonte	Proteína	Nome	Organismo	Estirpe
RefSeq	WP_035711482.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobium</i> sp. CCBAU 43298	CCBAU 43298
RefSeq	WP_028143548.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobium</i> sp. CCBAU 15615	CCBAU 15615
RefSeq	WP_028143548.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobium</i> sp. CCBAU 15635	CCBAU 15635
RefSeq	WP_028143548.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobiumjaponicum</i> USDA 4	USDA 4
RefSeq	WP_038951083.1	hypothetical protein [<i>Bradyrhizobium</i> sp. CCBAU 15544]	<i>Bradyrhizobium</i> sp. CCBAU 15544	CCBAU 15544
RefSeq	WP_036012195.1	hypothetical protein [<i>Bradyrhizobium</i> yuanmingense]	<i>Bradyrhizobiumyuanmingense</i> CCBAU 35157	CCBAU 35157
RefSeq	WP_038936579.1	hypothetical protein [<i>Bradyrhizobiumjaponicum</i>]	<i>Bradyrhizobiumjaponicum</i> CCBAU 25435	CCBAU 25435
RefSeq	WP_038936579.1	hypothetical protein [<i>Bradyrhizobiumjaponicum</i>]	<i>Bradyrhizobiumjaponicum</i> USDA 38	USDA 38
RefSeq	WP_044536925.1*	hypothetical protein [<i>Bradyrhizobium</i> sp. LTSP885]	<i>Bradyrhizobium</i> sp. LTSP885	LTSP885

*Apresentou identidade de 55% e similaridade de 68%

Tabela 8. Presença de proteínas hipotéticas similares a proteína BS08254 em outros genomas de estirpes de *Bradyrhizobium*

Fonte	Proteína	Nome	Organismo	Estirpe
RefSeq	WP_049806740.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobium</i> sp. CCBAU 43298	CCBAU 43298
RefSeq	WP_049806740.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobium</i> sp. CCBAU 15544	CCBAU 15544
RefSeq	WP_049806740.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobium japonicum</i> USDA 38	USDA 38
RefSeq	WP_051057125.1	hypothetical protein [<i>Bradyrhizobium</i> yuanmingense]	<i>Bradyrhizobium</i> yuanmingense CCBAU 35157	CCBAU 35157
RefSeq	WP_049802022.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobium</i> sp. CCBAU 15615	CCBAU 15615
RefSeq	WP_049802022.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobium</i> sp. CCBAU 15635	CCBAU 15635
RefSeq	WP_049802022.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobium japonicum</i> USDA 4	USDA 4
RefSeq	WP_052832958.1	hypothetical protein [<i>Bradyrhizobium</i> sp. LTSP885]	<i>Bradyrhizobium</i> sp. LTSP885	LTSP885

Tabela 9. Presença de proteínas hipotéticas similares a proteína BS08258 em outros genomas de estirpes de *Bradyrhizobium*

Fonte	Proteína	Nome	Organismo	Estirpe
RefSeq	WP_049813053.1	MULTISPECIES: aspartate aminotransferase [<i>Bradyrhizobium</i>]	<i>Bradyrhizobiumjaponicum</i> CCBAU 15618	CCBAU 15618
RefSeq	WP_049813053.1	MULTISPECIES: aspartate aminotransferase [<i>Bradyrhizobium</i>]	<i>Bradyrhizobium</i> sp. CCBAU 43298	CCBAU 43298
RefSeq	WP_041956560.1	aspartate aminotransferase [<i>Bradyrhizobiumjaponicum</i>]	<i>Bradyrhizobiumjaponicum</i>	Is-34
RefSeq	WP_049802023.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobiumjaponicum</i> CCBAU 25435	CCBAU 25435
RefSeq	WP_049802023.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobium</i> sp. CCBAU 15544	CCBAU 15544
RefSeq	WP_049802023.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobium</i> sp. CCBAU 15615	CCBAU 15615
RefSeq	WP_049802023.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobium</i> sp. CCBAU 15635	CCBAU 15635
RefSeq	WP_049802023.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobiumjaponicum</i> USDA 4	USDA 4
RefSeq	WP_028351390.1*	aspartate aminotransferase [<i>Bradyrhizobium elkanii</i>]	<i>Bradyrhizobium elkanii</i> WSM1741	WSM1741
RefSeq	WP_050424563.1**	aspartate aminotransferase [<i>Bradyrhizobium</i> sp. SEMIA 6148]	<i>Bradyrhizobium</i> sp. SEMIA 6148	SEMIA 6148
RefSeq	WP_024510015.1	aspartate aminotransferase [<i>Bradyrhizobium</i> sp. ARR65]	<i>Bradyrhizobium</i> sp. ARR65	ARR65

* Apresentou identidade de 54% e similaridade de 66%

** Apresentou identidade de 55% e similaridade de 67%

Tabela 10. Presença de proteínas hipotéticas similares a proteína BS08261 em outros genomas de estirpes de *Bradyrhizobium*

Fonte	Proteína	Nome	Organismo	Estirpe
RefSeq	WP_035711516.1	MULTISPECIES: aldehyde dehydrogenase [<i>Bradyrhizobium</i>]	<i>Bradyrhizobiumjaponicum</i> CCBAU 15618	CCBAU 15618
RefSeq	WP_035711516.1	MULTISPECIES: aldehyde dehydrogenase [<i>Bradyrhizobium</i>]	<i>Bradyrhizobiumjaponicum</i> CCBAU 25435	CCBAU 25435
RefSeq	WP_035711516.1	MULTISPECIES: aldehyde dehydrogenase [<i>Bradyrhizobium</i>]	<i>Bradyrhizobium</i> sp. CCBAU 43298	CCBAU 43298
RefSeq	WP_035711516.1	MULTISPECIES: aldehyde dehydrogenase [<i>Bradyrhizobium</i>]	<i>Bradyrhizobium</i> sp. CCBAU 15635	CCBAU 15635
RefSeq	WP_035711516.1	MULTISPECIES: aldehyde dehydrogenase [<i>Bradyrhizobium</i>]	<i>Bradyrhizobiumjaponicum</i> USDA 38	USDA 38
RefSeq	WP_038950855.1	MULTISPECIES: aldehyde dehydrogenase [<i>Bradyrhizobium</i>]	<i>Bradyrhizobium</i> sp. CCBAU 15544	CCBAU 15544
RefSeq	WP_038950855.1	MULTISPECIES: aldehyde dehydrogenase [<i>Bradyrhizobium</i>]	<i>Bradyrhizobium</i> sp. CCBAU 15615	CCBAU 15615
RefSeq	WP_038950855.1	MULTISPECIES: aldehyde dehydrogenase [<i>Bradyrhizobium</i>]	<i>Bradyrhizobiumjaponicum</i> USDA 4	USDA 4
RefSeq	WP_041956641.1	aldehyde dehydrogenase [<i>Bradyrhizobiumjaponicum</i>]	<i>Bradyrhizobiumjaponicum</i>	Is-34
RefSeq	WP_051378179.1	MULTISPECIES: aldehyde	<i>Bradyrhizobium</i> sp. USDA	USDA

		dehydrogenase [<i>Bradyrhizobium</i>]	3384	3384
RefSeq	WP_051378179.1	MULTISPECIES: aldehyde dehydrogenase [<i>Bradyrhizobium</i>]	<i>Bradyrhizobium</i> genosp. SA-4 str. CB756	CB756
RefSeq	WP_018647591.1	aldehyde dehydrogenase [<i>Bradyrhizobiumjaponicum</i>]	<i>Bradyrhizobiumjaponicum</i>	USDA 124
RefSeq	WP_028177772.1	MULTISPECIES: aldehyde dehydrogenase [<i>Bradyrhizobium</i>]	<i>Bradyrhizobium</i> liaoningense	CCBAU 05525
RefSeq	WP_028177772.1	MULTISPECIES: aldehyde dehydrogenase [<i>Bradyrhizobium</i>]	<i>Bradyrhizobium</i> liaoningense	CCBAU 83689
RefSeq	WP_028177772.1	MULTISPECIES: aldehyde dehydrogenase [<i>Bradyrhizobium</i>]	<i>Bradyrhizobiumjaponicum</i>	USDA 135
RefSeq	WP_028148451.1	aldehyde dehydrogenase [<i>Bradyrhizobiumjaponicum</i>]	<i>Bradyrhizobiumjaponicum</i>	CCBAU 15517
RefSeq	WP_028148451.1	aldehyde dehydrogenase [<i>Bradyrhizobiumjaponicum</i>]	<i>Bradyrhizobiumjaponicum</i>	USDA 123
RefSeq	WP_028169617.1	aldehyde dehydrogenase [<i>Bradyrhizobium japonicum</i>]	<i>Bradyrhizobium japonicum</i>	USDA 6
RefSeq	WP_028169617.1	aldehyde dehydrogenase [<i>Bradyrhizobium japonicum</i>]	<i>Bradyrhizobium japonicum</i>	E109
RefSeq	WP_028169617.1	aldehyde dehydrogenase [<i>Bradyrhizobium japonicum</i>]	<i>Bradyrhizobium japonicum</i>	FN1
RefSeq	WP_027571913.1	aldehyde dehydrogenase [<i>Bradyrhizobium</i> sp. WSM1743]	<i>Bradyrhizobium</i> sp. WSM1743	WSM1743
RefSeq	WP_038934240.1	aldehyde dehydrogenase [<i>Bradyrhizobium japonicum</i>]	<i>Bradyrhizobium japonicum</i>	CCBAU 15618
RefSeq	WP_038934240.1	aldehyde dehydrogenase [<i>Bradyrhizobium japonicum</i>]	<i>Bradyrhizobium japonicum</i>	CCBAU 25435
RefSeq	WP_028155693.1	aldehyde dehydrogenase [<i>Bradyrhizobium japonicum</i>]	<i>Bradyrhizobium japonicum</i>	USDA 38
RefSeq	WP_041955615.1	aldehyde dehydrogenase [<i>Bradyrhizobium japonicum</i>]	<i>Bradyrhizobium japonicum</i>	Is-34
RefSeq	WP_027558988.1	MULTISPECIES: aldehyde dehydrogenase [<i>Bradyrhizobium</i>]	<i>Bradyrhizobium</i> sp. USDA 3384	USDA 3384
RefSeq	WP_027558988.1	MULTISPECIES: aldehyde dehydrogenase [<i>Bradyrhizobium</i>]	<i>Bradyrhizobium</i> genosp. SA-4 str. CB756	CB756

Tabela 11. Presença de proteínas hipotéticas similares a proteína BS08267 em outros genomas de estirpes de *Bradyrhizobium*

Fonte	Proteína	Nome	Organismo	Estirpe
RefSeq	WP_018645251.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobiumjaponicum</i>	CCBAU 15618
RefSeq	WP_018645251.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobiumjaponicum</i>	CCBAU 25435
RefSeq	WP_018645251.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobium</i> sp. CCBAU 43298	CCBAU 43298
RefSeq	WP_018645251.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobium</i> sp. CCBAU 15615	CCBAU 15615
RefSeq	WP_018645251.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobium</i> sp. CCBAU 15635	CCBAU 15635
RefSeq	WP_018645251.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobiumjaponicum</i>	USDA 4
RefSeq	WP_018645251.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobiumjaponicum</i>	USDA 38
RefSeq	WP_018645251.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobiumjaponicum</i>	USDA 124
RefSeq	WP_049806726.1	hypothetical protein	<i>Bradyrhizobium</i> sp. CCBAU	CCBAU 15544

		[<i>Bradyrhizobium</i> sp. CCBAU 15544]	15544	
RefSeq	WP_052435595.1	hypothetical protein [Bradyrhizobiumjaponicum]	<i>Bradyrhizobiumjaponicum</i>	Is-34
RefSeq	WP_051380469.1	hypothetical protein [Bradyrhizobium sp. WSM1743]	<i>Bradyrhizobium</i> sp. WSM1743	WSM1743
RefSeq	WP_051378181.1	MULTISPECIES: hypothetical protein [Bradyrhizobium]	<i>Bradyrhizobium</i> sp. USDA 3384	USDA 3384
RefSeq	WP_051378181.1	MULTISPECIES: hypothetical protein [Bradyrhizobium]	<i>Bradyrhizobium</i> genosp. SA-4 str. CB756	CB756
RefSeq	WP_051379660.1	hypothetical protein [Bradyrhizobiummelkanii]	<i>Bradyrhizobiummelkanii</i>	WSM1741
RefSeq	WP_052363144.1	hypothetical protein [Bradyrhizobium sp. DOA9]	<i>Bradyrhizobium</i> sp. DOA9	DOA9
RefSeq	WP_024512700.1	hypothetical protein [Bradyrhizobium sp. ARR65]	<i>Bradyrhizobium</i> sp. ARR65	ARR65
RefSeq	WP_049823548.1	hypothetical protein [Bradyrhizobium sp. WSM2254]	<i>Bradyrhizobium</i> sp. WSM2254	WSM2254
RefSeq	WP_026233436.1	hypothetical protein [Bradyrhizobium sp. WSM4349]	<i>Bradyrhizobium</i> sp. WSM4349	WSM4349
RefSeq	WP_027515546.1	hypothetical protein [Bradyrhizobium sp. WSM1417]	<i>Bradyrhizobium</i> sp. WSM1417	WSM1417

Tabela 12. Presença de proteínas hipotéticas similares a proteína BS08317 em outros genomas de estirpes de *Bradyrhizobium*

Fonte	Proteína	Nome	Organismo	Estirpe
RefSeq	WP_011084660.1	MULTISPECIES: hypothetical protein [Bradyrhizobium]	<i>Bradyrhizobiumdiazoefficiens</i> USDA 110	USDA 110
RefSeq	WP_011084660.1	MULTISPECIES: hypothetical protein [Bradyrhizobium]	<i>Bradyrhizobiumjaponicum</i> USDA 6	USDA6
RefSeq	WP_011084660.1	MULTISPECIES: hypothetical protein [Bradyrhizobium]	<i>Bradyrhizobiumjaponicum</i> CCBAU 25435	CCBAU 25435
RefSeq	WP_011084660.1	MULTISPECIES: hypothetical protein [Bradyrhizobium]	<i>Bradyrhizobium</i> sp. CCBAU 43298	CCBAU 43298
RefSeq	WP_011084660.1	MULTISPECIES: hypothetical protein [Bradyrhizobium]	<i>Bradyrhizobium</i> sp. CCBAU 15544	CCBAU 15544
RefSeq	WP_011084660.1	MULTISPECIES: hypothetical protein [Bradyrhizobium]	<i>Bradyrhizobium</i> sp. CCBAU 15635	CCBAU 15635
RefSeq	WP_011084660.1	MULTISPECIES: hypothetical protein [Bradyrhizobium]	<i>Bradyrhizobium</i> sp. CCBAU 41267	CCBAU 41267
RefSeq	WP_011084660.1	MULTISPECIES: hypothetical protein [Bradyrhizobium]	<i>Bradyrhizobiumjaponicum</i> USDA 38	USDA 38
RefSeq	WP_011084660.1	MULTISPECIES: hypothetical protein [Bradyrhizobium]	<i>Bradyrhizobiumjaponicum</i>	E109
RefSeq	WP_011084660.1	MULTISPECIES: hypothetical protein [Bradyrhizobium]	<i>Bradyrhizobiumjaponicum</i>	FN1
RefSeq	WP_018648290.1	MULTISPECIES: hypothetical protein [Bradyrhizobium]	<i>Bradyrhizobium</i> diazoefficiens SEMIA 5080	SEMIA 5080
RefSeq	WP_018648290.1	MULTISPECIES: hypothetical protein [Bradyrhizobium]	<i>Bradyrhizobiumjaponicum</i> CCBAU 15618	CCBAU 15618

RefSeq	WP_018648290.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobiumjaponicum</i> USDA 122	USDA 122
RefSeq	WP_018648290.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobiumjaponicum</i> USDA 124	USDA 124
RefSeq	WP_018648290.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobiumjaponicum</i>	Is-1
RefSeq	WP_041959982.1	hypothetical protein [<i>Bradyrhizobiumjaponicum</i>]	<i>Bradyrhizobiumjaponicum</i>	Is-34
RefSeq	WP_028144278.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobium</i> sp. CCBAU 15615	CCBAU 15615
RefSeq	WP_028144278.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobiumjaponicum</i> USDA 4	USDA 4
RefSeq	WP_035680466.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobiumjaponicum</i> CCBAU 15517	CCBAU 15517
RefSeq	WP_035680466.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobiumjaponicum</i> CCBAU 83623	CCBAU 83623
RefSeq	WP_035680466.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobiumliaoningense</i> CCBAU 05525	CCBAU 05525
RefSeq	WP_035680466.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobium</i> liaoningense CCBAU 83689	CCBAU 83689
RefSeq	WP_028182298.1	hypothetical protein [<i>Bradyrhizobiumjaponicum</i>]	<i>Bradyrhizobiumjaponicum</i> USDA 135	USDA 135
RefSeq	WP_028154143.1	hypothetical protein [<i>Bradyrhizobiumjaponicum</i>]	<i>Bradyrhizobiumjaponicum</i> USDA 123	USDA 123
RefSeq	WP_018319537.1	hypothetical protein [<i>Bradyrhizobium</i> sp. WSM2793]	<i>Bradyrhizobium</i> sp. WSM2793	WSM2793
RefSeq	WP_036043218.1	hypothetical protein [<i>Bradyrhizobiumyuanmingense</i>]	<i>Bradyrhizobiumyuanmingense</i> CCBAU 05623	CCBAU 05623
RefSeq	WP_036014785.1	hypothetical protein [<i>Bradyrhizobiumyuanmingense</i>]	<i>Bradyrhizobiumyuanmingense</i> CCBAU 35157	CCBAU 35157
RefSeq	WP_036030832.1	hypothetical protein [<i>Bradyrhizobiumyuanmingense</i>]	<i>Bradyrhizobiumyuanmingense</i> CCBAU 25021	CCBAU 25021
RefSeq	WP_027563395.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobium</i> sp. USDA 3384	USDA 3384
RefSeq	WP_027563395.1	MULTISPECIES: hypothetical protein [<i>Bradyrhizobium</i>]	<i>Bradyrhizobium</i> genosp. SA-4 str. CB756	CB756
RefSeq	WP_050631893.1	hypothetical protein [<i>Bradyrhizobium</i> sp. SEMIA 690]	<i>Bradyrhizobium</i> sp. SEMIA 690	SEMIA 690
RefSeq	WP_050399944.1	hypothetical protein [<i>Bradyrhizobium</i> sp. SEMIA 6208]	<i>Bradyrhizobium</i> sp. SEMIA 6208	SEMIA 6208
RefSeq	WP_026192329.1	hypothetical protein [<i>Bradyrhizobiummelkanii</i>]	<i>Bradyrhizobiummelkanii</i> CCBAU 05737	CCBAU 05737
RefSeq	WP_026192329.1	hypothetical protein [<i>Bradyrhizobiummelkanii</i>]	<i>Bradyrhizobiummelkanii</i> CCBAU 43297	CCBAU 43297
RefSeq	WP_026192329.1	hypothetical protein [<i>Bradyrhizobiummelkanii</i>]	<i>Bradyrhizobiummelkanii</i> USDA 94	USDA 94
RefSeq	WP_026192329.1	hypothetical protein [<i>Bradyrhizobiummelkanii</i>]	<i>Bradyrhizobiummelkanii</i> USDA 76	USDA 76
RefSeq	WP_036031342.1	hypothetical protein [<i>Bradyrhizobiumyuanmingense</i>]	<i>Bradyrhizobiumyuanmingense</i> CCBAU 25021	CCBAU 25021
RefSeq	WP_050420728.1	hypothetical protein [<i>Bradyrhizobium</i> sp. SEMIA 6148]	<i>Bradyrhizobium</i> sp. SEMIA 6148	SEMIA 6148

RefSeq	WP_029083940.1	hypothetical protein [<i>Bradyrhizobium</i> sp. th.b2]	<i>Bradyrhizobium</i>sp. th.b2	th.b2
RefSeq	WP_025038801.1	hypothetical protein [<i>Bradyrhizobium</i> sp. DOA9]	<i>Bradyrhizobium</i>sp. DOA9	DOA9

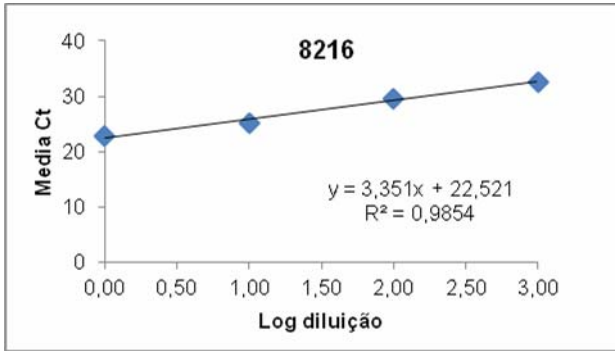
Tabela 13. Predição funcional das 7 proteínas hipotéticas localizadas na ilha simbiótica de estirpe CPAC 15 de *B. japonicum*

Proteína	NCBI ID	Predição de Funções em Bancos de Dados
BS08216	gi 627779337	
Transcriptional regulator, LysR family (<i>Burkholderia pseudomallei</i> 305) - 31,1% de identidade (Uniprot) Processo Biológico – Regulação da Transcrição (GO)		
Molecular chaperone Tir (<i>Burkholderia sp.</i> USM B20) – 28,4% de identidade (Uniprot) Processo Biológico – Transdução de Sinal (GO)		
TIR protein (<i>Methylobacterium extorquens</i>) – 28,0% de identidade (Uniprot) Transdução de Sinal (GO)		
Molecular chaperone Tir -(<i>Mesorhizobium sp.</i> LSJC280B00) - 29.4% de identidade (Uniprot) – Transdução de Sinal (GO)		
Alpha/beta hydrolase (<i>Mesorhizobium sp.</i> LNHC221B00) - 31,9% de identidade (Uniprot) Função Molecular: Hidrolase Atividade (GO)		
Zn finger protein HypA/HybF (Possibly regulating hydrogenase expression) - <i>Butyrivibrio fibrisolvens</i> - 35,7% de identidade (Uniprot) Processo de modificações de Proteínas Celulares (GO)		
Proteína	NCBI ID	Predição de Funções em Bancos de Dados
BS08251	gi 627779303	
Clã: Peptidase; Família: Trypsin_2; Domínio: Trypsin-like peptidase (Pfam) KEGG: EC: 3.4.21.107		
Processo Biológico: Proteólise (GO) Função Molecular: Atividade catalítica, serine-type endopeptidase atividade Categoria Funcional: Degradação de proteínas, peptídeos e glicopeptídeos Modificação Pós-Traducional, turnover de proteína, chaperonas (Rhizobase – Gene ID:AZLe02840)		
Putative serine protease - <i>Bradyrhizobium sp.</i> ORS 375 - 35.4% de identidade (Uniprot)		
Probable serine protease <i>Sorangium cellulosum</i> So ce56 – 36.9% de identidade (Uniprot)		
Proteína	NCBI ID	Predição de Funções em Bancos de Dados
BS08258	gi 627779297	
Clã: PLP_aminotran; Família: Aminotran_1_2 Domínio: Aminotran_1_2 (Pfam) KEGG: EC 2.6.1.1		
Processo Biológico: Processos de Biossíntese, como parte do metabolismo de substâncias (GO) Função Molecular: Atividade de Transaminase (GO) Categoria Funcional: Metabolismo e Transporte de Aminoácidos (Rhizobase – Gene ID:RHE_PB00131)		
Aspartate aminotransferase - <i>Mesorhizobium metallidurans</i> STM 2683 - 55.6% de identidade (Uniprot)		
Aspartate aminotransferase - <i>Rhizobium loti</i> (strain MAFF303099) - 53.3% de identidade (Uniprot)		
Aspartate aminotransferase- <i>Bradyrhizobium sp.</i> DOA9 - 49.9% de identidade (Uniprot)		
Proteína	NCBI ID	Predição de Funções em Bancos de Dados
BS08254	gi 627779300	
Clã: Peptidase_PA; Família: Trypsin; Domínio:Trypsin (Pfam) KEGG: EC 3.4.21.19		
Processo Biológico: Proteólise (GO) Função Molecular: serine-type endopeptidase activity (GO) Categoria Funcional Biossíntese de co-fatores, grupos prostéticos e carregadores - (Rhizobase – Gene ID: mll6878)		
Glutamyl endopeptidase*- <i>Bacillus licheniformis</i> ATCC 14580 - 39.6% de identidade (Uniprot)		
Proteína	NCBI ID	Predição de Funções em Bancos de Dados
BS08261	gi 627779294	
Clã ALDH-like; Família Aldedh - Aldehyde dehydrogenase; Domínio Aldedh (Pfam) Ald_DH/histidinol_DH.- IPR016161 (Interpro) Ald_DH_C. - IPR016163.(Interpro)		

Ald_DH_N. - IPR016162 (Interpro) Aldehyde_DH_dom - IPR015590 (Interpro) Subfamília: succinate-semialdehyde dehydrogenase (PANTHER) Domínio: ALDH_F5_SSADH_GabD - succinate-semialdehyde dehydrogenase –(NCBI) KEGG: EC=1.2.1.24		
Processo Biológico: Processo Metabólico (GO) Função Molecular: atividade de oxidoreductase que atuam no grupo aldeido ou oxo dos doadores (GO) Categoria Funcional: Metabolismo de Pequenas Moléculas (Rhizobase – Gene ID: RL0101)		
Succinate-semialdehyde dehydrogenase [NADP+] <i>Oligotropha carboxidovorans</i> - strain ATCC 49405 - 71.2% de identidade (Uniprot)		
Aldehyde dehydrogenase - <i>Burkholderia</i> sp. SJ98 - 71.8% de identidade (Uniprot)		
Proteína	NCBI ID	Predição de Funções em Bancos de Dados
BS08267	gi 627779288	
Clã MFS - Major Facilitator Superfamily; Família: MFS_1; Domínio: MFS_1 (Pfam) Domínios: Metabolite/H+ symporter, major facilitator superfamily (MFS); (NCBI) Domínios: Multidrug resistance protein (NCBI) subfamily: metabolite-H(+) symport (MHS) (NCBI) KEGG: [TC:2.A.1.6]		
Função Molecular: Atividade de Transporte Transmembrana (GO) Processo Biológico: Transporte (GO) Categoria Funcional: Transporte e Ligação de Proteínas (Rhizobase – Gene ID: GKPORF_B1827) Múltiplas Categorias – (Rhizobase – Gene ID: BJ6T_68620)		
Alpha-ketoglutarate permease (<i>Bradyrhizobium</i> sp. DOA9) - 81.9% de identidade (Uniprot)		
Major facilitator transporter - <i>Bradyrhizobium</i> sp. CCGE-LA001 - 79.9% de identidade (Uniprot)		
Proteína	NCBI ID	Predição de Funções em Bancos de Dados
BS08317	gi 627779238	
Superfamília ComA - Phosphosulfolactate synthase, CoM biossíntese proteína A [Coenzima transporte e metabolismo] (NCBI)		
Proteína Efetora do Sistema de Secreção do Tipo III - <i>Xanthomonas campestris</i> pv. <i>Campestris</i> - 31.1%de identidade (Uniprot)		
Type III effector protein XopP - <i>Xanthomonas campestris</i> pv. <i>raphani</i> 756C - 31.1%de identidade (Uniprot)		
Putative type III effector (HLK1) - <i>Ralstonia solanacearum</i> GMI1000 - 31%de identidade (Uniprot)		

Tabela 14. Dados utilizados para o cálculo da eficiência de amplificação do *primer* para o gene *bjs08216*

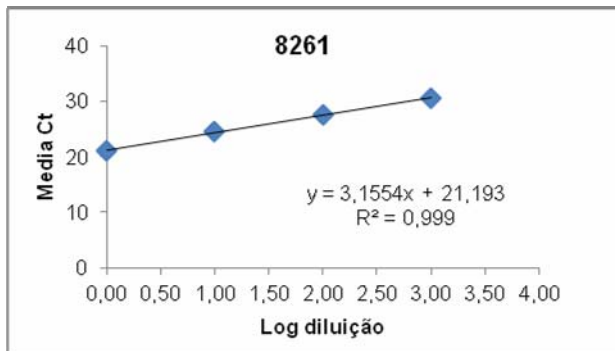
8216				
Well	Diluição	Ct	Média Ct	Log diluição
D1	1	22,72	22,89	0,00
D2		23,06		
D3	10	24,16	25,13	1,00
D4		26,09		
D5	100	29,58	29,60	2,00
D6		29,62		
D7	1000	32,07	32,57	3,00
D8		33,07		
D9	10000	33,96	33,98	4,00
D10		34,00		
D11	Bco	0,00	0	0
D12		0,00		



Eficiência
0,99

Tabela 15. Dados utilizados para o cálculo da eficiência de amplificação do *primer* para o gene *bjs08261*

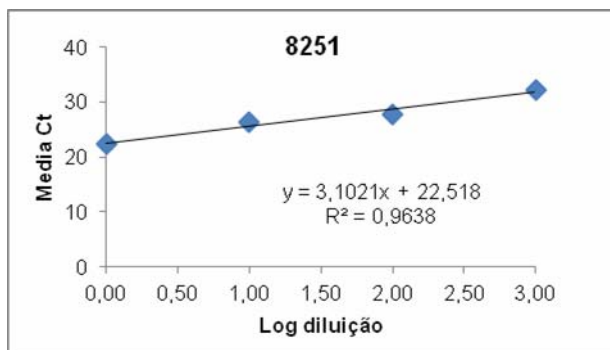
8261				
Well	Diluição	Ct	Média Ct	Log diluição
E1	1	21,14	21,07	0,00
E2		21,00		
E3	10	24,62	24,49	1,00
E4		24,36		
E5	100	27,64	27,59	2,00
E6		27,53		
E7	1000	30,88	30,56	3,00
E8		30,24		
E9	10000	35,19	35,19	4,00
E10		35,19		
E11	Bco	0,00	0,00	
E12		0,00		



Eficiência
0,93

Tabela 16. Dados utilizados para o cálculo da eficiência de amplificação do *primer* para o gene *bjs08251*

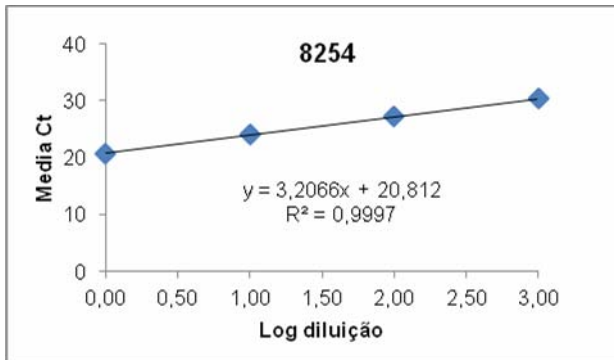
8251				
Well	Diluição	Ct	Média Ct	Log diluição
B1	1	22,51	22,42	0,00
B2		22,33		
B3	10	26,4722	26,29	1,00
B4		26,111		
B5	100	29,2225	27,67	2,00
B6		26,11		
B7	1000	32,7127	32,30	3,00
B8		31,8951		
B9	10000	35,7203	32,86	4,00
B10		30		
B11	Bco	0,00	0,00	
B12		0,00		



Eficiência
0,90

Tabela 17. Dados utilizados para o cálculo da eficiência de amplificação do *primer* para o gene *bjs08254*

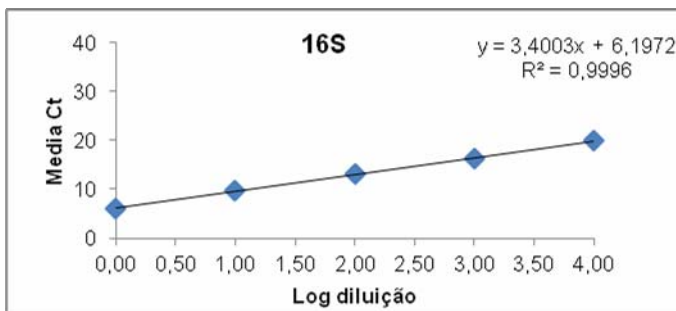
8254				
Well	Diluição	Ct	Média Ct	Log diluição
D1	1	20,07	20,76	0,00
D2		21,44		
D3	10	24,15	24,07	1,00
D4		23,98		
D5	100	27,33	27,30	2,00
D6		27,26		
D7	1000	30,48	30,37	3,00
D8		30,26		
D9	10000	33,23	31,12	4,00
D10		29,00		
D11	Bco	0,00		
D12		0,00		



Eficiência
0,95

Tabela 18. Dados utilizados para o cálculo da eficiência de amplificação do *primer* para o gene *16S*

16S				
Well	Diluição	Ct	Média Ct	Log diluição
E1	1	6,14	6,14	0,00
E2		6,14		
E3	10	9,58	9,66	1,00
E4		9,74		
E5	100	13,07	13,10	2,00
E6		13,13		
E7	1000	16,24	16,23	3,00
E8		16,22		
E9	10000	19,83	19,86	4,00
E10		19,88		
E11	Bco	0,00	0,00	
E12		0,00		

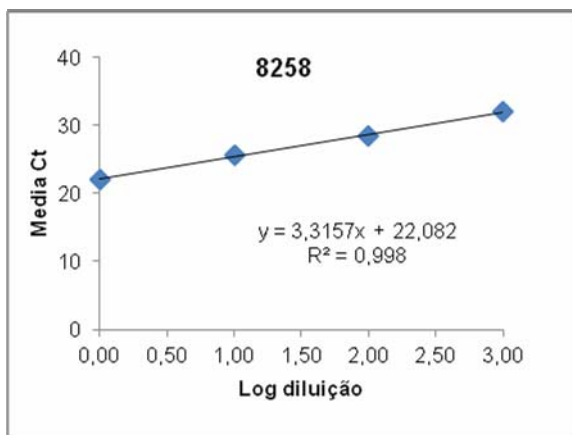


Eficiência
0,97

Tabela 19. Dados utilizados para o cálculo da eficiência de amplificação do *primer* para o gene *bjs08258*

8258				
Well	Diluição	Ct	Média Ct	Log diluição
B1	1	22,36	22,03	0,00

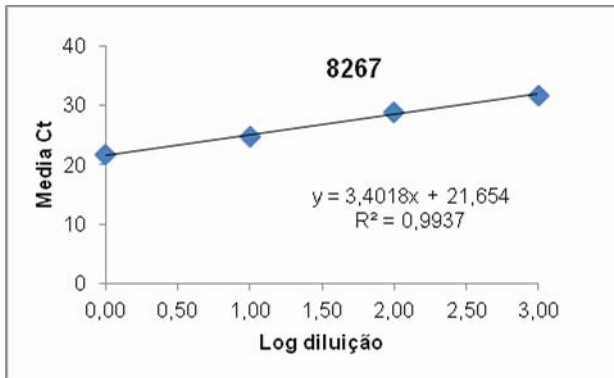
B2		21,70		
B3	10	26	25,60	1,00
B4		25,1934		
B5	100	28,6866	28,48	2,00
B6		28,2648		
B7	1000	31,9159	32,12	3,00
B8		32,328		
B9	10000	0	16,96	4,00
B10		33,9244		
B11	Bco	0,00	0,00	
B12		0,00		



Eficiência
1,00

Tabela 20. Dados utilizados para o cálculo da eficiência de amplificação do *primer* para o gene *bjs08267*

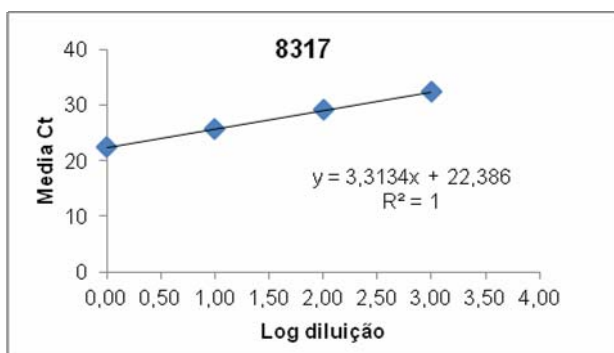
8267				
Well	Diluição	Ct	Média Ct	Log diluição
D1	1	22,27	21,76	0,00
D2		21,25		
D3	10	24,38	24,68	1,00
D4		24,98		
D5	100	28,80	28,89	2,00
D6		28,98		
D7	1000	32,00	31,70	3,00
D8		31,39		
D9	10000	34,85	34,85	4,00
D10				
D11	Bco	0,00		
D12		0,00		



Eficiência
0,97

Tabela 21. Dados utilizados para o cálculo da eficiência de amplificação do *primer* para o gene *bjs08317*

8317				
Well	Diluição	Ct	Média Ct	Log diluição
E1	1	22,5656	22,39	0,00
E2		22,2228		
E3	10	25,8463	25,67	1,00
E4		25,4921		
E5	100	28,8453	29,05	2,00
E6		29,2479		
E7	1000	32,31	32,31	3,00
E8		32,31		
E9	10000	33,26	33,81	4,00
E10		34,36		
E11	Bco	0,00	0,00	
E12		0,00		

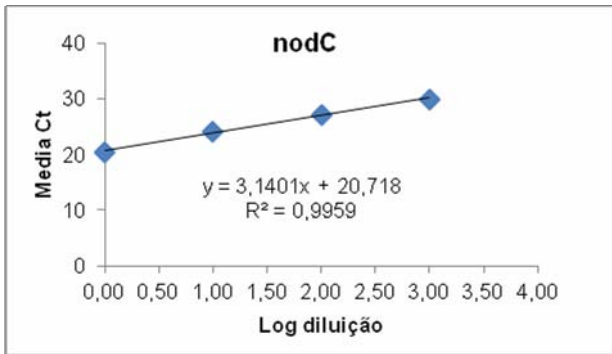


Eficiência
1,00

Tabela 22. Dados utilizados para o cálculo da eficiência de amplificação do *primer* para o gene *nodC*

<i>nodC</i>				
Well	Diluição	Ct	Média Ct	Log diluição
E1	1	20,831	20,47	0,00

E2		20,1187		
E3	10	23,9525	24,15	1,00
E4		24,3456		
E5	100	27,2874	27,14	2,00
E6		27		
E7	1000	29,40	29,94	3,00
E8		30,49		
E9	10000	34,22	35,65	4,00
E10		37,08		
E11	Bco	0,00	0,00	
E12		0,00		



Eficiência
0,92

Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenBank Graphics Distance tree of results

Description	Max score	Total score	Query cover	E value	Ident	Accession
Bradyrhizobium japonicum strain E109, complete genome	4145	5324	100%	0.0	100%	CP010313.1
Bradyrhizobium japonicum SEMIA 5079 genome	4145	5318	100%	0.0	100%	CP007569.1
Bradyrhizobium japonicum USDA 6 DNA, complete genome	4145	5324	100%	0.0	100%	AP012206.1
Bradyrhizobium japonicum USDA 110 DNA, complete genome	4145	5329	100%	0.0	100%	BA000040.2
Bradyrhizobium japonicum symbiotic gene region, part 1 of 2	4145	4145	100%	0.0	100%	AF322012.1
Bradyrhizobium japonicum USDA 122 DNA, type III secretion system cluster region, contig: contig530	4133	4133	100%	0.0	99%	AB758607.1
Bradyrhizobium diazoefficiens DNA, complete genome, strain: NKG	4111	5291	100%	0.0	99%	AP014685.1

Figura 8 - Alinhamentos da sequência do gene *bs08317* em outros genomas à partir do Blastn

type III effector protein XopP [Xanthomonas fragariae LMG 25863]	83.6	83.6	23%	2e-13	29%	EN295648.1
type III effector (hik1) [Ralstonia solanacearum]	84.7	84.7	18%	2e-13	34%	WP_013206096.1
type III effector protein XopP [Xanthomonas fragariae LMG 25863]	82.0	82.0	22%	2e-13	29%	EN295649.1
Putative type III effector protein (HIK2) (fragment) [Ralstonia solanacearum]	80.5	80.5	17%	3e-13	34%	CUV35194.1
conserved hypothetical protein [Xanthomonas campestris pv. campestris]	83.6	83.6	23%	4e-13	31%	CAP52420.1
type III effector protein XopP [Xanthomonas campestris]	83.6	83.6	23%	4e-13	31%	WP_050911902.1
type III effector protein XopP [Xanthomonas campestris]	83.6	83.6	23%	4e-13	31%	WP_040941606.1
type III effector protein XopP [Xanthomonas campestris]	83.6	83.6	23%	4e-13	31%	WP_014507127.1
hypothetical protein [Xanthomonas campestris]	83.6	83.6	23%	4e-13	31%	WP_011036442.1
conserved hypothetical protein [Xanthomonas campestris pv. campestris str. ATCC 33913]	83.6	83.6	23%	5e-13	31%	AAM40545.1

Figura 9 – Alinhamentos da sequência da proteína do gene *bs08317* em outros genomas à partir do Blastp