Universidade de São Paulo
Interunidades de Pós-Graduação em Bioinformática

Gesiele Almeida Barros de Carvalho

**Análise computacional dos genomas de duas estirpes brasileiras
de *Bradyrhizobium* de importância econômica**

– São Paulo, SP –
2016

Gesiele Almeida Barros de Carvalho

**Análise computacional dos genomas de duas estirpes brasileiras
de *Bradyrhizobium* de importância econômica**

Tese apresentada ao programa Interunidades de
Pós-Graduação em Bioinformática da Univer-
sidade de São Paulo, como requisito final para
a obtenção do título de doutora em Ciências.

Área de Concentração: Bioinformática

Orientadora: Prof^a. Dr^a. Marie-Anne Van Sluys

Coorientador: Prof. Dr. Fabrício Martins Lopes

Para o desenvolvimento deste trabalho, a aluna recebeu auxílio financeiro da CAPES.

– São Paulo, SP –

2016

*"Talvez não tenha conseguido fazer o
melhor, mas lutei para que o melhor fosse
feito. Não sou o que deveria ser, mas
Graças a Deus, não sou o que era antes."*

Marthin Luther King

*Aos meus queridos pais,
Mauro (in memoriam) e Élia,
ao meu irmão Jefferson e ao
amor da minha vida Amanda.*

Agradecimentos

Agradecer é admitir que houve um momento em que se precisou de alguém. Por isso, em primeiro lugar, agradeço a Deus por sempre me dar forças para superar os desafios impostos na minha vida, podendo chegar a mais esta conquista.

Agradeço também aos meus pais por terem muitas vezes abdicado de seus sonhos para que eu pudesse realizar os meus, muito obrigada por terem me apoiado em todas as fases da minha vida. Pai, pena ter te perdido antes de terminar o doutorado, você que sempre me apoiou, você que me acordava para ir à escola, você que fez minha mudança para São Paulo para que eu pudesse iniciar esta etapa. Espero que de onde estiver, sinta minha gratidão e saiba que você foi essencial para que eu conseguisse alcançar meu objetivo. Obrigada mãe, por ser meu porto seguro e por me apoiar nas minhas decisões.

Muito obrigada Jefferson, por ser esse irmão tão companheiro, compreensivo e paciente. Sei que em muitos momentos de estresse você me suportou. Quero agradecer também a todos meus familiares e à minha sogra pelas palavras de apoio, incentivo e pelo carinho de sempre.

Também agradeço profundamente à inspiração da minha vida, muito obrigada Amanda Rusiska Piovezani, pelo companheirismo, apoio, pela amizade e principalmente pela paciência de sempre. Saiba que amo muito você, esta conquista também te pertence.

Um agradecimento especial aos meus amigos, que de forma direta ou indireta me ajudaram e apoiaram nesta etapa, em especial à Francielle, Elton, Lígia e Ariadne.

Aos membros do GaTE lab, pela convivência e apoio diário, tanto para os membros atuais, quanto aos membros que já deixaram o grupo. Em especial à Tatiana, Paula, Geovani, Dani Quintanilha, Érika e Sára, pela amizade.

Um agradecimento especial à minha orientadora, Marie-Anne Van Sluys, por ter me acolhido em seu laboratório, sem me conhecer previamente, por acreditar em mim, pelas conversas e principalmente pelo conhecimento compartilhado.

Agradeço também ao meu coorientador, Fabrício Martins Lopes, pela paciência e prontidão em me auxiliar com os conceitos de ciências exatas. Muito obrigada também por me acalmar em vários momentos e por acreditar que seria possível.

Não poderia deixar de agradecer também à Mariangela Hungria, por ter me fornecido os dados para a execução deste trabalho, antes mesmo de serem publicados. Além disso, muito obrigada por

ter despertado em mim a paixão pela ciência.

Aos professores do programa Interunidades de Pós-Graduação em Bioinformática, que contribuíram para minha formação, e a todos demais membros deste programa.

Por último, mas não menos importante, agradeço à Universidade de São Paulo pelo espaço físico concedido para a realização das pesquisas, à FAPESP pelo auxílio financeiro dado ao laboratório GaTE lab, e à CAPES pela bolsa de doutorado concedida durante a execução desta tese.

BARROS-CARVALHO, GESIELE ALMEIDA. Análise computacional dos genomas de duas estirpes brasileiras de *Bradyrhizobium* de importância econômica. 2016. Tese (Doutorado em Bioinformática) - Universidade de São Paulo, São Paulo, Brasil.

RESUMO

B. diazoefficiens CPAC 7 e *B. japonicum* CPAC 15 são estirpes brasileiras de *Bradyrhizobium* que apresentam grande relevância para o cultivo da soja, pois são capazes de fornecer nitrogênio para a produção desta leguminosa através do processo de fixação biológica de nitrogênio (FBN), uma técnica sustentável e de baixo custo. Por esse motivo, tais bactérias são de grande interesse, e seu estudo contribui na compreensão do processo complexo e orquestrado por um conjunto de genes específicos que culmina no estabelecimento da simbiose. A estirpe CPAC 7 possui maior eficiência em fixar N₂, e a CPAC 15 destaca-se pela sua competitividade. Recentemente, o genoma de cada uma foi sequenciado na tentativa de conhecer seu conteúdo gênico e identificar os fatores genéticos responsáveis pelas diferenças no desempenho simbiótico. Apesar de ter sido encontrado alguns rearranjos, os genomas mostraram-se sintênicos na sua maioria. Entretanto, o fato de haver muitas transposases ao redor dos genes, principalmente na ilha simbiótica, e devido a presença de muitos genes hipotéticos, representando uma limitação no conhecimento, nos motivou a realizar o presente estudo, onde exploramos estes dois genomas. Portanto, os objetivos deste estudo foram de definir a população de elementos de transposição (TEs) que compõe estes genomas, avaliar se os elementos completos podem estar impactando os genes de alguma forma; explorar as proteínas hipotéticas, tentando identificar novas funções que possam estar associadas com a interação soja-*Bradyrhizobium* e apontá-las para estudos experimentais futuros; e ainda explorar os genes exclusivos das regiões atípicas dos genomas, sendo que para isso, nós também desenvolvemos uma nova metodologia, baseada na máxima entropia (ME), que pode ser utilizada em novos estudos genômicos a partir da simples sequência nucleotídica. Todas as análises deste estudo foram realizadas *in silico*. Estudando os TEs, identificamos 33 novas sequências de inserção, sendo que algumas destacaram-se por terem potencial impacto nos genes associados com a simbiose destas bactérias, como *nopAN*, *nopAG*, *rhcU*, *modC* e *hypB*. Explorar as proteínas hipotéticas nos permitiu reduzir a porcentagem de hipotéticas dos genomas. Adicionamos novas informações à 1.204 proteínas, das quais muitas apresentaram similaridade com proteínas comprovadamente associadas com a interação planta-bactéria, em condições de simbiose e/ou patogenicidade, como proteínas envolvidas na motilidade e adesão celular, fatores de virulência, proteínas secretoras e efetoras, entre outras. Além disso, a metodologia ME, desenvolvida neste estudo com o intuito de direcionar análises genômicas para regiões atípicas, quando comparada com outras ferramentas existentes, mostrou-se superior em termos de eficiência e tempo de execução computacional. Nas regiões genômicas apontadas pela ME nos dois genomas de interesse, identificamos 269 genes

exclusivos de CPAC 7 e 368 de CPAC 15, sendo que destacamos aqueles com potencial relação com as diferenças simbióticas das estirpes, como o gene *fixW*, *noeE*, *rtxA* e *nex18*. Assim, os resultados obtidos neste trabalho vêm expandir nosso conhecimento sobre os genomas destas estirpes. Destacando ainda, importantes diferenças que podem estar associadas com a habilidade simbiótica de cada bactéria.

Palavras-chave: Análise de sequências, Bioinformática, Fixação de nitrogênio, Genômica comparativa de procariotos, Interação planta-bactéria, Máxima entropia, Proteínas hipotéticas, Regiões atípicas, Sequências de inserção.

BARROS-CARVALHO, GESIELE ALMEIDA. Computational analysis of genomes of two Brazilian *Bradyrhizobium* strains of economic importance. 2016. Thesis (Bioinformatics) - University of São Paulo, São Paulo, Brazil.

ABSTRACT

B. diazoefficiens CPAC 7 and *B. japonicum* CPAC 15 are Brazilian *Bradyrhizobium* strains of great importance for soybean cultivation, since when in a symbiotic state they provide nitrogen for the crop through the biological nitrogen fixation process (BNF), a sustainable technique and low cost. For this reason, such bacteria represent great interest and have been widely studied, once the symbiotic establishment is a complex process and orchestrated by a specific set of genes. The CPAC 7 strain has a higher efficiency to fix N₂, while CPAC 15 stands out for its competitiveness. Recently, their genomes were sequenced in an attempt to gain knowledge about their gene content and to identify the genetic factors responsible for differences in their symbiotic performance. Despite having identified some rearrangements, the majority of genomes showed syntenic. However, the fact that there are many transposases around the genes, especially in symbiotic island, and due to the presence of many hypothetical genes, representing a limitation on knowledge, motivated us to conduct this study, which explored these two important genomes. Therefore, the objectives of this study were to define the population of transposable elements (TEs) present in these genomes and to verify whether such TEs could be impacting the genes somehow; to study the hypothetical proteins, trying to identify new features that may be associated with the soybean-*Bradyrhizobium* interaction and point them for future experimental studies; and to explore the exclusive genes from atypical regions of both genomes, and for that, we have also developed a new methodology, based on maximum entropy (ME), which can be used in new genomic studies. All analyzes in this study were performed *in silico*. Studying the TEs, we identified 33 new insertion sequences, and some stood out for having potential impact on genes associated with the symbiosis of these bacteria, such as *nopAN*, *nopAG*, *rhcU*, *modC* and *hypB*. As a consequence of improving the annotation of hypothetical proteins we were able to reduce the hypothetical percentage. Among these, we add new information to 1,204 proteins, many of which had similarity to proteins with involvement in the plant-bacteria interaction, in symbiosis and/or pathogenicity conditions, such as proteins involved in cell motility and adhesion, virulence factors, secretion proteins, effectors, among others. Moreover, the ME methodology developed in this study to direct genomic analysis to atypical regions, compared with other existing tools, it was superior in efficiency and execution time. In the genomic regions identified by the ME in both *Bradyrhizobium* genomes, we identified 269 exclusive genes of CPAC 7 and 368 of CPAC 15, we highlighted those with potential involvement with symbiotic differences of strains, as *fixW*, *noeE*, *rtxA* and *nex18*. Thus, the results obtained in this study come to expand our knowledge about the genomes of these important bacteria.

Finally, differences were identified as potential targets to be associated with the symbiotic ability of each strain to be further studied.

Key-words: Atypical regions, Bioinformatics, Comparative genomics of prokaryotes, Hypothetical proteins, Insertion sequences, Maximum entropy, Nitrogen fixation, Plant-bacteria interaction, Sequence analysis.

Lista de Figuras

1	Etapas do estabelecimento simbiótico soja- <i>Bradyrhizobium</i> para o processo de fixação biológica de nitrogênio.	5
2	Principais genes de nodulação e fixação de N ₂ presentes na ilha simbiótica de <i>Bradyrhizobium</i>	7
3	Panorama geral do estado de anotação da ilha simbiótica de <i>B. diazoefficiens</i> CPAC 7 e <i>B. japonicum</i> CPAC 15 publicadas após o sequenciamento dos genomas.	10
4	Representação de alguns tipos de impactos que os elementos de transposição podem promover dentro dos genomas.	12
5	Representação de como a contagem de oligonucleotídeos é computada para a análise de máxima entropia.	26
6	Cálculo da máxima entropia para as regiões genômicas.	27
7	Procedimento de análise da metodologia de máxima entropia para discriminação de regiões atípicas de genomas bacterianos.	28
8	Distribuição dos elementos ISs ao longo dos genomas de <i>B. diazoefficiens</i> CPAC 7 e <i>B. japonicum</i> CPAC 15.	33
9	Famílias dos ISs completos presentes nos genomas de <i>B. diazoefficiens</i> CPAC 7 e <i>B. japonicum</i> CPAC 15.	34
10	Contexto genômico dos ISs completos identificados em <i>B. diazoefficiens</i> CPAC 7 e <i>B. japonicum</i> CPAC 15.	38
11	Representação de genes importantes para a simbiose e localizados na ilha simbiótica que podem estar sendo afetados por ISs completos.	42
12	Comparação dos principais eventos de inserção identificados em outros genomas de <i>Bradyrhizobium</i>	44
13	Total de proteínas hipotéticas compartilhadas e exclusivas de <i>B. diazoefficiens</i> CPAC 7 e <i>B. japonicum</i> CPAC 15.	48
14	Diagrama de Venn mostrando que algumas proteínas anotadas tiveram similaridade com proteínas envolvidas na interação planta-bactéria.	49

15	Categorização funcional das 1.204 proteínas que tiveram função atualizada.	51
16	Alinhamento da corismato mutase secretada (AroQ*) de <i>Erwinia herbicola</i> e <i>Salmonella enterica</i> com corismato mutases exclusivas de <i>Bradyrhizobium japonicum</i> . . .	53
17	Estrutura de três tetratricopeptídeos identificados em <i>B. diazoefficiens</i> CPAC 7 e <i>B. japonicum</i> CPAC 15 que tiveram seus homólogos expressos em condição de simbiose	55
18	Proteínas com funções pertencentes à categoria processos celulares.	57
19	Avaliação dos parâmetros (tamanho das regiões e tamanho dos oligonucleotídeos) a serem utilizados na identificação de regiões atípicas de genomas bacterianos baseado na máxima entropia.	64
20	Comparação da performance das ferramentas utilizadas para detecção de regiões atípicas em genomas bacterianos.	67
21	Máxima de entropia aplicada ao genoma de <i>Xanthomonas campestris</i> pv. <i>campestris</i> ATCC 33913.	68
22	Perfil de máxima entropia aplicada aos genomas de <i>Bradyrhizobium diazoefficiens</i> CPAC 7 e <i>Bradyrhizobium japonicum</i> CPAC 15.	70
23	Ilha simbiótica de <i>B. diazoefficiens</i> CPAC 7 e <i>B. japonicum</i> CPAC 15 com as principais contribuições obtidas em comparação com o conhecimento anterior ao presente estudo	78
24	Representação de alguns aspectos genéticos específicos de cada estirpe de <i>Bradyrhizobium</i> identificados neste estudo que podem estar relacionados com suas diferenças fenotípicas	79

Lista de Tabelas

1	Comparação do número de cópias de cada sequência de inserção completa anotada nos genomas.	36
2	Genes impactados pelos ISs completos encontrados nas estirpes brasileiras de <i>Bradyrhizobium</i> , sendo que alguns homólogos destes genes são diferencialmente expressos em condição de simbiose em <i>B. diazoefficiens</i> USDA 110.	39
3	Proteínas apontadas como possíveis alvos de estudos experimentais, com potencial envolvimento no processo de interação planta-bactéria.	60
4	Características dos métodos disponíveis para detecção de regiões atípicas em genomas bacterianos comparados com a máxima entropia.	66
5	Genes de destaque selecionados pela máxima entropia nos genomas das estirpes brasileiras de <i>Bradyrhizobium</i>	72

Sumário

Resumo	vi
Abstract	viii
Lista de Figuras	x
Lista de Tabelas	xii
1 Introdução	1
2 Fundamentação Teórica	3
2.1 Interação simbiótica soja- <i>Bradyrhizobium</i>	3
2.2 Estirpes brasileiras de <i>Bradyrhizobium</i> : CPAC 7 e CPAC 15	8
2.3 Elementos transponíveis em procariotos e sua importância	10
2.4 A relevância de estudar proteínas hipotéticas	13
2.5 Explorando regiões atípicas dos genomas bacterianos	14
3 Objetivos	18
3.1 Objetivo Geral	18
3.2 Objetivos Específicos	18
4 Materiais e Métodos	19
4.1 Obtenção dos genomas	19
4.2 Análise dos elementos transponíveis	19
4.2.1 Identificação e anotação dos elementos de transposição	19
4.2.2 Contexto genômico dos elementos ISs completos	21
4.2.3 Identificação de homólogos dos genes impactados expressos em condição de simbiose	22
4.3 Explorando o conjunto de proteínas hipotéticas	22
4.3.1 Seleção das proteínas hipotéticas e adição de novas informações	22

4.3.2	Busca por proteínas similares envolvidas nos processos de interação planta-bactéria e FBN	23
4.4	Desenvolvimento da estratégia de máxima entropia para explorar regiões atípicas . .	24
4.4.1	Processamento das sequências	25
4.4.2	Calculando a máxima entropia	26
4.4.3	Avaliação da metodologia proposta	29
4.4.4	Máxima entropia utilizada na seleção de genes exclusivos de <i>B. diazoefficiens</i> CPAC 7 e <i>B. japonicum</i> CPAC 15	30
5	Resultados e Discussão	32
5.1	População de elementos transponíveis presentes nas estirpes brasileiras de <i>Bradyrhizobium</i>	32
5.1.1	Impacto local dos ISs completos nos genomas das estirpes CPAC 7 e CPAC 15	37
5.2	Descrição geral da análise das proteínas hipotéticas	47
5.2.1	Categorização funcional das proteínas que tiveram anotação atualizada . . .	50
5.2.2	Processos celulares: proteínas com potencial envolvimento no processo de interação planta-bactéria	57
5.3	Discriminando regiões atípicas com a máxima entropia	63
5.3.1	Máxima entropia aplicada aos genomas de <i>B. diazoefficiens</i> CPAC 7 e <i>B. japonicum</i> CPAC 15	69
6	Conclusões	76
A	Apêndice: Elementos ISs completos identificados em <i>B. diazoefficiens</i> CPAC 7	81
B	Apêndice: Elementos ISs completos identificados em <i>B. japonicum</i> CPAC 15	85
C	Apêndice: Gene <i>modC</i> compartilhando sequência com o ISBd2	89
D	Apêndice: Proteínas hipotéticas que tiveram novas informações adicionadas	90
E	Apêndice: Análise gráfica do resultado da máxima entropia aplicada a um genoma de <i>Xanthomonas</i>, utilizando diferentes tamanhos de regiões genômicas	119
F	Apêndice: Porcentagem de cobertura das regiões genômicas apontadas pela máxima entropia de CPAC 7 no genoma de CPAC 15	121
G	Apêndice: Porcentagem de cobertura das regiões genômicas apontadas pela máxima entropia de CPAC 15 no genoma de CPAC 7	122

H	Apêndice: Regiões genômicas de <i>B. diazoefficiens</i> CPAC 7 e de <i>B. japonicum</i> CPAC 15 apontadas pela máxima entropia a serem exploradas	123
I	Apêndice: Proteínas exclusivas de cada estirpe brasileira de <i>Bradyrhizobium</i> extraídas das regiões apontadas pela máxima entropia	124
J	Apêndice: Artigos científicos produzidos a partir dos resultados obtidos nesta tese	135
	Referências Bibliográficas	136

1

Introdução

Bactérias fixadoras de nitrogênio (N) são de extrema importância para o solo e para as plantas, pois são capazes de converter o nitrogênio atmosférico (N_2) em uma forma assimilável para as plantas (Amônia - NH_3). Este processo é conhecido como fixação biológica de nitrogênio (FBN), o qual, organismos eucariotos são incapazes de realizar. A FBN é especialmente importante para a cultura da soja, uma vez que seus grãos apresentam elevado teor proteico e portanto, necessita de grandes quantidades de N para seu cultivo. E com o uso da FBN é possível dispensarmos o uso dos fertilizantes químicos nitrogenados, trazendo maior economia para o agricultor, sem agredir o meio ambiente.

As bactérias *Bradyrhizobium diazoefficiens* CPAC 7 e *Bradyrhizobium japonicum* CPAC 15, conhecidas popularmente como rizóbios, são estirpes brasileiras autorizadas pelo Ministério da Agricultura, Pecuária e Abastecimento (MAPA) para a produção de inoculantes comerciais no Brasil, pois são capazes de estabelecer simbiose com a soja de forma eficiente, suprimindo o N que esta leguminosa precisa.

Por esta razão, tais bactérias vêm sendo bastante estudadas, sendo que *B. diazoefficiens* CPAC 7 possui maior eficiência em fixar N_2 e *B. japonicum* CPAC 15 apresenta maior competitividade. Na tentativa de tentar compreender os fatores genéticos responsáveis pelas diferenças na performance simbiótica, seus genomas foram sequenciados e comparados. Foram então observados alguns eventos de rearranjo, sendo que a maior parte dos genomas é sintênica. Em CPAC 15, a maioria dos genes

estão associados com transporte de nutrientes e metabolismo secundário, bactéria reconhecida pela sua competitividade. Em contrapartida, CPAC 7 possui muitos genes relacionados ao metabolismo de aminoácidos, estirpe conhecida pela sua eficiência em fixar N_2 .

Entretanto, na análise comparativa, foi verificada a presença de muitas transposases nestes genomas, principalmente ao redor dos genes da ilha simbiótica, região do genoma onde estão concentrados a maioria dos genes responsáveis pela nodulação e fixação de N_2 . Além disso, foi relatado que entre os genes presentes nos genomas, cerca de 50% correspondiam a genes hipotéticos, o que representa uma limitação no conhecimento.

Diante deste cenário, o presente estudo objetivou identificar e anotar os elementos transponíveis presente em ambos genomas, e verificar o impacto local que os elementos completos podem estar causando, com enfoque para aqueles genes com potencial envolvimento na simbiose.

Ademais, realizamos uma atualização da anotação dos genes hipotéticos e adicionamos novas informações à vários deles. Sendo que para muitos, identificamos similaridade a genes envolvidos em mecanismos de interação planta-bactéria e/ou na FBN.

Aditivamente, sabendo que genomas bacterianos são muito dinâmicos, capazes de transferir e adquirir genes que podem conferir-lhes novas características, e que tais regiões genômicas adquiridas podem apresentar características atípicas do restante do genoma, nós desenvolvemos uma nova metodologia com base na máxima entropia para nos direcionar rapidamente para tais regiões. Assim, as exploramos buscando por características distintivas de cada estirpe, focando em genes que possam estar associados com a capacidade simbiótica.

Portanto, o presente estudo analisa computacionalmente os genomas dessas duas bactérias sob três abordagens distintas, a análise de impacto local dos elementos de transposição, a busca por novas funções dentre os genes hipotéticos e a identificação de genes exclusivos presentes nas regiões atípicas, identificadas por uma nova metodologia, também desenvolvida neste trabalho. Todas as análises foram focadas visando aumentar o conhecimento sobre tais genomas e apontar potenciais alvos de estudos experimentais futuros, os quais podem estar relacionados com a interação soja-*Bradyrhizobium*.

Fundamentação Teórica

2.1 Interação simbiótica soja-*Bradyrhizobium*

A SIMBIOSE soja-*Bradyrhizobium* corresponde a uma interação planta-microrganismo, em que tais bactérias são capazes de disponibilizar nitrogênio (N) para a planta através do processo de fixação biológica de nitrogênio (FBN), recebendo em troca fontes de carbono (C) [Hungria *et al.*, 2006a; Ferguson *et al.*, 2010; Wagner, 2011]. Esse processo representa especial importância para o Brasil, que atualmente é o segundo maior produtor mundial de soja, sendo a FBN um processo de baixo custo, eficiente e sustentável [Hungria *et al.*, 2006a; Peoples *et al.*, 2009; CONAB, 2016].

Os organismos procariotos capazes de fixar N_2 são chamados de diazotróficos, os quais possuem um complexo enzimático, denominado nitrogenase, capaz de clivar a tripla ligação do N_2 [Ferguson *et al.*, 2010]. As bactérias capazes de estabelecer uma simbiose com a soja, de modo eficiente, são as espécies *Bradyrhizobium diazoefficiens*, *Bradyrhizobium japonicum* e *Bradyrhizobium elkanii* [Hungria *et al.*, 2006a,b; Delamuta *et al.*, 2013].

A simbiose envolve uma coordenada troca de sinais moleculares altamente específicos [Prell *et al.*, 2009]. Quando estabelecida, a simbiose pode ser facilmente identificada nas leguminosas pela presença de nódulos nas raízes, onde a fixação de N_2 acontece [Stougaard, 2000; Oldroyd & Downie, 2008].

Inicialmente, a planta hospedeira sintetiza e libera metabólitos secundários, dentre eles flavonoi-

des específicos, que atraem as bactérias fixadoras de N_2 por quimiotaxia (Figura 1). Os flavonoides com forte ação indutora sobre *Bradyrhizobium* liberados pela soja são a daidzeína e a genisteína [Subramanian *et al.*, 2006, 2007; Dolatabadian *et al.*, 2012]. Esses compostos ativam a expressão de genes da nodulação (*nodVW* e *nodD*). A proteína NodD ativa os genes *nodABC* que codificam moléculas de lipoquitino-oligossacarídeos, conhecidas como fatores de nodulação ou fatores Nod [Loh & Stacey, 2003; Sugiyama *et al.*, 2008] (Figura 1).

A percepção das bactérias e dos fatores Nod pelas células vegetais se dá através de receptores presentes na superfície celular das raízes (Ex.: LysM), os quais desencadeiam uma cascata de sinalização na planta levando à expressão das nodulinas, que por sua vez provocam uma série de alterações morfológicas e fisiológicas na planta para a formação dos nódulos, como o encurvamento dos pelos radiculares, degradação da parede celular, formação dos cordões de infecção (estruturas tubulares compostas por células da parede celular vegetal que conduzem os rizóbios para o interior das células do córtex da raiz), alteração no ciclo celular, reorganização do citoesqueleto, indução da divisão celular, alterações hormonais, entre outros [Blancaflor *et al.*, 2006; Oldroyd *et al.*, 2011; Dolatabadian *et al.*, 2012]. Além disso, as bactérias específicas suprimem as reações de defesa do hospedeiro, por exemplo, pela presença de compostos específicos como os exopolissacarídeos (EPS), lipopolissacarídeos (LPS) e β -glicanos cíclicos, ou derivados dessas moléculas [Mithöfer, 2002; Ferguson *et al.*, 2010; Oldroyd *et al.*, 2011] (Figura 1).

Ao atingirem o córtex radicular, as bactérias são liberadas no citoplasma da célula hospedeira. Nesse momento, através de um processo semelhante à endocitose, as bactérias são envolvidas por uma membrana derivada da planta, formando o simbiossomo. A bactéria envolvida por tal membrana continua se dividindo antes de se diferenciar em bacteroide e dar início à fixação de N_2 no interior dos nódulos [Ferguson *et al.*, 2010; Oldroyd *et al.*, 2011] (Figura 1).

Figura 1: Etapas do estabelecimento simbiótico soja-*Bradyrhizobium* para o processo de fixação biológica de nitrogênio. **1-** A planta libera flavonoides que atraem os rizóbios por quimiotaxia e induzem a expressão dos genes, presentes na ilha simbiótica (genes *nod*, *nol* e *noe*), que codificam os fatores Nod. Além desses genes, outros genes também têm a expressão induzida, os quais participam da interação com a planta, como os genes do sistema de secreção, efetores, genes de adesão, entre outros. Os fatores Nod ao serem percebidos pela planta, desencadeiam uma cascata de sinalização que levam a uma série de modificações morfo-fisiológicas na planta; **2-** Os rizóbios aderem-se aos pelos radiculares, que por sua vez sofrem um processo de curvatura; **3-** Os rizóbios invadem as células da raiz, formando o cordão de infecção até atingirem o córtex; **4-** Agora no córtex, inicia-se a divisão celular, levando à formação dos nódulos; **5-** Dentro do nódulo formado, as bactérias continuam se multiplicando e diferenciam-se em bacteroides; **6-** Os bacteroides estabelecidos nos nódulos estão aptos a fixar o N_2 , sendo que a cor avermelhada no interior do nódulo indica que o mesmo está ativo. (2-CR: dois componentes regulatórios, LPS: lipopolissacarídeos, EPS: exopolissacarídeos, Nop: Nodulation Outer Protein, NSP: Nodulation Signaling Pathway.)

Para que ocorra a FBN, deve haver um controle da presença de oxigênio (O_2) nos nódulos, exercido principalmente pela leghemoglobina (proteína produzida pela planta que possui uma alta afinidade por O_2) e pela membrana peribacteroide (barreira de difusão de O_2 no nódulo), os quais protegem o complexo enzimático da nitrogenase, que pode ser inativado pelo O_2 de forma irreversível, e garantem que os bacteroides recebam o O_2 necessário para sua respiração [Mylona *et al.*, 1995; Dixon & Kahn, 2004; Rees *et al.*, 2005].

A FBN propriamente dita, tem início com a ferredoxina reduzida transferindo um elétron para a unidade Fe-proteína da nitrogenase, que por sua vez, doa tal elétron recebido para a MoFe-proteína, unidade proteica responsável pelo acúmulo de elétrons. Depois de oito transferências e com os oito elétrons acumulados, a MoFe-proteína reduz o N_2 em amônia (NH_3). Cada molécula de N_2 reduzida resulta em um gasto energético de dezesseis moléculas de adenosina trifostato (ATP), pois na transferência de cada elétron são consumidos dois ATPs [Rees *et al.*, 2005].

A NH_3 quando em contato com o citoplasma dos bacteroides, é convertida rapidamente em íon amônio (NH_4^+), que por conseguinte, inibe a fixação de N_2 por não poder acumular-se no interior dos bacteroides. Assim, logo após formado, o NH_4^+ é exportado para o citoplasma da célula radicular onde é imediatamente convertido em aminoácidos, pois também é prejudicial para a célula vegetal. As enzimas essenciais para essa conversão são a glutamina sintetase (GS) e a glutamato sintase (GOGAT), que originam a glutamina e o glutamato, respectivamente. Esses levam à formação de asparagina ou ureídeos (alantoína e ácido alantóico), as formas de transporte do nitrogênio para toda a planta. A soja, frequentemente, transporta ureídeos como fonte de nitrogênio ao longo das células vegetais [Mylona *et al.*, 1995; Dixon & Kahn, 2004; Rees *et al.*, 2005].

Como é possível observar, a interação simbiótica soja-*Bradyrhizobium* passa basicamente pelo processo de infecção, nodulação e fixação de N_2 , os quais contam com a presença de genes específicos presentes nas bactérias. Dentre esses, temos os genes *exo* (exopolissacarídeos - EPS), *lps* (lipopolissacarídeos - LPS) e *ndv* (desenvolvimento nodular) envolvidos na formação da superfície celular bacteriana e na alteração da nodulação. Os genes *nod*, *nol* e *noe* correspondem aos genes da nodulação propriamente dita e os genes *hsn* são os genes específicos do hospedeiro. Os genes *nif* estão envolvidos diretamente na FBN e os genes *fix* correspondem as moléculas que regulam o processo

de fixação [Hennecke, 1990; Göttfert *et al.*, 2001; Loh & Stacey, 2003].

Além desses genes específicos, muitos outros também participam do estabelecimento da simbiose, como os genes envolvidos na mobilidade da bactéria, genes que codificam proteínas dos sistemas de secreção, genes envolvidos no metabolismo do hidrogênio, genes que interagem com a parede celular vegetal, genes que codificam proteínas de membrana, efetores, enzimas hidrolíticas, genes que codificam proteínas de adesão, entre outros, os quais contribuem para o sucesso dessa interação [Göttfert *et al.*, 2001; Loh & Stacey, 2003; Soto *et al.*, 2006, 2009].

Em *Bradyrhizobium*, a maioria desses genes está concentrada em uma região do genoma conhecida como ilha simbiótica (Figura 2). A estrutura dessa ilha foi primeiramente caracterizada em *B. japonicum* USDA 110 por Göttfert *et al.* [2001], bactéria reclassificada recentemente como *B. diazoefficiens* USDA 110 [Delamuta *et al.*, 2013]. Posteriormente, essa mesma bactéria foi a primeira desse gênero a ter o genoma completamente sequenciado [Kaneko *et al.*, 2002]. Alguns anos mais tarde foi a vez da estirpe *B. japonicum* USDA 6 a ter seu genoma completo sequenciado [Kaneko *et al.*, 2011]. Ambas são estirpes americanas, sendo que a USDA 110 é referência nos estudos de interação soja-*Bradyrhizobium* [Pessi *et al.*, 2007; Delmotte *et al.*, 2010; Libault *et al.*, 2010; Franck *et al.*, 2014].

Figura 2: Principais genes de nodulação e fixação de N_2 presentes na ilha simbiótica de *Bradyrhizobium*. Além destes, temos os demais genes presentes na ilha simbiótica, os quais envolvem genes específicos do hospedeiro, genes dos sistemas de secreção, proteases, genes hipotéticos, transposases, entre outros. (Estrutura da ilha simbiótica baseada em Göttfert ET AL. [2001].)

Além dos fatores genéticos, condições ambientais também podem interferir no processo de simbiose, tais como salinidade, acidez do solo, temperatura, deficiência de nutrientes, seca, entre outros [Hungria & Franco, 1993; Hungria & Vargas, 2000; Ferguson *et al.*, 2013].

2.2 Estirpes brasileiras de *Bradyrhizobium*: CPAC 7 e CPAC 15

O gênero *Bradyrhizobium* compreende bactérias gram-negativas aeróbicas presentes no solo, pertencentes à classe *Alfa-proteobactéria*, sendo algumas espécies capazes de fixar N₂. Tais bactérias têm o formato bacilos e possuem um único flagelo na sua região polar. Bactérias desse gênero possuem taxas de crescimento lento, característica que resultou no nome *Bradyrhizobium*, onde "*bradus*" vem do grego e significa "lento" [Jordan, 1982].

Essas bactérias não estão presentes de forma natural em solos brasileiros, foram introduzidas através da inoculação. Programas de seleção de estirpes (microrganismos de uma mesma espécie com características genéticas e bioquímicas distintas) competitivas e eficientes são os responsáveis por selecionar tais rizóbios [Hungria *et al.*, 1996, 2006a, 2014].

B. diazoefficiens CPAC 7 (SEMIA 5080) e *B. japonicum* CPAC 15 (SEMIA 5079) foram isoladas em Brasília (CPAC: Centro de Pesquisa Agropecuária dos Cerrados) [Hungria *et al.*, 1996; Batista *et al.*, 2007]. A estirpe CPAC 7 foi selecionada pela sua elevada eficiência em fixar N₂ [Peres *et al.*, 1993]. Já a estirpe CPAC 15 possui alta capacidade de nodulação, além de ser a mais competitiva dentre as autorizadas para a produção de inoculantes [Mendes *et al.*, 2004; Batista *et al.*, 2007].

Além da competitividade e capacidade na fixação, elas possuem outras diferenças morfológicas, fisiológicas e genéticas, como por exemplo os antibióticos ao qual são resistentes, enquanto CPAC 15 é resistente a estreptomicina, carbenicilina e cloranfenicol, CPAC 7 é sensível [Nishi *et al.*, 1996; Boddey & Hungria, 1997]. Além disso, ao contrário da CPAC 7, CPAC 15 desencadeia sintomas de clorose foliar em algumas variedades de soja pela produção de rizobiotoxina, entre outros [Hungria *et al.*, 1996; Nishi *et al.*, 1996; Boddey & Hungria, 1997].

Sabendo da importância da FBN para o Brasil e na tentativa de identificar novos fatores que possam estar contribuindo para as diferenças relacionadas à simbiose das estirpes brasileiras, *B. dia-*

zoefficiens CPAC 7 e *B. japonicum* CPAC 15 também tiveram seus genomas sequenciados e uma análise comparativa realizada. Ambas apresentam cromossomo circular único e ausência de plasmídeo, sendo que a estirpe CPAC 7 possui cerca de 9.085.545 pares de bases (pb) disponível em treze *contigs* e CPAC 15 possui 9.582.287 pb em um *scaffold* [Siqueira *et al.*, 2014].

Tratando-se mais especificamente da ilha simbiótica, que está completa nos dois genomas, seu tamanho em CPAC 7 e CPAC 15 é de 688.007 pb e 700.213 pb, respectivamente [Siqueira *et al.*, 2014]. Nesses genomas, as ilhas estão localizadas em diferentes regiões do cromossomo, mas segundo Siqueira *et al.* [2014], encontram-se bastante conservadas entre as duas estirpes, sendo que os principais genes conhecidos envolvidos na nodulação (Figura 2) encontram-se presentes em ambas.

Apesar desses genomas serem sintênicos na sua maioria, apresentam alguns rearranjos, inclusive na ilha simbiótica. Além disso, algumas ilhas genômicas foram identificadas nos genomas [Siqueira *et al.*, 2014]. Comparando as categorias funcionais dos genes identificados, em CPAC 7, a maioria está envolvida com o metabolismo de aminoácidos, o que pode estar relacionado com sua eficiência na fixação de N₂. Em contrapartida, a maioria dos genes de CPAC 15 está associado com o metabolismo secundário e transporte de nutrientes, podendo estar ligado à sua competitividade (capacidade de superar outras bactérias do solo por nutrientes, e na rizosfera por superar outras bactérias da mesma espécie pela raiz do hospedeiro para o início da simbiose) [Loh & Stacey, 2003; Siqueira *et al.*, 2014].

Alguns resultados obtidos com o sequenciamento e análise comparativa desses genomas nos despertaram o interesse em explorá-los. Primeiramente por terem identificado muitas transposases distribuídas nos cromossomos, especialmente ao redor dos genes da ilha simbiótica [Siqueira *et al.*, 2014], e sabendo da importância que os elementos de transposição têm para os organismos, ficamos interessados em entender se esses elementos podem estar impactando de alguma forma o processo de simbiose.

Em segundo lugar, o fato dos autores terem relatado que cerca de 50% das proteínas identificadas nos genomas não tiveram função atribuída, sendo assim definidas como hipotéticas [Siqueira *et al.*, 2014]. Isso nos mostra que o conhecimento sobre esses genomas ainda é limitado e a busca por novas funções torna-se fundamental. Na Figura 3, apresentamos o estado da anotação da ilha simbiótica publicadas de CPAC 7 e CPAC 15. Aqui destacamos 3 tipos de genes identificados, genes com função

definida, genes hipotéticos e genes que codificam transposases. Como podemos observar, o conteúdo gênico é variável dentro da ilha, com isso nosso estudo vem melhorar a compreensão do conteúdo gênico do genoma, com interesse especial na ilha simbiótica e ainda tentar entender se os elementos transponíveis podem interferir de alguma forma na simbiose.

Figura 3: Panorama geral do estado de anotação da ilha simbiótica de *B. diazoefficiens* CPAC 7 e *B. japonicum* CPAC 15, após o sequenciamento dos genomas.

Além disso, também consideramos que genomas de procariotos são bastante dinâmicos, podendo passar por modificações (aquisições ou perdas de genes) durante os processos evolutivos [Kunin & Ouzounis, 2003; Darmon & Leach, 2014]. Dessa forma, visamos explorar tais regiões genômicas, pois podem nos fornecer informações importantes, que ainda não foram destacadas na análise comparativa tradicional.

2.3 Elementos transponíveis em procariotos e sua importância

Elementos transponíveis (TEs, do inglês, *Transposable Elements*) são sequências genéticas móveis capazes de se mover dentro do genoma e estão presentes na maioria dos organismos [Mahillon & Chandler, 1998; Casacuberta & González, 2013; Siguier *et al.*, 2014]. Dentre os elementos que compõem o mobiloma, conjunto de elementos genéticos móveis de um genoma, em procariotos, estão os transposons e os elementos ISs, sendo esses últimos predominantes [Siguier *et al.*, 2006a, 2014].

Elementos ISs são sequências de DNA (< 3,5 kb) com uma organização genética simples, pois codificam apenas as proteínas necessárias para sua transposição, carregando, na maioria das vezes e dependendo da família, apenas uma fase aberta de leitura (ORF, do inglês, *Open Reading Frame*), correspondente a transposase (Tpase). Além disso, a ORF é, geralmente, flanqueada por curtas sequências de repetições invertidas (IRs, do inglês, *Inverted Repeats*), com tamanho < 40 pb, as quais são reconhecidas pela Tpase no momento da transposição [Cerveau *et al.*, 2011; Siguier *et al.*, 2014]. Ainda, ao se inserirem em um novo sítio do genoma, podem gerar repetições diretas (DRs, do inglês, *Direct repeats*), que variam entre 2 e 14 pb, dependendo da família [Cerveau *et al.*, 2011; Siguier *et al.*, 2014].

Atualmente, a classificação desses elementos está distribuída em 26 famílias [Siguier *et al.*, 2014]. Cada elemento é classificado, principalmente, com base na sequência primária da Tpase [Mahillon & Chandler, 1998; Cerveau *et al.*, 2011; Siguier *et al.*, 2014]. O domínio catalítico da Tpase define à qual família pertence, sendo que o domínio predominante nas proteínas até agora identificadas é o DDE (*Asp-Asp-Glu*). Portanto, uma família de IS é definida como um grupo de ISs com Tases relacionadas e repetições invertidas parecidas [Siguier *et al.*, 2006a].

Os transposons bacterianos (> 3 kb) distinguem-se dos ISs por carregarem genes adicionais em sua estrutura, que não estão envolvidos com sua transposição, como por exemplo, genes de resistência a antibióticos. Eles ainda podem ser divididos em transposon simples e composto, o Tn simples possui suas ORFs flanqueadas por IRs, já o Tn composto contém dois ISs da mesma família, em fitas opostas, nas suas terminações [Siguier *et al.*, 2006a, 2014].

Os TEs são importantes moduladores dos genomas, sendo muitas vezes considerados agentes da evolução, capazes de influenciar na adaptação do organismo [Casacuberta & González, 2013]. Os tipos de impactos mediados pelos TEs podem ser divididos em dois: impacto global e impacto local.

Impacto global trata-se de alterações causadas na arquitetura do cromossomo. Dentro deste, temos a variação no tamanho dos replicons, conforme mostrado por Touchon & Rocha [2007], o tamanho do genoma está diretamente relacionado com o número de cópias de ISs. Também são capazes de promover vários tipos de rearranjos: deleção, inversão e junção de replicons. Tais rearranjos podem levar à grupos de genes com funções especializadas como, resistência a antibióticos, virulência, sim-

biose, entre outros. Além disso, facilitam a evolução de novas combinações de genes [Zinser *et al.*, 2003; Ling & Cordaux, 2010; Cerveau *et al.*, 2011]. Na Figura 4-a, apresentamos um exemplo de uma inversão de genes que pode ser promovida pela mobilização dos elementos *ISa* e *ISb*.

Figura 4: Representação de alguns tipos de impactos que os elementos de transposição podem promover dentro dos genomas. **a-** Impacto global: Inversão de genes promovida pela transposição dos elementos *ISa* e *ISb* (isoformas presentes em fitas opostas do DNA); **b-** Impacto local: Elementos *ISs* inseridos na região codificadora do gene e na região regulatória do gene. *IR*: Inverted Repeat.

Os TEs podem ainda interromper regiões codificantes do genoma e alterar a expressão de genes, o que caracteriza um impacto local (Figura 4-b). A inativação de genes devido a inserção de IS na região codificante é bastante observada em genomas bacterianos, conforme mostrado em *Escherichia coli* [Rodriguez *et al.*, 1992] e *Rickettsia peacockii* [Simser *et al.*, 2005]. Ao se inserirem em regiões regulatórias podem alterar a expressão dos genes vizinhos (Figura 4-b). Podendo interromper a sequência promotora inibindo a expressão, ativar ou ainda aumentar a expressão de genes [Zhang & Saier, 2009; Al Safadi *et al.*, 2010; Cerveau *et al.*, 2011].

Apesar da importância desses elementos, não temos na literatura muitos trabalhos realizados para identificar o impacto desses elementos em genomas de *Bradyrhizobium*. Quando o primeiro genoma desse gênero foi sequenciado, *B. diazoefficiens* USDA 110 [Kaneko *et al.*, 2002; Delamuta *et al.*, 2013], foram identificadas muitas transposases nesse genoma, principalmente na ilha simbiótica, entretanto nenhum trabalho voltado para esses elementos foi realizado.

Posteriormente, no genoma de *B. japonicum* USDA 6 também foram identificadas várias transposases, contudo em menor número quando comparada com a estirpe USDA 110 [Kaneko *et al.*, 2011].

Mais recentemente, Iida *et al.* [2015] identificaram que os elementos ISs promoveram rearranjos na ilha simbiótica de *B. diazoefficiens* NK6, estirpe isolada no Japão de extra crescimento lento. Além disso, mostraram que o *ISRj1* interrompeu genes de mobilidade e genes que pode estar comprometendo o metabolismo central dessa bactéria. Os elementos ISs estão distribuídos por todo seu genoma [Iida *et al.*, 2015].

Diante disso, depois que Siqueira *et al.* [2014] identificaram uma alta concentração de transposases na ilha simbiótica das estirpes brasileiras de *Bradyrhizobium*, ficamos interessados em saber se tais elementos poderiam estar impactando os genes presentes nesses genomas.

2.4 A relevância de estudar proteínas hipotéticas

Em um genoma, as proteínas hipotéticas partem das ORFs, que são definidas por um códon iniciador e um códon terminador durante a predição de genes. As proteínas são preditas a partir da sequência de nucleotídeos dessas ORFs, e aquelas onde não foram encontradas homologia de função definida durante o processo de anotação são, geralmente, denominadas como proteínas hipotéticas (*hypothetical protein*), proteínas de função desconhecida (*unknown function*) ou proteínas não-caracterizadas (*uncharacterized protein*) [Lubec *et al.*, 2005; Mazandu & Mulder, 2012], aqui trataremos todas como proteínas hipotéticas (PHs).

Muitas dessas PHs são encontradas em organismos filogeneticamente próximos, mas que não foram caracterizadas funcionalmente ou bioquimicamente, por isso são definidas como hipotéticas conservadas, já outras não apresentam similaridade com PHs depositadas nos bancos de dados [Galperin & Koonin, 2004; Lubec *et al.*, 2005].

Estudar PHs representa extrema importância, visto que, sua presença nos genomas torna o conhecimento, sobre o mesmo, limitado. Voltar a atenção para essas proteínas pode levar à identificação de funções relevantes dependendo do objetivo do estudo, como por exemplo, função com potencial biotecnológico ou alvos para fármacos [Galperin, 2001; Lubec *et al.*, 2005; Shahbaaz *et al.*, 2013].

Por isso, muitos estudos têm sido voltados mais especificamente para essas proteínas, onde estão as chances de novas descobertas importantes. Entre eles temos a análise computacional de uma PH de

Streptococcus mutans associada com a resistência de antibióticos [Nan *et al.*, 2009]. Também temos estudos com PHs de *Mycobacterium tuberculosis*, no qual Doerks *et al.* [2012] adicionaram informações a mais da metade das PHs presentes nesse genoma e Mazandu & Mulder [2012] realizaram uma predição funcional das PHs e concluíram que muitas dessas proteínas estão associadas à habilidade de invadir e de adaptação no hospedeiro. Shahbaaz *et al.* [2013] exploraram as PHs de *Haemophilus influenzae* Rd KW20, apontando ao final da análise potenciais alvos terapêuticos. Nesse mesmo sentido, *Candida dubliniensis* teve várias de suas PHs anotadas funcionalmente, na qual Kumar *et al.* [2014] identificaram proteínas relacionadas com a patogenicidade do microrganismo.

Sabendo da importância que a caracterização de PHs possui, muitos métodos e ferramentas de bioinformática têm sido empregados na tentativa de identificá-las [Lubec *et al.*, 2005]. Sendo que, o método mais comumente usado para inferir função é por homologia de sequência através de alinhamentos, onde o programa mais utilizado é o BLAST (*Basic Local Alignment Search Tool* - Altschul *et al.* [1990, 1997]).

Além disso, a análise de motivos (conjunto de aminoácidos conservados importantes para a função) e domínios proteicos (unidade tridimensional independente e compacta que compõe proteínas) é uma etapa obrigatória na caracterização de PHs, e para isso contamos com bancos de dados e ferramentas para essa classificação, como o CATH [Sillitoe *et al.*, 2013], Interpro [Hunter *et al.*, 2009], Pfam [Finn *et al.*, 2010], ProDom [Bru *et al.*, 2005], PROSITE [Sigrist *et al.*, 2012], entre outros.

Nesse tipo de análise é importante priorizar alvos de estudo, para que possam ser apontados para uma validação experimental [Galperin & Koonin, 2004]. Nosso interesse em explorar as PHs das estirpes CPAC 7 e CPAC 15 é aumentar o conhecimento sobre o aparato proteico de cada bactéria, identificar proteínas que possam estar associadas com a capacidade simbiótica e ainda apontar alvos para estudos experimentais posteriores.

2.5 Explorando regiões atípicas dos genomas bacterianos

Genomas bacterianos apresentam uma estrutura simples, são compactos, sendo que cerca de 90% de seu conteúdo é informativo, o restante corresponde a regiões não codificantes de proteínas, em

geral constituem regiões regulatórias [Casjens, 1998; Rocha, 2008].

Os genes procarióticos são compostos, basicamente, pela região codificadora flanqueada por regiões regulatórias, à montante e à jusante [Casjens, 1998]. Diferente de organismos eucariotos, a presença de íntrons é extremamente rara [Casjens, 1998; Rocha, 2008]. Muitos desses genes podem ainda estar organizados em operons, que são unidades funcionais, na qual duas ou mais regiões codificadoras ocupam posições adjacentes, com a mesma orientação e têm sua transcrição controlada pelas mesmas sequências regulatórias [Rocha, 2008].

Algumas bactérias podem contar com a presença de plasmídeos, que são elementos genéticos móveis extracromossomais, cuja replicação é independente do DNA cromossomal, e são constituídos, entre outros, de genes que podem conferir características especiais às bactérias, como genes de resistência à antibióticos [Casjens, 1998; Rocha, 2008].

São nos cromossomos que as bactérias carregam os genes essenciais à manutenção celular, ou seja, o genoma central [Casjens, 1998; Bentley & Parkhill, 2004; Rocha, 2008]. Além das sequências de DNA dos genes essenciais, alguns cromossomos podem carregar outros DNAs integrados, como DNAs de bacteriófagos, ilhas genômicas (simbióticas ou patogênicas), elementos transponíveis, entre outros [Koonin *et al.*, 2001; Lawrence & Hendrickson, 2005]. Tais elementos acessórios dos genomas contribuem para a variabilidade na estrutura genômica das bactérias, o que faz com que o genoma seja dinâmico, sujeito à inúmeras alterações como reorganização de sequências, fusão de replicons, transferência gênica horizontal, entre outros, mesmo entre linhagens de uma mesma espécie [Casjens, 1998; Koonin *et al.*, 2001; Darmon & Leach, 2014].

Uma característica importante dos cromossomos bacterianos é que a distribuição de oligonucleotídeos é relativamente estável ao longo de todo DNA, sendo essa distribuição particular de cada organismo [Karlin & Burge, 1995; Bohlin & Skjerve, 2009]. Regiões que possuem desvios nessa distribuição podem ser consideradas regiões atípicas ou anômalas, e frequentemente correspondem a regiões de DNA integrados no genoma [Bentley & Parkhill, 2004; Rocha, 2004; Darmon & Leach, 2014].

Portanto, sabendo das características desses genomas e de sua plasticidade, consideramos que as regiões atípicas dos genomas que passaram por modificações durante os processos evolutivos, possam nos fornecer conhecimentos interessantes [Dufraigne *et al.*, 2005]. Dessa forma, essa é mais

uma abordagem utilizada neste trabalho para identificar características distintas das estirpes CPAC 7 e CPAC 15. Na análise comparativa tradicional, foram identificadas várias ilhas genômicas nos cromossomos [Siqueira *et al.*, 2014], entretanto nenhuma análise aprofundada foi realizada.

Além disso, acreditamos que essa abordagem deva ser mais empregada no direcionamento das análises de genomas procariotos. Assim, esse trabalho também propõe uma nova metodologia para discriminar regiões atípicas dos genomas bacterianos, no intuito de utilizar essa característica para guiar o pesquisador, de forma rápida, direto para regiões onde possa ter maior chance de identificar características exclusivas do genoma de interesse.

Há muito tempo os pesquisadores sabem da importância que essas regiões atípicas possuem para o microrganismo, por isso várias formas diferentes vêm sendo empregadas na tentativa de identificá-las [Karlin & Burge, 1995; Dufraigne *et al.*, 2005]. Por exemplo, Garcia-Vallve *et al.* [2003] desenvolveram o banco de dados HGT-DB, que possui genes possivelmente transferidos horizontalmente, em genomas bacterianos, com base no uso de códons e no conteúdo GC do genoma. Uma outra abordagem disponível é o IslandPath, uma ferramenta-*web* utilizada para identificar ilhas genômicas baseada na porcentagem GC e no viés de dinucleotídeos [Hsiao *et al.*, 2003]. Um ano depois, mais um banco de dados com ilhas integrativas também foi disponibilizado (Islander), mas nesse caso, a identificação é baseada nos genes de tRNA [Mantri & Williams, 2004]. Posteriormente, Waack *et al.* [2006] desenvolveram a ferramenta SIGI-HMM para prever ilhas genômicas usando modelo ocultos de Markov. E finalmente, Vernikos & Parkhill [2006] introduziram o método IVOMs (*Interpolated Variable Order Motifs*), conhecido também como Alien Hunter, com o mesmo objetivo.

Apesar de todos esses métodos disponíveis, o desafio é melhorar a performance, reduzir o tempo de execução computacional e os requisitos necessários para a análise (tipos de arquivos de entrada). Pois alguns deles precisam do arquivo de anotação para realizar as análises, ou no caso dos bancos de dados, onde só é possível analisar os genomas disponíveis nos mesmos. Sendo que seria interessante executar tal análise, também em genomas recém-sequenciados.

Por isso, esse estudo também propõe uma estratégia alternativa, simples e rápida computacionalmente para discriminar regiões nos genomas que devem ser priorizadas durante as análises.

A metodologia proposta é baseada na máxima entropia (ME), a qual parte do princípio introduzido

por Shannon [1948], que aponta a distribuição de probabilidade que mede a informação de uma fonte [Shannon, 1948].

A ME foi proposta originalmente por Jaynes [1957a,b], onde basicamente determina a distribuição que representa o máximo de incerteza [Guisu & Shenitzer, 1985]. E neste trabalho, utilizamos a ME com base no conhecimento prévio da frequência de oligonucleotídeos, para discriminar as regiões atípicas dos genomas, ou seja, as regiões que podem conter novas informações.

Portanto, primeiramente desenvolvemos uma nova metodologia, que pode ser aplicada na análise de outros genomas. E em seguida, a utilizamos para discriminar regiões dos genomas de *B. diazoefficiens* CPAC 7 e *B. japonicum* CPAC 15, a fim de explorá-las.

3

Objetivos

3.1 Objetivo Geral

Explorar os genomas das estirpes brasileiras de *Bradyrhizobium* buscando por fatores genéticos que possam contribuir para a compreensão de suas diferenças simbióticas.

3.2 Objetivos Específicos

- Analisar o impacto dos elementos de transposição;
- Identificar novas funções entre as proteínas hipotéticas;
- Desenvolver uma nova metodologia baseada na máxima entropia para discriminação de regiões atípicas;
- Explorar os genes exclusivos das regiões atípicas apontadas pela máxima entropia.

4

Materiais e Métodos

4.1 Obtenção dos genomas

Os genomas no formato FASTA de *B. diazoefficiens* CPAC 7 e *B. japonicum* CPAC 15 foram obtidos do banco de dados de genomas do NCBI (*National Center for Biotechnology Information*), recentemente sequenciados [Siqueira *et al.*, 2014]. O genoma da CPAC 7 possui 9.085.533 pb disponibilizado em 13 *contigs* (Número de Acesso: ADOU02) e a estirpe CPAC 15 tem 9.583.027 pb disponível em um *scaffold* (Número de Acesso: CP007569).

Para a execução das análises, os *contigs* da estirpe CPAC 7 foram ordenados com o alinhador MAUVE [Rissman *et al.*, 2009] utilizando como referência o genoma de *B. diazoefficiens* USDA 110 [Kaneko *et al.*, 2002]. A partir daí, também obtivemos um *scaffold* para CPAC 7.

4.2 Análise dos elementos transponíveis

4.2.1 Identificação e anotação dos elementos de transposição

Os processos de identificação e anotação dos elementos de transposição neste trabalho envolveu duas etapas principais, uma etapa automática e uma etapa manual.

A etapa automática teve início com a submissão dos genomas no formato FASTA ao ISSaga, ferramenta para anotação de TEs de procariotos [Varani *et al.*, 2011]. Primeiramente ocorreu a predição

de todas as ORFs utilizando o Glimmer [Salzberg *et al.*, 1998], e posteriormente foram identificadas as ORFs associadas a elementos de transposição através do BLASTx (*e*-valor 10^{-5}) contra o banco de dados ISfinder (principal banco de dados de TEs de procariotos), assim obtivemos uma lista de ORFs candidatas [Siguier *et al.*, 2006b; Varani *et al.*, 2011]. Em seguida, ocorreu um BLASTn (*e*-valor 10^{-5}) das ORFs contra seu próprio genoma para a predição das outras partes de um IS (IRs: repetições invertidas e DRs: repetições diretas). Quando não identificado, novamente um BLASTn foi realizado do genoma contra o ISfinder. Nessa etapa, quando presente, foi possível predizer os ISs parciais [Varani *et al.*, 2011].

Além disso, cada ORF foi classificada em família, definida pela similaridade com as famílias de elementos transponíveis presentes no ISfinder. A identificação das transposases baseia-se principalmente no seu sítio catalítico [Siguier *et al.*, 2012, 2014]. Todas essas fases foram realizadas pelo ISSaga [Varani *et al.*, 2011].

Ao final dessa predição, obtivemos uma lista de ORFs candidatas e suas respectivas coordenadas. Para os elementos já depositados no banco de dados ISfinder, a anotação foi automaticamente realizada, identificando se o IS estava completo ou parcial, com as coordenadas das ORFs, das IRs e se houve DRs para os elementos completos. Ao contrário, cada ORF predita foi manualmente anotada.

Na etapa manual da anotação, para cada ORF não depositada no ISfinder, a partir da coordenada predita, foram estendidos cerca de 800 nucleotídeos a montante da coordenada inicial e 800 nucleotídeos a jusante da coordenada final, e extraídos do genoma. Assim, foi possível identificar a presença de IRs, seu tamanho, uma ou mais ORFs presentes com seus tamanhos e coordenadas, se durante a inserção houve a formação de DRs, qual seu tamanho e se o IS estava completo ou parcial. Aqui, partimos das características das famílias para a anotação, e além do ISSaga, utilizamos o BLAST [Altschul *et al.* [1990, 1997]] contra o RefSeq [Pruitt *et al.*, 2005; Tatusova *et al.*, 2014], BLAST2Seq [Tatusova & Madden, 1999] e códigos Perl *in-house* para manipulação das sequências.

Foram consideradas isoformas (cópias) de um mesmo IS, se as sequências de aminoácidos da T_pase apresentaram uma identidade > 98% e a sequência de nucleotídeos do IS > 95% de identidade com um IS já depositado no banco de dados. Caso contrário, um novo IS foi identificado, onde lhe foi atribuído um novo nome com base na nomenclatura do ISfinder. Os novos ISs anotados para *B.*

diazoefficiens CPAC 7 receberam o nome de ISBd seguido por um número de acordo com a ordem em que foram anotados, e para *B. japonicum* CPAC 15 foi ISBja.

Ao final da anotação obtivemos o número de ISs completos (com ORF(s) e ambas terminações) e parciais presentes em cada genoma, sua distribuição, o número de cópias para cada IS e suas respectivas famílias.

4.2.2 Contexto genômico dos elementos ISs completos

Com a população de elementos ISs já anotadas, nosso interesse foi verificar quais poderiam estar interferindo localmente nos genomas. Para isso, partimos das coordenadas dos ISs completos de cada genoma, na qual foram mapeadas as ORFs flanqueadoras, calculado a distância existente entre os ISs e suas ORFs vizinhas e em qual sentido (senso e anti-senso) da fita de DNA se encontravam. Esse processo foi realizado utilizando códigos Perl *in-house*.

A partir disso, verificamos manualmente cada caso, inclusive confirmando a anotação das proteínas de interesse através dos bancos de dados RefSeq [Pruitt *et al.*, 2005; Tatusova *et al.*, 2014], UniProt [Consortium, 2008] e Interpro [Hunter *et al.*, 2009].

Cada IS completo analisado recebeu uma classificação de acordo com seu contexto genômico: *i- InterORF*: quando inserido em região intergênica, longe dos genes; *ii- Interrupção de ORFs*: quando o IS completo foi identificado inserido na região codificadora de genes; *iii- Região promotora*: ISs presentes na região promotora do gene (distância < 100 pb a partir do códon iniciador, uma vez que essas regiões possuem tamanhos variados [Estrem *et al.*, 1999; Gordon *et al.*, 2006; Davis *et al.*, 2011]); *iv- Região terminadora*: ISs presentes na região terminadora do gene (distância < 25 pb a partir do códon terminador [Mitra *et al.*, 2010; Peters *et al.*, 2011]); *v- ORFs modificadas*: quando um IS foi identificado sobrepondo um gene quando em fitas opostas, ou IS e ORF compartilhando sequência no mesmo sentido da fita de DNA.

Adicionalmente, para os principais eventos de inserção encontrados nos genomas de CPAC 7 e CPAC 15, verificamos se tais eventos também ocorrem em outros 31 genomas de *Bradyrhizobium*, das espécies *B. diazoefficiens*, *B. japonicum* e *B. elkanii*, todos obtidos do banco de dados de genomas

do NCBI. Essa análise foi realizada utilizando o BLASTn com parâmetros padrões [Altschul *et al.*, 1990, 1997].

4.2.3 Identificação de homólogos dos genes impactados expressos em condição de simbiose

No intuito de tentar entender se os genes impactados podem ter participação no processo de simbiose, buscamos por homólogos desses genes nos estudos de *B. diazoefficiens* USDA 110. Portanto, nessa etapa, realizamos BLASTn dos nossos genes de interesse contra os genes de *B. diazoefficiens* USDA 110 identificados como diferencialmente expressos em condição de simbiose.

Os estudos utilizados nessa análise foram: um estudo de transcrito 21 dias após a inoculação (DAI) em raízes de soja [Pessi *et al.*, 2007], um estudo de proteoma 21 DAI [Delmotte *et al.*, 2010] e um estudo de transcrito 28 DAI [Franck *et al.*, 2014].

4.3 Explorando o conjunto de proteínas hipotéticas

4.3.1 Seleção das proteínas hipotéticas e adição de novas informações

Os genomas de *B. diazoefficiens* CPAC 7 e *B. japonicum* CPAC 15 possuem dois arquivos de anotação disponíveis. Um corresponde a anotação original, realizada pelos autores da publicação dos genomas, depositada no GenBank (CPAC 7: ADOU02 e CPAC 15: CP007569) [Siqueira *et al.*, 2014]. E a outra versão da anotação diz respeito àquela depositada no banco de dados do RefSeq (CPAC 7: NZ_ADOU000000000.2 e CPAC 15: NZ_CP007569.1) [Tatusova *et al.*, 2014].

As proteínas hipotéticas consideradas na análise foram aquelas anotadas como hipotéticas em ambos arquivos de anotação de cada genoma. Portanto, foram extraídas 2.255 PHs de um total de 8.100 ORFs da estirpe CPAC 7; e 2.447 PHs da estirpe CPAC 15 de um total de 8.606 ORFs.

A partir desse conjunto, identificamos quais das proteínas hipotéticas são compartilhadas entre as duas estirpes e quais são específicas, e para isso utilizamos a ferramenta *get_homologues* [Contreras-Moreira & Vinuesa, 2013], a qual é baseada no algoritmo *Bidirectional Best Hits* (BBH) [Overbeek

et al., 1999]. Este algoritmo considera o melhor resultado recíproco entre os dois conjuntos de dados, utilizando cobertura mínima de 75% e *e-valor* mínimo de 10^{-5} .

A adição de novas informações às proteínas hipotéticas foi baseada na homologia de sequências, e utilizamos neste processo de atualização de anotação a ferramenta Blast2GO [Conesa *et al.*, 2005], que integra a busca por similaridade. Utilizamos o BLASTx contra o banco de dados de proteínas não-redundantes, realizamos análise de ontologia (atribuição de termos GO), atribuição do código de enzimas baseado no banco de dados do KEGG [Kanehisa & Goto, 2000] e anotação funcional com o InterProScan [Zdobnov & Apweiler, 2001]. As análises foram realizadas utilizando os parâmetros padrões do Blast2GO [Conesa *et al.*, 2005].

O InterProScan consulta no banco de dados InterPro, por informações de proteínas, como famílias, domínios, sítios funcionais, entre outros. Esse banco integra dados dos principais repositórios de proteínas conhecidos como: CDD, PROSITE, HAMAP, Pfam, PRINTS, ProDom, Smart, TIGRFAMs, PIRSF, SUPERFAMILY, GENE3D e PANTHER [Zdobnov & Apweiler, 2001; Hunter *et al.*, 2009].

Também consultamos a ferramenta Phobius para a confirmar a predição de peptídeo sinal [Käll *et al.*, 2004], o PSORT-B (versão 3.0) para prever a localização celular dos peptídeos sinais [Gardy *et al.*, 2003] e a ferramenta TPRpred para confirmar a predição de tetratricopeptídeos [Karpenahalli *et al.*, 2007].

Ao final do processo de atualização da anotação, as proteínas foram categorizadas funcionalmente, para facilitar na apresentação dos resultados.

4.3.2 Busca por proteínas similares envolvidas nos processos de interação planta-bactéria e FBN

Procurando atribuir maior valor às proteínas, até então hipotéticas, buscamos por proteínas similares às nossas em outros bancos de dados e estudos publicados, com potencial envolvimento no processo de interação planta-bactéria e/ou na fixação de nitrogênio, através do BLASTx [Altschul *et al.*, 1990, 1997].

Primeiramente, identificamos as proteínas do banco de dados PHI-base (*Pathogen-Host Interacti-*

ons) que tiveram identidade $\geq 20\%$ com as sequências de aminoácidos das PHs [Urban *et al.*, 2015]. Este banco de dados armazena proteínas de bactérias patogênicas que já foram experimentalmente estudadas e que possuem participação no processo de patogenicidade [Urban *et al.*, 2015].

Depois, buscamos por homólogos das proteínas de interesse em estudos de simbiose realizados com *B. diazoefficiens* USDA 110, no intuito de verificar se homólogos de tais proteínas são diferenciatamente expressos (DE) em condição de simbiose.

Tais estudos envolvem uma análise transcricional da estirpe USDA 110 21 dias após a inoculação (DAI) na soja [Pessi *et al.*, 2007], um estudo de proteoma nas mesmas condições que o estudo transcricional (21 DAI) [Delmotte *et al.*, 2010], e finalmente, um estudo transcricional de USDA 110 28 DAI em soja [Franck *et al.*, 2014]. Para esta análise, as proteínas DE foram recuperadas, as quais identificamos aquelas que tiveram identidade e cobertura $\geq 60\%$ com as sequências de aminoácidos das PHs estudadas.

Além disso, as estirpes CPAC 7 e CPAC 15 possuem estudos de proteoma em condições de vida-livre, e aqui nós também verificamos se as PHs exploradas no presente estudo já haviam sido identificadas nessa condição [Batista *et al.*, 2010; Gomes *et al.*, 2014].

4.4 Desenvolvimento da estratégia de máxima entropia para explorar regiões atípicas

O genoma ($\mathcal{G}$) de um organismo corresponde basicamente a específicas e finitas sequências de DNA, contendo a distribuição de quatro nucleotídeos $\mathcal{G}$: Adenina (A), Timina (T), Guanina (G) e Citosina (C). Cada genoma possui uma frequência na distribuição de pequenas combinações desses oligonucleotídeos (di, tetra e hexanucleotídeos), onde essa distribuição é relativamente estável e particular de cada genoma bacteriano [Karlin & Burge, 1995; Bohlin & Skjerve, 2009; Bohlin *et al.*, 2010].

O método de máxima entropia foi empregado neste trabalho para discriminar regiões dos genomas que possuem as maiores diferenças nesta distribuição ao longo do genoma, também chamadas de

regiões atípicas ou anômalas. O objetivo de se desenvolver esta nova metodologia foi para agilizar as análises, direcionando o pesquisador para regiões consideradas candidatas a serem exploradas mais detalhadamente, podendo conter informações distintivas de cada bactéria.

Para isso, foi necessário estabelecer o tamanho das regiões a serem analisadas (Ex.: 2 ou 5 kilobase (kb)). Um outro parâmetro necessário foi o tamanho do oligonucleotídeo (*word size* - w) a ser utilizado nas análises, para isso diferentes tamanhos de palavras foram testados (2, 4 e 6).

A metodologia proposta aqui consiste basicamente em duas etapas, o processamento das sequências e o cálculo da máxima entropia.

4.4.1 Processamento das sequências

A metodologia desenvolvida requer apenas o genoma no formato FASTA como arquivo de entrada, a definição do tamanho das regiões (sub-sequências) a serem analisadas e o tamanho dos oligonucleotídeos (2, 4 ou 6).

Primeiramente, o genoma é particionado no tamanho pré-definido, gerando sub-sequências (FASTA) e um arquivo que armazena as coordenadas de cada região. O programa foi desenvolvido para iniciar esse processamento a partir de qualquer coordenada do genoma, a ser escolhido no momento da execução.

Em seguida, para cada sub-sequência, uma janela deslizante é aplicada, caminhando uma base por vez e contabilizando a presença de cada di, tetra ou hexanucleotídeo, dependendo do tamanho de oligonucleotídeo (w) definido.

Na Figura 5-a, representamos como a metodologia contabiliza a presença dos oligonucleotídeos em cada sub-sequência de tamanho desejado. Os valores são armazenados em uma matriz de adjacência. Portanto, se 10 regiões são analisadas, 10 matrizes são geradas.

O final desta etapa resulta, basicamente, em um arquivo com as coordenadas das regiões analisadas, as sequências de cada sub-região genômica (formato FASTA) e as matrizes de adjacência, que são utilizadas para o cálculo da máxima entropia.

Figura 5: Representação de como a contagem de oligonucleotídeos é computada para a análise de máxima entropia. No exemplo temos um genoma (G) analisado em duas regiões. A janela deslizante é aplicada à sub-sequência 1, caminhando uma base por vez. O exemplo é baseado em hexanucleotídeo, dessa forma é contabilizado a relação de cada trinucleotídeo com as demais trincas, ou seja, quantas vezes AAA aparece ao lado de AAA, AAC, AAG, e assim todas as combinações. O mesmo processo ocorre para cada trinucleotídeo, por isso considerado um tamanho de palavra igual a seis. Os valores são armazenados em uma matriz de adjacência, que neste caso tem uma dimensão de 4^3 .

4.4.2 Calculando a máxima entropia

A máxima entropia (ME) [Jaynes, 1957a,b] no presente estudo, foi utilizada para apontar as regiões dentro do genoma que fogem do padrão de distribuição de oligonucleotídeos, ou seja, que aponte a máximo de desordem.

Para se obter esse perfil, a ME utiliza os valores armazenados em cada matriz de adjacência, as quais correspondem as regiões do genoma. Assim, a ME fornecerá o valor máximo de entropia obtido para cada coordenada específica.

ME é dada por Jaynes [1957a], sendo definida pela somatória de H_1 e H_2 :

$$H_1(t) = - \sum_{i=0}^t \frac{h(i)}{\sum_{j=0}^t h(j)} \log_2 \frac{h(i)}{\sum_{j=0}^t h(j)}, \quad (4.1)$$

onde, t corresponde a posição na janela e $h(i)$ é a frequência de oligonucleotídeos, todos apresentados na primeira parte, i é a posição do oligonucleotídeo e $h(j)$ é a frequência total.

E,

$$H_2(t) = - \sum_{i=t+1}^{i_{max}} \frac{h(i)}{\sum_{j=t+1}^{i_{max}} h(j)} \log_2 \frac{h(i)}{\sum_{j=t+1}^{i_{max}} h(j)}, \quad (4.2)$$

assim, existirá um conjunto de valores

$$ME_t = H_1 + H_2, \quad (4.3)$$

na qual a máxima entropia é escolhida para representar cada região genômica.

Na Figura 6, exemplificamos como esse processo ocorre utilizando um tamanho de palavra igual a seis.

Figura 6: Cálculo da máxima entropia para as regiões genômicas. Neste caso a análise é baseada no hexanucleotídeo (relação de cada trinucleotídeo com os demais), assim todas as combinações na matriz de adjacência resulta em 64 valores. Sobre esses valores são aplicadas as equações 4.1 e 4.2. Exemplificando, em ME_1 , as relações de AAA representa H_1 (primeira parte) e a soma dos demais corresponde a H_2 (segunda parte). Em ME_2 a soma de AAA e AAC representa H_1 e o restante H_2 , e assim sucessivamente. Ao final, teremos 63 valores resultantes das somatórias, na qual o maior valor (ME_{max}) é escolhido para representar aquela região. o mesmo procedimento é aplicado à todas regiões genômicas.

A distribuição dos valores ME que representam cada região ao longo do cromossomo podem

ser visualizados em um gráfico, sendo que quanto menor o tamanho da região analisada, pior será a resolução da visualização no gráfico. Por esse motivo, essa metodologia foi testada utilizando diferentes tamanhos de regiões (2kb, 5kb, 10kb e 50kb). Além disso, para cada tamanho de região, aplicamos a metodologia com diferentes tamanhos de palavra (w : 2, 4 e 6), a fim de avaliarmos a performance da ME em discriminar regiões usando cada parâmetro.

Considerando que estamos buscando por regiões atípicas, os valores ME menores e ao redor da mediana podem ser consideradas regiões mais estáveis. Assim, as regiões com valor ME maior ou igual ao terceiro quartil (Q_3) são as regiões candidatas a serem exploradas, ou seja, preditas como regiões atípicas ($ME \geq Q_3$). Os principais resultados fornecidos por essa análise são: o gráfico contendo a distribuição dos valores ME ao longo do genoma, suas respectivas coordenadas e uma lista das regiões atípicas preditas (Figura 7).

Figura 7: Procedimento de análise da metodologia de máxima entropia para discriminação de regiões atípicas de genomas bacterianos. A análise consiste em duas etapas principais, a partir do genoma no formato FASTA: o processamento das sequências e o cálculo da máxima entropia. Os resultados são um gráfico com a distribuição dos valores ME, as sub-regiões no formato FASTA, suas coordenadas e as regiões atípicas preditas ($ME \geq Q_3$). Além disso, para os genomas que possuem os arquivos de anotação, a lista de proteínas presentes nas regiões preditas e informações sobre a presença de ribossomais nas mesmas também podem ser obtidas.

Além disso, caso o genoma analisado já possua seus arquivos de anotação disponíveis no banco de dados do NCBI (Arquivo de proteínas (.ptt) e de RNAs (.rnt)), a lista de proteínas presentes nas

regiões atípicas preditas e informações sobre a presença de ribossomais nessas regiões podem ser obtidas (Figura 7).

A implementação da metodologia ME foi realizada utilizando as linguagens Perl e R de programação, a qual está disponível em um pacote que pode ser executado em linha de comando.

4.4.3 Avaliação da metodologia proposta

Os genomas adotados para avaliar a metodologia foram *Xanthomonas axonopodis* pv. *citri* 306 (XAC) e *Xanthomonas campestris* pv. *campestris* ATCC 33913 (XCC), os quais já possuem suas regiões atípicas bem caracterizadas dentro do seu cromossomo [Lima *et al.*, 2005, 2008]. Dessa forma, as CDSs (sequências codificantes) presentes nas regiões atípicas foram definidas como verdadeiro positivo (VP) [Lima *et al.*, 2005, 2008].

A partir daí, foi possível calcular a sensibilidade (%) (Equação 4.4), a especificidade (%) (Equação 4.5), e consequentemente a eficiência (%) (Equação 4.6) e o F-score (%) da metodologia proposta [Powers, 2011].

A sensibilidade (SENS) corresponde a capacidade da metodologia em predizer corretamente as CDSs de regiões atípicas, e se dá por:

$$SENS = \frac{VP}{(VP + FN)} \quad (4.4)$$

onde, VP corresponde aos verdadeiros positivos e FN aos falsos negativos.

A especificidade (ESP) diz respeito à capacidade da metodologia em predizer de forma correta as CDSs que não correspondem a regiões atípicas, dada por:

$$ESP = \frac{VN}{(VN + FP)} \quad (4.5)$$

na qual, VN é o número de verdadeiros negativos e FP o número de falsos positivos.

Pelo fato de métodos preditores raramente conseguirem obter sensibilidade e especificidade com 100% de acertos, um equilíbrio deve ser atingido, por isso calculamos também a eficiência (EF) da

metodologia, que corresponde a média entre a sensibilidade e a especificidade:

$$EF = \frac{(SENS + ESP)}{2} \quad (4.6)$$

O F-score é uma medida que combina precisão e sensibilidade, muito utilizado para avaliar métodos preditivos, onde F-score: 1 é considerada uma predição perfeita e F-score: 0 corresponde a pior predição possível [Powers, 2011].

Além disso, analisamos os genomas de XAC e XCC com outros métodos disponíveis para a detecção de regiões atípicas (Alien Hunter [Vernikos & Parkhill, 2006], HGT-DB [Garcia-Vallve *et al.*, 2003], Islander [Mantri & Williams, 2004], IslandPath [Hsiao *et al.*, 2003] e SIGI-HMM [Waack *et al.*, 2006]) e comparamos os resultados obtidos com a metodologia ME. Todas as análises foram realizadas no mesmo computador (Processador Intel Core i3-540 e 8GB RAM), para que os resultados pudessem ser comparados também considerando o tempo de execução da análise. Para os métodos disponíveis como banco de dados (HGT-DB e Islander), os resultados já computados foram baixados.

Ademais, selecionamos regiões do genoma de *X. campestris* pv. *campestris* ATCC 33913 (XCC) que tiveram valores ME abaixo da mediana, na mediana e no terceiro quartil. Em seguida, verificamos se tais regiões são compartilhadas por outras bactérias do mesmo gênero e da mesma espécie. Para isso utilizamos o BLASTn [Altschul *et al.*, 1990] e observamos sua porcentagem (%) de cobertura.

4.4.4 Máxima entropia utilizada na seleção de genes exclusivos de *B. diazoefficiens* CPAC 7 e *B. japonicum* CPAC 15

Para cada sequência FASTA (genoma), a máxima entropia foi aplicada, baseando-se em regiões genômicas de 25 kb e utilizando hexanucleotídeos. Portanto, obtivemos um perfil de máxima entropia, sendo que aquelas regiões com valores ME $\geq$ ao terceiro quartil (Q_3) foram as regiões consideradas como de interesse.

Pelo fato deste estudo ter sido focado em apontar as particularidades de cada genoma, somente as regiões preditas pela máxima entropia que possuem uma cobertura $\leq 75\%$ no genoma de CPAC 7 em comparação com CPAC 15, e vice-versa, seguiram nas análises. Para identificar quais seriam

estas regiões, utilizamos a ferramenta BLASTn com parâmetros padrões. Em seguida, de cada região selecionada, foram extraídas as ORFs exclusivas de cada bactéria, também com base nos resultados do BLASTn. As ORFs exclusivas que já haviam sido abordadas na análise das sequências de inserção e das proteínas hipotéticas foram desconsideradas, no intuito de não duplicar resultados.

Dessa forma, obtivemos uma lista de proteínas exclusivas para cada estirpe, onde foi possível destacar suas principais particularidades. O enfoque maior foi dado para aquelas que não haviam sido destacadas na publicação da análise comparativa dos genomas, e que poderiam contribuir para as diferenças simbióticas das estirpes.

Resultados e Discussão

5.1 População de elementos transponíveis presentes nas estirpes brasileiras de *Bradyrhizobium*

Os genomas das estirpes brasileiras de *Bradyrhizobium* foram submetidos à ferramenta ISSaga, para predição de ORFs relacionadas a TEs, a seguir foi realizado uma anotação manual dessas ORFs, focando na análise dos elementos ISs completos (ORFs e ambas terminações invertidas). Foram identificados 77 ISs completos em *B. diazoefficiens* CPAC 7 e 81 em *B. japonicum* CPAC 15. A anotação detalhada desses elementos pode ser acessada nos Apêndices A e B.

Após o processo de anotação, obtivemos a distribuição desses elementos nos dois genomas (Figura 8), sendo facilmente observado a concentração dentro da ilha simbiótica, conforme previamente mencionado por Siqueira *et al.* [2014]. Nessa distribuição podemos visualizar tanto os ISs completos, bem como os elementos parciais. A predominância dos elementos ISs na ilha simbiótica também já foi relatada em outros genomas de rizóbios como, *Mesorhizobium loti* MAFF303099 [Kaneko *et al.*, 2000], *B. diazoefficiens* USDA 110 [Kaneko *et al.*, 2002], *B. japonicum* USDA 6 [Kaneko *et al.*, 2011] e *B. diazoefficiens* NK6 [Iida *et al.*, 2015].

Acredita-se que a ilha simbiótica seja proveiente de um plasmídeo simbiótico (pSym) que tenha sido integrado ao cromossomo via transferência horizontal, como é o caso das bactérias em estudo [Sullivan & Ronson, 1998; Göttfert *et al.*, 2001; Finan, 2002; Okubo *et al.*, 2016]. Essa seria uma

possível explicação para a alta concentração de ISs da ilha simbiótica, pois segundo Siguier *et al.* [2006a], a presença de elementos ISs em plasmídeos auxilia na inclusão de novos de genes, aumentando o repertório de genes não essenciais para sua manutenção. Assim, possivelmente o pSym tenha trazido consigo muitos dos ISs no momento da integração, pois conforme apresentado no pSym de *Rhizobium* sp. NGR234, tais elementos estão presentes [Freiberg *et al.*, 1997].

No estudo de Uchiumi *et al.* [2004], analisaram a expressão dos genes da ilha simbiótica em bacteroides de *M. loti* MAFF303099 e observaram que a maioria das transposases, bem como os genes da simbiose, estavam altamente expressos (*up-regulated*), enquanto que os genes fora da ilha mostraram-se *down-regulated*. Resultados semelhantes também foram obtidos em *B. diazoefficiens* USDA 110 [Wei *et al.*, 2008]. Portanto, tais estudos fortalecem a hipótese de que os elementos de transposição também podem desempenhar algum papel no processo da FBN.

Figura 8: Distribuição dos elementos ISs ao longo dos genomas de *B. diazoefficiens* CPAC 7 e *B. japonicum* CPAC 15.

Os elementos completos identificados pertencem a 15 famílias diferentes. Apesar do número de ISs nas duas bactérias serem bem próximos (77 em CPAC 7 e 81 em CPAC 15), quando analisamos

as famílias a que pertencem, temos algumas variações (Figura 9). As principais diferenças são em relação as famílias IS481, IS5 e IS110, onde a estirpe CPAC 15 possui maior número de elementos dessas famílias. Já a estirpe CPAC 7, tem mais elementos das famílias IS21 e IS1380. Além disso, a CPAC 7 possui um elemento IS representante da família IS701 e IS91, que não estão presentes na CPAC 15 (Figura 9). Observar tais diferenças são importantes, uma vez que as famílias dos ISs possuem características diferentes [Mahillon & Chandler, 1998; Siguier *et al.*, 2014].

Figura 9: Famílias dos ISs completos presentes nos genomas de *B. diazoefficiens* CPAC 7 e *B. japonicum* CPAC 15.

Por exemplo, a família IS3 é amplamente distribuída, sendo uma das famílias que mais se tem conhecimento, carregam duas ORFs na sua estrutura e seu mecanismo de transposição é "copia-e-cola", ou seja, por replicação [Siguier *et al.*, 2014, 2015]. A família IS481 possui características muito semelhantes à família IS3, como o sítio catalítico da transposase e as terminações invertidas [Siguier *et al.*, 2015]. Além disso, também parece ter o mesmo mecanismo de transposição da IS3. Muitos pesquisadores acreditam que a família IS481 seja derivada da família IS3, entretanto possuem apenas uma ORF [Siguier *et al.*, 2015].

Quanto às famílias exclusivas de CPAC 7, temos IS701 que possui uma ORF, contudo seu me-

canismo de transposição ainda é desconhecido; e a família IS91 que também possui uma ORF, cuja transposição é replicativa, conhecida como "círculo rolante" (*rolling circle*), devido a estrutura formada na sua transposição [Cerveau *et al.*, 2011; Siguier *et al.*, 2015].

A família IS5, a mais representativa entre os ISs completos de CPAC 15 (Figura 9), possui alguns sub-grupos [Siguier *et al.*, 2014, 2015]. Entre os elementos identificados neste trabalho e pertencentes à esta família, temos representantes do sub-grupo IS5 ssgr IS5, os quais possuem uma ORF e com tamanhos variando de 1.509 a 1.535 pb; e o sub-grupo IS5 ssgr IS427 que carregam duas ORFs e possuem um tamanho na faixa de 830 pb. Destes, destacamos o ISBd23 com 14 cópias em CPAC 15 e oito cópias na estirpe CPAC 7 (Tabela 1).

A família IS21 é considerada bastante homogênea e é caracterizada por carregar duas ORFs na sua estrutura [Siguier *et al.*, 2014, 2015]. Mas no ISBd13, cópia única identificada neste trabalho, detectamos três ORFs, sendo esse elemento exclusivo de *B. diazoefficiens* CPAC 7 (Tabela 1).

A família IS110, mais especificamente IS110 ssgr IS1111, também está bem representada nas estirpes brasileiras, principalmente na estirpe CPAC 15, que possui 14 elementos ISs completos pertencentes a esta família (Figura 9), que por característica, possui uma ORF e não possuem IRs bem definidas [Siguier *et al.*, 2014, 2015].

Outra família que se destaca é a IS1380, cujo único representante dessa família é o ISBd5, que possui 12 cópias na CPAC 7 e apenas uma em CPAC 15 (Tabela 1). Membros dessa família geralmente possuem uma ORF e seu mecanismo de transposição ainda permanece desconhecido [Siguier *et al.*, 2014, 2015].

Na Tabela 1 apresentamos todos os ISs completos identificados, juntamente com seu respectivo número de cópias, tamanho, número de ORFs e quais estão localizados na ilha simbiótica. As características dos ISs são baseadas na cópia do IS referência de acordo com o ISSaga [Varani *et al.*, 2011]. Entre os 39 elementos ISs completos, 33 são elementos novos, somente seis (ISBj2, ISBj4, ISBj5, ISBj6, ISRj1 e ISRj2) já haviam sido anotados e depositados no ISfinder (Tabela 1). Assim, com nosso estudo também aumentamos o número de elementos conhecidos.

As estirpes CPAC 7 e CPAC 15 compartilham 22 elementos, sendo o número de elementos exclusivos maior em CPAC 7. Além disso, considerando todos os elementos completos, cerca de 61% em

Tabela 1: Comparação do número de cópias de cada sequência de inserção completa anotada nos genomas.

ISs completos	Famílias	Bd CPAC 7		Bj CPAC 15		Tam. do IS referência (pb)	No. de ORFs	Referência
		No. de cópias	Na ilha simbiótica	No. de cópias	Na ilha simbiótica			
ISBd1	ISNCY ssgr ISLbi1	2	1	1	0	1.743	1	Presente estudo
ISBd2	IS256	2	2	4	3	1.395	1	Presente estudo
ISBd3	IS256	1	1	1	1	1.303	1	Presente estudo
ISBd4	IS481	1	1	3	2	1.240	1	Presente estudo
ISBd5	IS1380	12	4	1	1	1.622	1	Presente estudo
ISBd6	ISL3	1	1	2	0	1.654	1	Presente estudo
ISBd7	IS1182	1	0	-	-	1.697	1	Presente estudo
ISBd8	IS5 ssgr IS5	1	0	-	-	1.535	1	Presente estudo
ISBd9	IS21	2	1	3	2	2.532	2	Presente estudo
ISBd10	IS21	1	1	-	-	2.606	2	Presente estudo
ISBd11	IS91	1	0	-	-	1.283	1	Presente estudo
ISBd12	IS21	4	2	1	0	2.479	2	Presente estudo
ISBd13	IS21	1	1	-	-	2.245	3	Presente estudo
ISBd14	IS91	1	0	-	-	1.103	1	Presente estudo
ISBd15	IS6	1	1	4	4	816	1	Presente estudo
ISBd16	IS110 ssgr IS1111	1	1	1	0	1.410	1	Presente estudo
ISBd17	IS110 ssgr IS1111	1	1	1	1	1.339	1	Presente estudo
ISBd18	IS110 ssgr IS1111	1	1	-	-	1.357	1	Presente estudo
ISBd19	IS110 ssgr IS1111	1	1	-	-	1.263	1	Presente estudo
ISBd20	IS630	2	0	2	0	1.185	1	Presente estudo
ISBd21	IS630	1	0	-	-	1.044	2	Presente estudo
ISBd22	IS630	1	0	-	-	1.044	2	Presente estudo
ISBd23	IS5 ssgr IS427	8	7	14	8	832	2	Presente estudo
ISBd24	IS3 ssgr IS150	1	1	-	-	1.284	2	Presente estudo
ISBd25	IS3 ssgr IS3	1	1	1	1	643	2	Presente estudo
ISBj2	IS5 ssgr IS5	2	2	2	2	1.509	1	Isfinder
ISBj4	IS110 ssgr IS1111	3	0	4	0	1.576	1	Isfinder
ISBj5	IS5 ssgr IS5	2	0	2	0	1.524	1	Isfinder
ISBj6	IS701	1	1	-	-	1.448	1	Isfinder
ISBja1	IS481	-	-	3	0	1.271	1	Presente estudo
ISBja2	IS1182	-	-	3	0	1.564	1	Presente estudo
ISBja3	IS110 ssgr IS1111	1	1	2	1	1.406	1	Presente estudo
ISBja4	IS110 ssgr IS1111	1	0	4	0	1.370	1	Presente estudo
ISBja5	IS110 ssgr IS1111	-	-	1	0	1.352	1	Presente estudo
ISBja6	IS3 ssgr IS407	-	-	2	1	1.268	2	Presente estudo
ISBja7	IS110 ssgr IS1111	-	-	1	0	1.154	1	Presente estudo
ISBja8	IS66	1	1	1	1	1.584	3	Presente estudo
ISRj1	IS630	9	7	10	8	1.118	1	Isfinder
ISRj2	IS3 ssgr IS150	7	6	7	4	1.364	2	Isfinder

CPAC 7 e 50% em CPAC 15 estão na ilha simbiótica (Tabela 1). A descrição detalhada da estrutura de cada elemento completo, juntamente com sua localização nos dois genomas pode ser encontrada nos Apêndices A e B.

A seguir, apresentamos o impacto local que tais elementos anotados podem estar exercendo sobre os genomas analisados.

5.1.1 Impacto local dos ISs completos nos genomas das estirpes CPAC 7 e CPAC 15

No presente estudo, o impacto local dos elementos completos foi classificado em cinco categorias de acordo com seu contexto genômico. InterORF corresponde aos elementos inseridos na região intergênica, sem impactar ORFs; Interrupção de ORFs quando identificado inserido na região codificadora; Região promotora, quando localizado a uma distância de até 100 nucleotídeos a partir do códon iniciador da ORF; Região terminadora quando presente a uma distância de até 25 nucleotídeos após o códon de parada; e ORF modificada, quando identificamos elemento IS e ORF em sentidos opostos na fita de DNA se sobrepondo ou quando no mesmo sentido da fita, mas compartilhando sequência. Na Figura 10, mostramos um resumo comparativo do impacto local identificado em ambos genomas das estirpes brasileiras de *Bradyrhizobium*.

Conforme o esperado, a maioria das inserções foi encontrada em regiões intergênicas (Figura 10), por ser menos prejudicial ao organismo [Plague, 2010]. De acordo com Zaghloul *et al.* [2007], há uma seleção negativa contra a inserção de elementos ISs em *Shigella flexneri* 2457T, o que foi confirmado neste estudo nos genomas de *B. diazoefficiens* CPAC 7 e *B. japonicum* CPAC 15.

O número de elementos interrompendo ORFs, presente na região terminadora e modificando ORFs é muito similar nas duas estirpes. A diferença mais evidente é que a CPAC 7 possui mais elementos completos inseridos na região promotora do que a CPAC 15 (Figura 10). A lista completa dos genes impactados estão apresentados na Tabela 2.

B. diazoefficiens CPAC 7 possui 13 genes interrompidos por ISs completos, enquanto que *B. japonicum* CPAC 15 tem 15. Entre esses genes encontramos transposases, genes hipotéticos, genes

Figura 10: Contexto genômico dos ISs completos identificados em *B. diazoefficiens* CPAC 7 e *B. japonicum* CPAC 15. **InterORF**- Região intergênia: ISs em regiões intergênicas, longe dos genes **Interrupção de ORFs**: ISs interrompendo a região codificadora de genes; **Região promotora**: ISs presentes na região promotora do gene (distância < 100 pb a partir do códon iniciador); **Região terminadora**: ISs presentes na região terminadora do gene (distância < 25 pb a partir do códon terminador); **ORFs modificadas**: IS sobrepondo um gene quando em fitas opostas, ou IS e ORF compartilhando sequência no mesmo sentido da fita de DNA.

envolvidos na conjugação, metiltiotransferase, oxidoredutase, ABC transportador, entre outros (Tabela 2). A maior parte dos genes interrompidos estão localizados na ilha simbiótica, sendo que quatro têm seus homólogos em *B. diazoefficiens* USDA 110 identificados como diferencialmente expressos em condição de simbiose [Pessi *et al.*, 2007; Delmotte *et al.*, 2010]. Destacamos aqui dois genes interrompidos: a metiltiotransferase interrompida em CPAC 7 e em CPAC 15; e o ABC transportador interrompido somente em CPAC 15 (Tabela 2).

O gene que codifica a metiltiotransferase é cópia única nos genomas, o qual possui um tamanho de 1.980 pb. Pode estar envolvido, entre outras funções, em fornecer o radical SAM para a formação da nitrogenase [Wiig *et al.*, 2012; Hu & Ribbe, 2013; Ribbe *et al.*, 2014]. Interessantemente, esta interrupção identificada em ambas bactérias foi causada por diferentes elementos (ISBd23 e ISRj1) e eventos de inserção diferentes (Tabela 2). Uma representação desse evento é mostrada na Figura 11-a.

Tabela 2: Genes impactados pelos ISs completos encontrados nas estirpes brasileiras de *Bradyrhizobium*, sendo que alguns homólogos destes genes são diferencialmente expressos em condição de simbiose em *B. diazoefficiens* USDA 110.

Impacto	Elemento IS	Gene	CPAC 7 Locus Tag	CPAC 15 Locus Tag	Distância / Modificando (pb)	Ilha Simb.	Homólogos em simbiose
Interrupção	ISBd1	Transposase	BJA5080_07124	BJS_05811	-	Não	-
Interrupção	ISBd2	ABC transportador	-	BJS_08733	-	Sim	blr1679
Interrupção	ISBd4	Transposase	BJA5080_08164	BJS_08315	-	Sim	bls1856
Interrupção	ISBd4	Transcriptase reversa	-	BJS_08833	-	Sim	-
Interrupção	ISBd5	Hipotético	BJA5080_06299	-	-	Não	-
Interrupção	ISBd6	Fosfoglicolato fosfatase	-	BJS_07033	-	Não	-
Interrupção	ISBd9	Transposase	BJA5080_08268	BJS_08831	-	Não	-
Interrupção	ISBd10	Hipotético	BJA5080_08345	-	-	Sim	-
Interrupção	ISBd12	TrbI -conjugação	BJA5080_08064	-	-	Sim	-
Interrupção	ISBd12	DEAD- helicase	BJA5080_06102	-	-	Não	-
Interrupção	ISBd15	Transposase	-	BJS_08948	-	Sim	-
Interrupção	ISBd17	Transposase	BJA5080_08069	BJS_08415	-	Sim	-
Interrupção	ISBd23	TrbI -conjugação	BJA5080_08366	BJS_08864	-	Sim	-
Interrupção	ISBd23	Transposase	-	BJS_08412	-	Sim	-
Interrupção	ISBd23 (CPAC15) ISRj1 (CPAC 7)	Metiltiotransferase/radical SAM	BJA5080_07602	BJS_08822	-	Sim	bl11977
Interrupção	ISRj1	TraG	BJA5080_08181	BJS_08775	-	Sim	-
Interrupção	ISRj1	Oxidoredutase	BJA5080_00853	-	-	Não	-
Interrupção	ISRj1	Hipotético	-	BJS_07667	-	Sim	bl11979
Interrupção	ISBja3	Transposase	BJA5080_07660	BJS_08809	-	Sim	-
Interrupção	ISBja6	Transposase	-	BJS_08844	-	Sim	-
Região Promotora	ISBd5	Hipotético	BJA5080_08087	-	95	Sim	-
Região Promotora	ISBd5	Hipotético	BJA5080_00952	-	61	Não	-
Região Promotora	ISBd5	Hipotético	BJA5080_04252	-	16	Não	-
Região Promotora	ISBd5	Pilus tipo IV- montagem PilZ	BJA5080_08401	-	67	Não	blr5568
Região Promotora	ISBd5	Hipotético	BJA5080_08259	-	91	Não	-
Região Promotora	ISBd11	PriA-Primosomal N	BJA5080_01519	-	18	Não	-
Região Promotora	ISBd12	ATPase	BJA5080_06101	-	71	Não	-
Região Promotora	ISBd18	Transportador	BJA5080_07988	-	81	Sim	-
Região Promotora	ISBd23	Hipotético	BJA5080_07587	-	97	Sim	-
Região Promotora	ISBd23	Hipotético	BJA5080_07732	-	44	Sim	-
Região Promotora	ISBd23	Transglicosilase lítica de mureína	-	BJS_07214	48	Não	bl12447
Região Promotora	ISBd23	Multi-sensor de transdução de sinal	-	BJS_07014	19	Não	-

Tabela 2 – Continuação

Impacto	Elemento IS	Gene	CPAC 7 Locus Tag	CPAC 15 Locus Tag	Distância / Modificando (pb)	Ilha Simb.	Homólogos em simbiose
Região Promotora	ISBd23	Glucosiltransferase	-	BJS_02329	35	Não	-
Região Promotora	ISBj4	Hipotético	BJA5080_00921	-	80	Não	-
Região Promotora	ISBja2	Putativo peptídeo sinal	-	BJS_06214	81	Não	-
Região Promotora	ISBja3	Peptídeo sinal	-	BJS_00582	9	Não	-
Região Promotora	ISBja8	Hipotético	BJA5080_08031	-	59	Sim	-
Região Promotora	ISBja8	Hipotético	BJA5080_08028	-	50	Sim	-
Região Promotora	ISRj1	Efeto Tipo III - nopAN	N/A	BJS_08797	7	Sim	blr1912
Região Promotora	ISRj1	Amidohidrolase	BJA5080_06025	-	40	Não	-
Região Promotora	ISRj2	PilA2 – montagem de pilus	N/A	-	15	Sim	-
Região Promotora	ISRj2	Exonuclease SbcD	BJA5080_08362	-	24	Sim	-
Região Promotora	ISRj2	Peptidase C58	BJA5080_07710	-	24	Sim	blr2118
Região Terminadora	ISBd5	Efeto Tipo III – nopAG	BJA5080_08310	BJS_09091	1	Sim	bl11862
Região Terminadora	ISBd5	Domínio Sell	BJA5080_08421	-	1	Não	-
Região Terminadora	ISBd5	Transglutaminase	BJA5080_08400	-	1	Não	blr5569
Região Terminadora	ISBd12	Hipotético	-	BJS_08586	2	Não	-
Região Terminadora	ISBd23	Efeto Tipo III - RhcU	BJA5080_08074	BJS_08411	1	Sim	-
Região Terminadora	ISBd23	Adenilato/guanilato ciclase	-	BJS_01536	1	Não	-
Região Terminadora	ISBj4	HTH- tipo- lysR	-	BJS_04020	7	Não	-
Modificando ORF	ISBd2	modC-Molibdênio ABC transp.	BJA5080_08107	BJS_08743	48	Sim	-
Modificando ORF	ISBd3	Hipotético	BJA5080_07730	BJS_07527	24	Sim	-
Modificando ORF	ISBd5	Hipotético	BJA5080_08396	-	40	Não	-
Modificando ORF	ISBd5	Hipotético	BJA5080_08161	BJS_08313	18-15	Sim	-
Modificando ORF	ISBd9	Hipotético	BJA5080_08296	BJS_08483	3	Sim	-
Modificando ORF	ISBd11	Hipotético	BJA5080_01521	-	121	Não	blr0447
Modificando ORF	ISBd15	Hipotético	BJA5080_07526	BJS_08509	4	Sim	-
Modificando ORF	ISBd16	Hipotético	BJA5080_02655	-	28	Sim	blr1546
Modificando ORF	ISBd20	Putativo regulador transcricional	BJA5080_06196	BJS_07670	23	Não	-
Modificando ORF	ISBd23	Multi-sensor de transdução de sinal	-	BJS_08694	16	Não	-
Modificando ORF	ISRj1	Lisina 2,3-aminomutase	-	BJS_08612	69	Não	bl18232
Modificando ORF	ISRj2	HypB-hidrogenase	BJA5080_08448	BJS_08967	50	Sim	blr1732
Modificando ORF	ISRj2	Exonuclease – subunit SbcD	BJA5080_08346	-	42	Sim	bl11924

No genoma de CPAC 7, esse gene teve sua coordenada corrigida (BJA5080_07602, Cont24: Parte I 1595750:1595779 e Parte II 1596900:1598849). A proteína codificada pelo gene metiltiotransferase foi interrompida pelo ISRj1 intacto, contudo, seu domínio funcional parece estar intacto, segundo uma busca realizada no banco de dados do Interpro (Figura 11-a). Como a metiltiotransferase e o ISRj1 estão localizadas em sentidos opostos da fita de DNA, não podemos afirmar que esse elemento está inativando esse gene, para isso, estudos experimentais são necessários. Pois conforme observado em *Escherichia coli*, a inserção de um elemento IS na região codificadora do gene nem sempre causa inativação do mesmo [Skorski *et al.*, 2007].

Contrariamente, o domínio funcional da metiltiotransferase em CPAC 15 está interrompido (Parte I BJS_08822 - 224174:224900 e Parte II BJS_08163 - 218724:219979) pelo ISBd23 defectivo. Adicionalmente, foi encontrado entre as partes I e II desse gene, um IS parcial (BJS_08711), que não corresponde ao IS presente em CPAC 7, seguido por um íntron do grupo II (BJS_08710), conforme mostrado na Figura 11-a. Com isso, podemos concluir que os eventos de inserção que causaram a interrupção da metiltiotransferase ocorreram em momentos distintos nessas estirpes. Este gene permanece intacto em *B. japonicum* USDA 6 (BJ6T_78720) e *B. diazoefficiens* USDA 110 (bll1977), conforme mostra o resultado da análise comparativa dos eventos de inserção em outros genomas de *Bradyrhizobium* (Figura 12). Além disso, na USDA 110 foi identificado como diferencialmente expresso 21 DAI [Delmotte *et al.*, 2010]. Portanto, estudos experimentais focados nesse gene precisam ser realizados para entendermos se o domínio funcional intacto da metiltiotransferase em CPAC 7 poderia estar relacionado com suas alta eficiência em fixar N₂, em comparação com a CPAC 15, que possui o domínio funcional interrompido.

Também dentro da ilha simbiótica, encontramos um gene que codifica um ABC transportador (BJS_08733 - Parte I 9436782:9436851 e Parte II 9438256:9438848) interrompido pelo ISBd2 intacto (Tabela 2). Esse gene faz parte de um operon cópia única presente no genoma das quatro estirpes de *Bradyrhizobium* (CPAC 7, CPAC 15, USDA 110 e USDA 6), mas a interrupção ocorre apenas em CPAC 15 (Figura 12). É possível que isso possa interferir de alguma forma o processo de simbiose nessa estirpe, uma vez que seu homólogo (blr1679) é requerido na simbiose 21 DIA [Pessi *et al.*, 2007]. A identificação desse tipo de impacto nos genomas de interesse é fundamental, já que é

frequente e pode ser deletério para a bactéria, conforme apontado por Plague [2010].

Figura 11: Representação de genes importantes para a simbiose e localizados na ilha simbiótica que podem estar sendo afetados por ISs completos. **a-** Metiltransferase interrompida por diferentes eventos de inserção em *B. diazoefficiens* CPAC 7 e *B. japonicum* CPAC 15; **b-** Gene *nopAN* (Nodulation Outer Protein) contendo *ISrj1* (Família IS630) dentro da sua região promotora.

Nosso trabalho também identificou elementos ISs em regiões promotoras dos genes, os quais podem inativar a expressão dos mesmos ou alterar seu nível de expressão [Nagy & Chandler, 2004; Cerveau *et al.*, 2011]. Por exemplo, Zhang & Saier [2009] mostraram que a expressão de genes essenciais para o uso do glicerol foram ativados pela presença de um IS5 em *E. coli*. Em *Streptococcus*, o gene *lmb* (laminin-binding protein) normalmente expresso, teve seu nível de expressão aumentado quando na presença do IS1548 na sua região promotora [Al Safadi *et al.*, 2010].

No genoma de *B. diazoefficiens* CPAC 7 encontramos o maior número de elementos ISs na região promotora de genes (distância de até 100 pb a partir do códon iniciador) (Figura 10). A distância

exata entre os ISs e o códon iniciador destes genes pode ser verificada na Tabela 2. Dentre os eventos de inserção identificados nas regiões promotoras em ambas estirpes brasileiras, somente um é compartilhado entre elas. Esta inserção corresponde ao gene *nopAN*, um efetor do tipo III localizado na ilha simbiótica, que possui o ISRj1 (Família IS630) inserido a uma distância de sete nucleotídeos do códon iniciador (Figura 11-b). O ISRj1 está intacto em CPAC 15, por outro lado, em CPAC 7 ele parece estar defectivo. Uma outra importante observação é que no genoma de CPAC 7, o gene *nopAN* não tinha sido anotado, e no presente estudo nós o identificamos (Cont20 – 113141:113977), por essa razão na Tabela 2, não há *Locus tag* disponível para esse gene.

Genes *nops* (*Nodulation Outer Proteins*) são essenciais para o processo de infecção durante o estabelecimento simbiótico [Kimbrel *et al.*, 2013], e neste estudo identificamos que um elemento IS pode estar interferindo de alguma forma na regulação da sua expressão, uma vez que está tão próximo de seu códon iniciador. Tentamos reconstruir *in silico* o promotor desse gene na ausência do IS e identificar o promotor na presença do IS, mas não obtivemos resultados relevantes. Portanto, para a confirmação desta interferência outros estudos são necessários. Conforme apresentado na (Figura 12), o ISRj1 presente na região promotora do gene *nopAN* é conservado em *B. diazoefficiens* CPAC 7, *B. diazoefficiens* USDA 110, *B. japonicum* CPAC 15, *B. japonicum* E109 e *B. japonicum* USDA 6. Além disso, em *B. diazoefficiens* USDA 110, esse gene (blr1912) foi identificado como diferencialmente expresso 21 e 28 DAI, indicando a importância desse gene durante o processo de simbiose e que o ISRj1 pode ter participação nessa regulação (Figura 12).

Um outro evento de inserção na região promotora que chamou nossa atenção neste trabalho, foi a presença do ISRj2 a uma distância de 24 pb do códon iniciador do gene que codifica a peptidase C58 (Tabela 2). Apesar deste gene estar presente também nos genomas das estirpes USDA 110, NK6, CPAC 15, E109 e USDA 6, tal evento de inserção ocorre exclusivamente em *B. diazoefficiens* CPAC 7 (Figura 12). A peptidase C58 pode ser amplamente encontrada em bactérias patogênicas e em bactérias simbióticas. Essa família de protease desempenha um papel central após a invasão da célula hospedeira e pode estar envolvida com os mecanismos de resistência [Hotson & Mudgett, 2004; Niño *et al.*, 2014]. Assim, é muito provável que esta inserção interfira na performance da estirpe CPAC 7 durante a simbiose, uma vez que seu homólogo (blr2118) é expresso 21 DAI [Delmotte *et al.*, 2010].

Gene/Impacto	Sequências de Inserção	Sequências de Inserção																												Ilha Simbiótica?										
		B. diazoefficiens CPAC 7	B. diazoefficiens USDA 110	B. diazoefficiens NK6	B. japonicum CPAC 15				B. japonicum CCBau 15354	B. japonicum CCBau 15517	B. japonicum CCBau 15618	B. japonicum CCBau 25435	B. japonicum CCBau 83623	B. japonicum E109	B. japonicum FN1	B. japonicum Is-1	B. japonicum Is-34	B. japonicum is5	B. japonicum USDA 122	B. japonicum USDA 123	B. japonicum USDA 124	B. japonicum USDA 135	B. japonicum USDA 22	B. japonicum USDA 38	B. japonicum USDA 4	B. japonicum USDA 6	B. japonicum USDA in8p8	B. elkanii 587	B. elkanii CCBau 05737		B. elkanii CCBau 43297	B. elkanii UASWS1015	B. elkanii USDA 3254	B. elkanii USDA 3259	B. elkanii USDA 76	B. elkanii USDA 94	B. elkanii WSM1741	B. elkanii WSM2783		
ABC transportador*	ISBd2																																							Sim
modC-Molibdênio ABC transportador	ISBd2																																						Sim	
Hipotético	ISBd3																																						Sim	
Hipotético	ISBd5																																						Sim	
Pilus tipo IV- montagem PilZ*	ISBd5																																						-	
Efetor Tipo III – nopAG*	ISBd5																																						Sim	
Transglutaminase*	ISBd5																																						-	
Hipotético	ISBd5																																						Sim	
Hipotético	ISBd9																																						Sim	
Hipotético	ISBd10																																						Sim	
PriA-Primosomal protein N	ISBd11																																						-	
Hipotético*	ISBd11																																						-	
TrbI -conjugação	ISBd12																																						Sim	
Hipotético	ISBd15																																						Sim	
Hipotético*	ISBd16																																						Sim	
Transportador	ISBd18																																						Sim	
Putativo regulador transcricional	ISBd20																																						-	
TrbI -conjugação	ISBd23																																						Sim	
Hipotético	ISBd23																																						Sim	
Hipotético	ISBd23																																						Sim	
Transglicosilase Lítica de Mureína*	ISBd23																																						-	
Efetor Tipo III - RhcU	ISBd23																																						Sim	
Metiltiotransferase/radical SAM*	ISBd23																																						Sim	
ISRj1	ISRj1																																						Sim	
TraG	ISRj1																																						Sim	
Hipotético*	ISRj1																																						Sim	
Efetor Tipo III - nopAN*	ISRj1																																						Sim	
Lisina 2,3-aminomutase*	ISRj1																																						-	
PilA2 – montagem de pilus	ISRj2																																						Sim	
Exonuclease SbcD	ISRj2																																						Sim	
Exonuclease – SbcD*	ISRj2																																						Sim	
Peptidase C58*	ISRj2																																						Sim	
HypB-hidrogenase*	ISRj2																																						Sim	
Hipotético	ISBja8																																						Sim	
Hipotético	ISBja8																																						Sim	

Figura 12: Comparação dos principais eventos de inserção identificados em outros genomas de *Bradyrhizobium*. **Verde:** genes interrompidos; **Vermelho:** IS presente na região promotora; **Amarelo:** IS presente na região terminadora; **Azul:** IS modificando ORF; **Cinza:** Ausência da inserção nos genomas de *Bradyrhizobium*; *: Homólogos expressos em condição de simbiose.

Estes dois genes discutidos aqui (*nopAN* e peptidase C58) estão localizados na ilha simbiótica. Além destes, também identificamos genes hipotéticos, genes envolvidos na montagem de pilus, ATPase, transportadores, uma transglicosilase lítica, transdutor de sinal, amidohidrolase, glucosiltransferase, entre outros (Tabela 2).

A presença de elementos ISs na região terminadora dos genes também é importante, uma vez que estes podem interferir na terminação da transcrição [Mitra *et al.*, 2010; Peters *et al.*, 2011]. Neste estudo, os casos identificados estão concentrados principalmente no códon de parada dos genes (5 entre os 7 casos). Portanto, destacamos aqui dois genes localizados na ilha simbiótica (*nopAG* e *rhcU*) (Tabela 2). O gene *nopAG*, efetor do tipo III, em CPAC 7 (BJA5080_08310) e em CPAC 15 (BJS_09091) estava previamente anotado como hipotético. A análise dos elementos ISs nos permitiu identificar que as DRs geradas pelo ISBd5 (Família IS1380) faz parte do códon de parada deste gene (Tabela 2). A transposase do ISBd5, inserido no *nopAG*, possui uma deleção de cerca de 470 nucleotídeos em comparação com o IS referência, portanto, sua transposase parece estar inativa. O mesmo foi observado no gene *rhcU* (sistema de secreção do tipo III), o qual possui o ISBd23 (Família IS5 ssgr IS427) inserido no seu códon de parada (Tabela 2).

Ambos eventos de inserção são conservados em outros genomas de *Bradyrhizobium*, além da CPAC 7 e CPAC 15 (Figura 12). Por isso, estudos futuros são requeridos para verificar se há alguma influência na simbiose, uma vez que estes genes participam de tal processo [Kimbrel *et al.*, 2013], pois foram identificados como diferencialmente expressos em condição de simbiose [Pessi *et al.*, 2007]. Conforme já mencionado, genes do sistema de secreção do tipo III e efetores estão envolvidos com o processo de infecção da célula hospedeira [Dale *et al.*, 2002; Alfano & Collmer, 2004].

A presença dos ISs no códon terminador dos genes pode ser explicada pela preferência por sítios alvos específicos que algumas famílias de elementos ISs possuem [Siguier *et al.*, 2015]. Como é o caso da família IS1380 (ISBd5), inserida na sequência "CTAG", caso também observado em *Acetobacter pasteurianus* NC11380. Tal família, também possui preferência pelos alvos "TCGG" e "TCGA" [Takemura *et al.*, 1991]. A família IS5 ssgr IS427, frequentemente prefere a sequência alvo "CTAG" [Mahillon & Chandler, 1998; Siguier *et al.*, 2015], que foi exatamente o que aconteceu com o ISBd23 (Família IS5 ssgr IS427) inserido no códon de parada do gene *rhcU* (Tabela 2).

Um outro tipo de impacto destacado neste trabalho foi um elemento IS modificando ORF, pela sobreposição de sequências quando em fitas opostas do DNA ou ainda, compartilhando sequência quando no mesmo sentido da fita de DNA. Entre as 13 ORFs encontradas nesta condição, 8 estão localizadas na ilha simbiótica. Tais ORFs compreendem hipotéticas, um putativo regulador transcricional, um transdutor de sinal, uma lisina 2,3-aminomutase, exonuclease, e os genes *modC* e *hypB* (Tabela 2). Nosso estudo destaca os genes *modC* e *hypB*, ambos estão na ilha simbiótica, que estão compartilhando sequência com um elemento IS.

Curiosamente, nós observamos que o gene *modC* compartilha 48 nucleotídeos com o ISBd2 (Família IS256). Esta sequência compartilhada está presente na proteína codificada pelo gene *modC* (Apêndice C). A transposase do ISBd2 em *B. diazoefficiens* CPAC 7 apresenta uma mutação *indel*, levando a um códon de parada prematuro. Enquanto que em *B. japonicum* CPAC 15, a transposase apresenta-se intacta, portanto com maior potencial de estar ativa. Este compartilhamento está presente em outros genomas de *Bradyrhizobium* (Figura 12), contudo ainda não havia sido relatado.

O gene *modC* codifica uma proteína que pode ter participação na formação do complexo ABC transportador, responsável pelo transporte de molibdênio para dentro da célula bacteriana, juntamente com os genes *modA* e *modB*. Sabendo da importância do molibdênio na síntese da nitrogenase, ressaltamos a importância de tal gene. Pessi *et al.* [2007] identificaram a expressão de *modB* em bacteroides. Contudo, não podemos afirmar que este gene (*modC*), em CPAC 7 (BJA5080_08107) e em CPAC 15 (BJS_08743) estão ativos. Uma vez que *B. diazoefficiens* USDA 110 possui três cópias do gene *modC* (bll1780, blr8162 e bll6953), bem como CPAC 7 e CPAC 15. Portanto, outros estudos são necessários para entendermos melhor o efeito deste evento de inserção.

O mesmo ocorre com o gene *hypB* que compartilha 50 nucleotídeos com o ISRj2 (Família IS3 ssgR IS150), um evento conservado nas estirpes CPAC 7, USDA 110, CPAC 15, E109 e USDA 6 (Figura 12). Este gene é parte do operon *hypABFCDE*, presente na ilha simbiótica, e o ISRj2 está inserido entre o *hypB* e *hypF*. Este operon pode estar envolvido na síntese da hidrogenase [Hansel *et al.*, 2001]. A atividade da hidrogenase aumenta a eficiência da fixação de nitrogênio nos nódulos de soja, o qual é evidenciado pelas altas taxas de N fixado na sua presença [Albrecht *et al.*, 1979; Bothe *et al.*, 2010]. Siqueira *et al.* [2014] mostraram que o sistema de obtenção de hidrogenase nas estirpes brasileiras

de *Bradyrhizobium* localizado na ilha simbiótica parece ser não-funcional, certamente pela presença de uma transposase no meio do operon. Contudo, seu homólogo em *B. diazoefficiens* USDA 110 (blr1732) é diferencialmente expresso 21 DAI [Pessi *et al.*, 2007], o que nos chama atenção uma vez que esse evento de inserção é conservado e o elementos ISRj2 parece estar compartilhando sequência e não interrompendo o gene. Por isso, só um estudo experimental focado nesse operon poderá nos informar se ele é mesmo não-funcional ou se o IS participa de alguma forma na sua regulação.

O compartilhamento de sequência entre genes e sequências de inserção que resulta em uma proteína funcional não parece ser um evento comum, não encontramos nenhum relato na literatura. Isso é muito interessante, já que tais genes parecem ter participação na fixação de nitrogênio e estes eventos serem conservados em alguns genomas de *Bradyrhizobium* (Figura 12).

Como podemos perceber, estudos voltados para os elementos de transposição podem nos acrescentar muita informação sobre os organismos de interesse. Zerillo *et al.* [2008] caracterizaram novos elementos ISs em duas espécies de *Leifsonia xyli*, sugerindo que tais elementos estão associados com a diversificação de *L. xyli* subsp. *xyli* e *L. xyli* subsp. *cynodontis*. Já o estudo de Bickhart *et al.* [2009], verificou que o conteúdo de elementos ISs reflete na plasticidade dos genomas das estirpes de *Frankia*, uma actinobacteria que nodula e fixa nitrogênio em plantas não-leguminosas.

Portanto, com este estudo conhecemos quais elementos estão presentes nestes dois genomas, além de apontarmos, pela primeira vez, várias evidências de que tais elementos podem influenciar no processo de simbiose soja-*Bradyrhizobium*, por isso, devem ser levados em consideração.

5.2 Descrição geral da análise das proteínas hipotéticas

A proposta deste estudo foi atualizar a anotação de genes considerados hipotéticos no momento da publicação dos genomas. Um total de 2.255 PHs de *B. diazoefficiens* CPAC 7 e 2.447 PHs de *B. japonicum* CPAC 15 foram analisadas. Estas proteínas encontravam-se anotadas como hipotéticas nos dois arquivos de anotação disponíveis para estes genomas (versão GenBank e RefSeq), e correspondem a menos de 30% do total de proteínas presentes em ambos. Esta observação é importante, uma vez que baseando-se apenas na anotação original, a versão publicada, apontava cerca de 50%

de hipotéticas nos genomas [Siqueira *et al.*, 2014], e no início do presente estudo já nos deparamos com uma porcentagem menor. Isso pode ter ocorrido na publicação dos genomas pelo fato da busca por homologia de função ter ocorrido utilizando o BLASTp ao invés de BLASTx, pois caso a ORF tenha sido predita em coordenada errada, o BLASTx consegue fazer a análise baseando-se nos seis diferentes quadros de leitura [McGinnis & Madden, 2004].

Mesmo assim, estes 30% representam uma limitação no conhecimento das funções presentes dentro de cada genoma, por isso, dispendemos esforços para adicionar o máximo de informação possível sobre tais proteínas. Tentando, principalmente apontar aquelas que possam ter envolvimento com a simbiose e/ou fixação de nitrogênio, para que possam ser estudadas experimentalmente.

Na Figura 13, observamos que a maioria das PHs são compartilhadas por ambas estirpes, mas ainda temos 917 PHs deste conjunto, presentes apenas em CPAC 7 e 1.101 PHs exclusivas de CPAC 15. Portanto, exploramos neste trabalho 3.361 proteínas diferentes.

Figura 13: Total de proteínas hipotéticas compartilhadas e exclusivas de *B. diazoefficiens* CPAC 7 e *B. japonicum* CPAC 15.

Com a estratégia de anotação adotada, das 3.361 proteínas diferentes, conseguimos atribuir função para 1.204, um incremento de informação para 35% das PHs. Finalmente, cabe notar que cerca de 340 das 2.157 proteínas que permanecem sem função atribuída, possuem similaridade com genes expressos em condição de simbiose, portanto, são alvos interessantes para estudos futuros.

No intuito de agregar maior valor às proteínas com função atualizada, buscamos por proteínas que tiveram similaridade com outras comprovadamente associadas ao processo de interação planta-

bactéria, através do banco de dados PHI-base [Urban *et al.*, 2015]. E ainda, buscamos por homólogos das proteínas deste trabalho em estudos de simbiose com *B. diazoefficiens* USDA 110. Tais estudos envolvem transcrito e proteoma 21 DAI e um transcrito 28 DAI em raízes de soja [Pessi *et al.*, 2007; Delmotte *et al.*, 2010; Franck *et al.*, 2014]. Além disso, também apontamos aquelas proteínas de CPAC 7 e CPAC 15 identificadas em condição de vida livre [Batista *et al.*, 2010; Gomes *et al.*, 2014]. Portanto, entre as 1.204 proteínas com anotação atualizada, 44 tiveram similaridade com proteínas de bactérias patogênicas (PHI-base), 215 tiveram seus homólogos diferencialmente expressos (DE) em condição de simbiose no estudo de transcrito (T: 21 DAI), 92 homólogos foram DE no estudo de proteoma (P: 21 DAI); nove homólogos DE 28 DAI (Transcrito) e 12 destas proteínas foram identificadas em condição de vida livre (Figura 14).

Figura 14: Diagrama de Venn mostrando que algumas proteínas anotadas tiveram similaridade com proteínas envolvidas na interação planta-bactéria. PHI: Pathogen-Host Interactions, T: estudo de transcrito, P: estudo de proteoma, DAI: Dias após a inoculação.

O interessante é que algumas proteínas estão presentes em mais de uma condição estudada, por exemplo, das 44 que tiveram similaridade com proteínas do banco de dados PHI-base, 10 são com-

partilhadas por outras condições (Figura 14). Das nove proteínas com homólogos no transcrito 28 DAI, nenhuma é específica desta condição, as mesmas aparecem também em outras condições (Figura 14). Nenhuma destas proteínas é compartilhada por todos estudos analisados, sendo que o estudo que mais possui homólogos com as proteínas anotadas é o transcrito de *B. diazoefficiens* USDA 110 21 DAI (Figura 14). Esta análise pode auxiliar na escolha de novos alvos para estudos experimentais futuros. Discutiremos a seguir algumas destas proteínas, sendo que a relação completa das proteínas com função predita e com similaridade com proteínas presentes nos estudos abordados está apresentada no Apêndice D.

5.2.1 Categorização funcional das proteínas que tiveram anotação atualizada

As 1.204 proteínas, agora com função predita, foram agrupadas em nove categorias funcionais: biossíntese geral; processos metabólicos; funções regulatórias; processos celulares; metabolismo de DNA; integrases, trasposases e fagos; transporte e proteínas de membrana; proteínas de ligação e outras. Na Figura 15, verificamos as proporções de cada categoria, e dentro de cada uma, quantas proteínas são compartilhadas e exclusivas de cada estirpe. A descrição detalhada dessas proteínas pode ser encontrada no Apêndice D.

Na classe biossíntese geral foram agrupadas um total de 65 proteínas (Figura 15). As quais podem estar associadas à biossíntese de lipídeos, de metabólitos secundários, de aminoácidos, de antibióticos, entre outros (Apêndice D). Nesta categoria, encontramos uma ent-caureno sintase (BJA5080_RS26465 e BJS_RS01535) e a uma ent-copalil difosfato sintase (BJA5080_RS26470 e BJS_RS01530), até então anotadas como hipotéticas nestes genomas. Tais proteínas estão presentes em outros genomas de rizóbios (*Bradyrhizobium*, *Sinorhizobium*, *Mesorhizobium* e *Rhizobium*) [Hershey *et al.*, 2014], além de *Xanthomonas oryzae* pv. *oryzicola* BLS256 (XOC_0077 e XOC_0078). Estas proteínas são diterpeno sintases e estão envolvidas na biossíntese de metabólitos secundários, sendo que a ent-caureno sintase teve sua estrutura e função caracterizada em *B. diazoefficiens* USDA 110 (blr2150) [Liu *et al.*, 2014], onde mostraram que sua estrutura é similar à estrutura do domínio α de diterpenos ciclases em plantas. Além disso, esta mesma proteína é considerada um fator de virulência em *Xanthomonas ory-*

Figura 15: Categorização funcional das 1.204 proteínas que tiveram função atualizada. Compartilhado: proteínas presentes em ambos genomas; CPAC 7 e CPAC 15 correspondem as proteínas exclusivas de cada estirpe.

zae pv. *oryzicola*, pois uma mutação na mesma causou uma redução da virulência desta bactéria [Lu *et al.*, 2015]. Estes dois genes foram identificados como diferencialmente expressos em *B. diazoefficiens* USDA 110 (blr2149 e blr2150) 21 e 28 DAI em soja, além de estarem presentes no PHI-base como fator de virulência em *Xanthomonas oryzae* pv. *oryzicola* (Apêndice D), portanto são proteínas importantes para o processo de interação planta-bactéria, tanto simbióticas quanto patogênicas.

Ainda na categoria de biossíntese geral, destacamos também a corismato mutase (EC: 5.4.99.5), uma enzima chave na biossíntese de aminoácidos aromáticos, converte corismato em prefenato. Seu papel é manter um balanço desses aminoácidos na célula, podendo ser encontrada em bactérias, fungos e plantas [Lee *et al.*, 1995; Schneider *et al.*, 2008]. A estrutura desta enzima varia em diferentes organismos, podendo ser considerada monofuncional, quando apresenta um sítio catalítico (AroH ou AroQ) ou bifuncional por apresentarem dois sítios catalíticos, formando uma outra classe de AroQ [Lee *et al.*, 1995; Schneider *et al.*, 2008]. O interessante é que dentro da classe AroQ, podemos ter aquelas que permanecem no citoplasma (monofuncionais), e aquelas que podem ser secretadas para

o periplasma (AroQ*), conforme identificado em *Salmonella typhimurium* (ou *Salmonella enterica*) e *Pseudomonas aeruginosa* [Calhoun *et al.*, 2001]. Tais corismato mutases secretadas podem estar associadas com a patogenicidade e virulência bacteriana, conforme apresentado em *Erwinia herbicola* [Xia *et al.*, 1993], em *Mycobacterium tuberculosis* [Sasso *et al.*, 2005] e em *X. oryzae* pv. *oryzicola* XKK.12 [Degrassi *et al.*, 2010].

Uma observação que nos chamou atenção neste trabalho é que a espécie *B. japonicum* possui duas corismato mutases diferentes, enquanto que bactérias da espécie *B. diazoefficiens* apenas uma. Tais proteínas pertencem à classe AroQ monofuncionais. A corismato mutase anotada neste estudo é compartilhada por ambas estirpes (BJA5080_RS34850 e BJS_RS32955), além de todos genomas de *Bradyrhizobium* disponíveis em bancos de dados, sendo DE 21 DAI em *B. diazoefficiens* [Pessi *et al.*, 2007; Delmotte *et al.*, 2010]. O que nos despertou interesse foi a corismato mutase monofuncional presente apenas em *B. japonicum* (BJS_RS42065), pelo fato de poderem ser um tipo de AroQ secretada e ainda participar da interação planta-bactéria, sendo que estudos posteriores serão necessários para entendermos exatamente seu o papel em *B. japonicum*. Uma vez que tais proteínas exclusivas, quando comparadas com duas AroQ* caracterizadas [Xia *et al.*, 1993; Calhoun *et al.*, 2001] possuem conservados quatro dos cinco sítios catalíticos (38, 49, 58, 62,) desta proteína (Figura 16).

O grupo processos metabólicos engloba o maior número de proteínas anotadas (272). Aqui estão presentes entre outras, proteínas envolvidas no metabolismo central, no metabolismo de ácidos graxos, metabolismo de aminoácidos, enzimas dehidrogenases, ATPases, glicosil hidrolases e glicosil transferases (Figura 15). Destacamos as glicosil hidrolases (GHs) e as glicosil transferases (GTs). As GHs são enzimas amplamente distribuídas que hidrolisam ligações glicosídicas, possuindo diversas funções, sendo que tais enzimas podem participar da interação planta-bactéria, remodelando ou degradando a parede celular [Faure, 2002; Izquierdo *et al.*, 2010]. Conforme apresentado no Apêndice D, cinco das GHs identificadas tiveram seus homólogos expressos em condição de simbiose [Pessi *et al.*, 2007; Delmotte *et al.*, 2010; Franck *et al.*, 2014].

As GTs constituem uma família de enzimas que catalisam a transferência de açúcares ativados para uma variedade de moléculas aceptoras, sendo de extrema importância na síntese de carboidratos e glicoconjugados [Breton *et al.*, 2006; Zhu *et al.*, 2013; Gloster, 2014]. Essas reações de glicosilação

Eh ATCC33243	1	MTHFVAIFSSLFMC SNVFAGSVSSVSLGSLSSALNERMQVMKAVAGYALHHPEDLP	60
Se ATCC15277	1	MIRHIAIFLCSLMCSTTFADSVTSVSLGALLTALNERMLMKDVAAYMKHLPEDFT	60
Bj ⁻ CPAC15	1	MTDDDDQEFASDLARL----RRSIDN-LDAAMVHLLAERFRCTDAIGELKARHGLESDPR	55
Bj ⁻ USDA6	1	MTDDDDQEFASDLARL----RRSIDN-LDAAMVHLLAERFRCTDAIGELKARHGLESDPR	55
		* * * * *	
Eh ATCC33243	61	REOVVLDHMLQNAQQAGLEPHSVEPVFV----HALMNASKTIQYRYRADWLSSPDSAVPVR	116
Se ATCC15277	61	REONVFAEAEELAKNGLDPHSITPFI----RSLMDASKAIQYRYLAQWRTGSEPSFPIQ	116
Bj ⁻ CPAC15	56	REAAQLKRLRLIAAQSGLDPDFTRKLHALVVREVITNHEAVKFESGAG-----	103
Bj ⁻ USDA6	56	REAAQLKRLRLIAAQSGLDPDFTRKLHALVVREVITNHEAVKFESGAG-----	103
		** : . * : ** : : : : *	
Eh ATCC33243	117	DLTETRQQIQQLDTQLLTAISQRLMTGAFSQEDKEFLMSHLTAPHLSKNSLFASLSR	176
Se ATCC15277	117	TL SVTRQRIRQLDNQMLIIISQRLMVGAFSHDDMVWLRQFNAPNLNESDISNVLAALS	176
Bj ⁻ CPAC15	104	-----	103
Bj ⁻ USDA6	104	-----	103
Eh ATCC33243	177	IQRQH	181
Se ATCC15277	177	VRRAR	181
Bj ⁻ CPAC15	104	-----	103
Bj ⁻ USDA6	104	-----	103

Figura 16: Alinhamento da corismato mutase secretada (AroQ*) de *Erwinia herbicola* e *Salmonella enterica* com corismato mutases exclusivas de *Bradyrhizobium japonicum*. Em vermelho destacamos os cinco sítios catalíticos, os quais quatro são conservados nas estirpes CPAC 15 e USDA 6. Eh: *Erwinia herbicola*; Se: *Salmonella enterica*; Bj: *Bradyrhizobium japonicum*.

são essenciais para vários processos biológicos como adesão, sinalização e biossíntese da parede celular [Breton *et al.*, 2006; Zhu *et al.*, 2013]. Pelo fato de muitas glicoproteínas bacterianas serem fatores de virulência de bactérias patogênicas, as GTs possuem um papel importante no processo de infecção [Zhu *et al.*, 2013; Jank *et al.*, 2015]. Dentre as que anotamos nas estirpes de *Bradyrhizobium*, algumas apresentaram similaridade com GTs de bactérias patogênicas (Apêndice D).

As proteínas agrupadas na categoria de funções regulatórias podem ter papéis importantes na interação planta-bactéria. Das 101 proteínas, a maioria corresponde a proteínas da estirpe CPAC 15 (Figura 15). Dentre elas, encontramos principalmente, adenilato ciclases, diguanilato ciclases, vários reguladores transcricionais e proteínas quinases (Apêndice D). Adenilato ciclase é uma enzima com função regulatória essencial na célula, catalisam a conversão do ATP (adenosina trifosfato) em cAMP (adenosina monofosfato cíclico) [Ahuja *et al.*, 2004; Lory *et al.*, 2004]. cAMP é uma das mais abundantes moléculas de sinalização nas células, responsável por controlar inúmeros processos regulatórios e metabólicos, tanto em eucariotos como em procariotos [Ahuja *et al.*, 2004; Lory *et al.*, 2004]. Mais especificamente em bactérias, algumas adenilato ciclases podem estar envolvidas na

interação hospedeiro-microrganismo, tanto no processo de patogenicidade, quanto no processo simbiótico [Terakado *et al.*, 1997; Ahuja *et al.*, 2004; Lory *et al.*, 2004; McDonough & Rodriguez, 2011]. Por exemplo, na simbiose entre *Sinorhizobium meliloti* e *Medicago*, onde verificou-se que a adenilato ciclase participa do processo de infecção [Tian *et al.*, 2012].

Outra proteína regulatória que destacamos é a diguanilato ciclase, importante molécula de sinalização, podendo participar de vários processos biológicos, incluindo virulência, motilidade, adesão, secreção e formação de biofilme [Newell *et al.*, 2011; Whiteley & Lee, 2015]. Mais especificamente na simbiose, podem estar envolvidos na motilidade e formação de exopolissacarídeos, como mostrado em *S. meliloti* [Schäper *et al.*, 2015].

Ainda em funções de regulatórias, também apontamos vários reguladores transcricionais, que são proteínas que se ligam ao DNA regulando a expressão gênica, sendo portanto essenciais para a maioria dos processos biológicos, inclusive no processo de simbiose [Janczarek & Skorupska, 2007; Balleza *et al.*, 2009], ainda mais pelo fato de que muitos desses reguladores tiveram similaridade com proteínas expressas em condições de simbiose. Além disso, anotamos várias histidinas quinases, as quais participam do processo de transdução de sinais, uma via de sinalização bastante comum no processo de simbiose [Kistner & Parniske, 2002; Popp & Ott, 2011], o que também desperta interesse.

Proteínas de ligação são as proteínas que se ligam a outras moléculas, sendo representadas neste estudo por 52 proteínas, as quais destacamos os tetratricopeptídeos. Repetições tetratricopeptídeos (TPRs) são motivos estruturais presentes em uma ampla gama de proteínas, responsáveis por mediar a interação proteína-proteína, podendo ser identificados em proteínas presentes em procariotos e eucariotos [Blatch & Lässle, 1999; Zeytuni & Zarivach, 2012]. Mais especificamente em bactérias, TPRs têm se mostrado essenciais em muitas vias, como montagem de membrana externa e patogenicidade [Zeytuni & Zarivach, 2012; Cervený *et al.*, 2013], por isso estudar essas proteínas vem sendo tão importante.

Já se sabe que as TPRs estão diretamente envolvidas em funções associadas a virulência em bactérias patogênicas, como na translocação de fatores de virulência e adesão bacteriana nas células hospedeiras [Edqvist *et al.*, 2006; Cervený *et al.*, 2013]. Portanto, pelo fato de poderem participar de vários processos biológicos nas bactérias, por terem participação no processo de patogenicidade e

ainda por serem expressos em condições de simbiose em *Bradyrhizobium* (Apêndice D), estas proteínas representam interessantes e potenciais alvos para novos estudos, uma vez que sua participação na simbiose não é comumente descrita. Na Figura 17 apresentamos uma representação da estrutura de três TPRs presentes em CPAC 7 e CPAC 15, cujos homólogos foram identificados em condições de simbiose. Tais sequências apresentam diferenças entre as duas estirpes como podemos observar pela porcentagem de identidade, pelo tamanho e pela posição das repetições previstas.

Figura 17: Estrutura de três tetratricopeptídeos identificados em *B. diazoefficiens* CPAC 7 e *B. japonicum* CPAC 15 que tiveram seus homólogos expressos em condição de simbiose. **TPR 1:** posição das repetições igual em ambas estirpes; **TPR 2 e TPR 3:** posição das repetições é diferente as estirpes. Porcentagem de identidade baseada na sequência de aminoácido.

Em metabolismo de DNA temos 94 proteínas, as quais podem estar envolvidas no processo de reparo de DNA, e ainda nos processos de replicação, transcrição e tradução (Figura 15). Também anotamos 69 proteínas relacionadas a integrases, transposases e fagos (Figura 15). No Apêndice D estão estas proteínas, as quais contribuem para aumentar o conhecimento sobre conteúdo gênico destes genomas.

A absorção de nutrientes e a interação do microrganismo com o meio ambiente dependem, entre

outros, de proteínas envolvidas no transporte e também de proteínas integrais de membrana (internas e externas) [Burkovski & Krämer, 2002]. Além disso, as proteínas de membrana estão associadas com os processos de sinalização e reconhecimento [Bernstein, 2000; Dalbey *et al.*, 2011]. Do conjunto de proteínas que tiveram função atribuída, identificamos 159 proteínas que podem estar relacionadas a estes processos (Apêndice D). Dentre elas, destacam-se principalmente os ABC transportadores, transportadores MFS (*Major Facilitator Superfamily*), proteínas de membrana em geral, proteínas de membrana externa, entre outras (Figura 15).

A categoria de processos celulares será abordada a seguir, pois corresponde a categoria de maior destaque neste trabalho, já que envolvem proteínas que podem estar diretamente relacionadas com o processo de interação planta-bactéria.

Na categoria outras estão 139 proteínas que não se encaixaram nas demais (Figura 15). Destacamos aqui uma proteína anotada como possível nitrogenase, específica de CPAC 15 (BJS_RS31000), a qual possui 42% de similaridade com uma nitrogenase (I0G8B6) predita em *Bradyrhizobium* sp. S23321, esta proteína não está localizada na ilha simbiótica e parece ser específica desta estirpe, ausente em outros genomas de *B. japonicum* e *B. diazoefficiens*.

Outra proteína interessante em destaque é específica de CPAC 7 (BJA5080_RS08765), e pela busca em banco de dados, presente apenas em genomas de *B. diazoefficiens*. Esta proteína possui homologia com o domínio da proteína NodB segundo o Interpro (IPR002509), o que é extremamente interessante, pois NodB está envolvida na síntese dos fatores Nod, sinais que os rizóbios enviam para a planta hospedeira para que ocorra o estabelecimento da simbiose [John *et al.*, 1993]. Este gene que identificamos é diferencialmente expresso 21 DAI em *B. diazoefficiens* USDA 110 (bll3646) [Pessi *et al.*, 2007], por isso seria interessante estudar mais profundamente se sua presença só em CPAC 7 pode influenciar a simbiose da CPAC 15. Nessa mesma categoria também encontram-se duas proteínas, ambas presentes nas duas estirpes brasileiras (BJA5080_RS00595 e BJA5080_RS35985 em CPAC7, BJS_RS44125 e BJS_RS43425 em CPAC 15), anotadas como proteínas de nodulação, apresentaram cerca de 56% de similaridade com a proteína NopA (C4ALD0) de *Rhizobium* sp. NGR234, uma possível proteína de secreção do tipo III [Deakin *et al.*, 2005].

5.2.2 Processos celulares: proteínas com potencial envolvimento no processo de interação planta-bactéria

Além das várias proteínas importantes identificadas e já citadas neste trabalho, as quais podem ter participações no processo de interação planta-bactéria, destacamos aqui a categoria de processos celulares, na qual estão concentradas proteínas que podem estar diretamente relacionadas ao processo de colonização e infecção do microrganismo na planta. Esta categoria conta com um total de 254 proteínas, a segunda categoria mais representativa (Figura 15). Sendo que 130 proteínas são compartilhadas por ambas bactérias, 46 específicas da estirpe CPAC 7 e 78 de CPAC 15 (Apêndice D). Tais proteínas estão associadas, principalmente, com os processos de conjugação, motilidade e adesão, fatores de virulência e anti-toxinas, proteínas secretoras, proteínas efetoras, estresse oxidativo, proteases e peptídeos sinais (Figura 18-a).

Figura 18: Proteínas com funções pertencentes à categoria processos celulares. **a-** Categorização das 254 proteínas presentes em processos celulares; **b-** Diagrama de Venn mostrando o número das proteínas de processos celulares que apresentaram similaridade com proteínas envolvidas na interação planta-bactéria.

Sabe-se que tanto na interação patogênica, quanto na simbiótica, a bactéria deve possuir habilidades para reconhecer e chegar até seu hospedeiro, aderindo-se à ele. E na sequência, é preciso ainda que ocorra a infecção para que finalmente, o microrganismo possa estabelecer um nicho e se multiplicar dentro do hospedeiro [Jones *et al.*, 2007; Soto *et al.*, 2009]. Em ambos tipos de interações (patogênica e simbiótica), as bactérias são percebidas como intrusas pelas plantas, as quais utilizam estratégias de

infecção similares [Hentschel *et al.*, 2000; Soto *et al.*, 2006, 2009]. Contudo, os rizóbios possuem a capacidade de suprimir a resposta de defesa de seu hospedeiro específico [Spaink, 2000; Jones *et al.*, 2007; Luo & Lu, 2014].

Portanto, entre estas proteínas, até então consideradas hipotéticas, identificamos várias que podem ser importantes em tais processos. Sendo que, das 254 proteínas anotadas nesta categoria, 76 tiveram similaridade com proteínas associadas à interação planta-bactéria. E ainda, são compartilhadas por mais de um estudo, por exemplo, identificadas em bactérias patogênicas e em condição de simbiose (PHI-base e 21 DAI), ou ainda em condição de simbiose e em vida-livre (Figura 18-b). Dentro desta categoria não houve proteínas com homólogos expressos 28 DAI.

Assim, além de uma visão geral das proteínas anotadas, selecionamos algumas proteínas que representam potenciais alvos para estudos experimentais e que podem estar diretamente ligadas ao processo de interação simbiótica em estudo (Tabela 3). A lista completa das proteínas desta categoria encontra-se no Apêndice D.

As proteínas relacionadas à conjugação representam 4% das proteínas da categoria processos celulares (Figura 18-a). Conjugação é o processo pelo qual ocorre a transferência de material genético entre células bacterianas através do contato direto. Essa transferência se dá através do pilus, um apêndice filiforme que pode ser encontrado na superfície de muitas bactérias [Sauer *et al.*, 2000; Proft & Baker, 2009]. Além da conjugação, o pilus é uma estrutura muito importante pois auxilia na adesão ao hospedeiro, participam da formação de biofilmes e da motilidade da célula [Sauer *et al.*, 2000; Proft & Baker, 2009].

Proteínas envolvidas com a motilidade e adesão das bactérias são essenciais para que as mesmas entrem em contato com a planta hospedeira e deem início à troca de sinais moleculares entre eles [Hentschel *et al.*, 2000; Soto *et al.*, 2006], e das proteínas agrupadas em processos celulares, 4% as representam (Figura 18-a). A motilidade dos rizóbios se dá, principalmente, pela presença do flagelo, no processo de quimiotaxia até a planta [Subramanian *et al.*, 2007; Bernabéu-Roda *et al.*, 2015]. E entre as proteínas selecionadas, temos 2 proteínas flagelares anotadas em ambas estirpes (Tabela 3), que tiveram seus homólogos DE em 21 DAI [Pessi *et al.*, 2007]. Além disso, também identificamos uma adesina específica de CPAC 7 (BJA5080_RS18675) e uma adesina específica de

CPAC 15 (BJS_RS26170), ambas tiveram similaridade com adesinas de bactérias patogênicas (Tabela 3). As adesinas são proteínas especializadas que atuam na aderência bacteriana na superfície das células hospedeiras [Niemann *et al.*, 2004; Sanchez, 2011]. Estas proteínas são bastante estudadas nas bactérias patogênicas, e em bactérias simbióticas sua importância para o estabelecimento simbiótico também vem sendo demonstrada [Smit *et al.*, 1989; Niemann *et al.*, 2004; Mongiardini *et al.*, 2008; Nigmatullina *et al.*, 2015].

Os fatores de virulência e anti-toxinas representam 8% do total anotado (Figura 18-a). Dentre os fatores de virulência, alguns são compartilhados e outros específicos de cada estirpe (Tabela 3). Tais proteínas são comumente encontradas em bactérias patogênicas e contribuem para a colonização do hospedeiro [Casadevall & Pirofski, 2001; Soto *et al.*, 2006], mas também podem ser encontradas em rizóbios [Van Sluys *et al.*, 2002; Soto *et al.*, 2006]. A antitoxina (BrnA), presente em ambas estirpes, cujo seu homólogo (bsl4681) é expresso em condição de simbiose (Tabela 3), parece estar relacionada com a adaptação ao estresse conforme descrito em *Brucella abortus* [Heaton *et al.*, 2012]. Talvez fosse interessante entendermos sua função no estabelecimento simbiótico.

Proteínas secretoras contam com 6% das 254 proteínas e as proteínas efetoras com 7% (Figura 18-a). O envolvimento e a importância das mesmas para o estabelecimento simbiótico são bastante conhecidos [Soto *et al.*, 2006; Deakin & Broughton, 2009; Nelson & Sadowsky, 2015]. Apesar de sua relevância, vimos neste trabalho que muitas permaneciam anotadas como hipotéticas, ressaltando mais uma vez a importância de explorá-las. Conseguimos adicionar informações à várias destas importantes proteínas, as quais participam potencialmente da simbiose (Tabela 3). As proteínas de secreção, principalmente aquelas do sistema de secreção do tipo III e tipo IV, são responsáveis por transportar proteínas efetoras diretamente para o citoplasma do hospedeiro ou para o meio externo, as quais auxiliam no processo de infecção [Hentschel *et al.*, 2000; Deakin & Broughton, 2009; Soto *et al.*, 2009; Costa *et al.*, 2015]. Tais proteínas efetoras ao entrarem em contato com a célula hospedeira, interferem em seus processos biológicos, beneficiando o microrganismo, por exemplo, suprimindo a resposta de defesa da planta [Gürlebeck *et al.*, 2006; Kay & Bonas, 2009; Okazaki *et al.*, 2010]. Assim mostramos a importância das proteínas efetoras encontradas neste estudo, as quais correspondem principalmente a efetores do tipo III.

Tabela 3: Proteínas apontadas como possíveis alvos de estudos experimentais, com potencial envolvimento no processo de interação planta-bactéria.

Proc. Celulares	Proteína	Estirpes de <i>Bradyrhizobium</i>		PHI-base		T: 21 DAI	P: 21 DAI
		CPAC 7	CPAC 15	Organismo	Fenótipo	ID	ID
Motilidade e adesão	Proteína flagelar	BJA5080_RS14670	BJS_RS09410	-	-	blI4781	-
	Proteína flagelar	BJA5080_RS20870	BJS_RS22765	-	-	blI5823	-
	Adesina	BJA5080_RS18675	-	<i>B. cenocepacia</i>	Red. da virul.	-	-
	Adesina	-	BJS_RS26170	<i>V. vulnificus</i>	Red. da virul.	-	-
Fatores de virul. e Anti-toxinas	BA14k - fator de virulência	BJA5080_RS36865	BJS_RS07750	-	-	-	-
	BA14k - fator de virulência	-	BJS_RS25240	-	-	-	-
	BrnA antitoxina	BJA5080_RS14165	BJS_RS09945	-	-	bsI4681	-
	Fator de virulência	BJA5080_RS18060	-	-	-	-	-
	Virulence factor BrkB	BJA5080_RS19330	BJS_RS11675	-	-	-	-
	Fator de virulência E	BJA5080_RS34075	-	-	-	-	-
	Fator de virulência E	BJA5080_RS40845	-	-	-	-	-
Proteínas secretoras	Secretora rica em cisteína	BJA5080_RS35335	BJS_RS42825	-	-	blr1486	-
	Proteína de secreção	BJA5080_RS18820	BJS_RS06110	-	-	blr5567	-
	Proteína de secreção IV	BJA5080_RS27455	BJS_RS43615	-	-	-	-
	Proteína de secreção	-	BJS_RS23720	<i>X. fastidiosa</i>	Aum. da virul.	-	-
	Proteína de secreção I	-	BJS_RS20395	-	-	-	-
Proteínas efetoras	Efector tipo III	BJA5080_RS26815	BJS_RS01075	-	-	bsI2050	-
	Efector tipo III	BJA5080_RS00190	BJS_RS00110	-	-	blr1869	-
	Efector tipo III	BJA5080_RS27535	BJS_RS00010	-	-	blI1840	-
	Efector tipo III	BJA5080_RS27565	BJS_RS00040	-	-	blI1848	-
	Efector tipo III	BJA5080_RS27615	BJS_RS40610	-	-	blI1862	-
	Efector tipo III	BJA5080_RS30960	-	-	-	blI1840	-
	Efector tipo III - LRR	BJA5080_RS27425	BJS_RS43585	<i>Y. pestis</i>	Avirulência	blr1676	blr1676
	Proteína secretada	BJA5080_RS24460	BJS_RS27130	-	-	blI6609	-
	Proteína secretada	BJA5080_RS15340	BJS_RS08715	-	-	bsI4900	-
	Proteína secretada	BJA5080_RS30785	BJS_RS37205	-	-	-	blr0628
Estresse oxidativo	Domínio-AhpD	BJA5080_RS28755	-	-	-	-	blI0235
	Alquil hidroperóxido reductase	BJA5080_RS10270	BJS_RS14505	-	-	-	blr3933
	Peroxidase	BJA5080_RS13455	BJS_RS11120	-	-	blr4509	-
Proteases	Metalopeptidase	BJA5080_RS02485	BJS_RS03160	-	-	blI2452	-
	Peptidase A26	-	BJS_RS01215	-	-	blI2085	-
	Peptidase c14	BJA5080_RS30775	-	-	-	blI0626	-
	Protease	-	BJS_RS27500	<i>M. oryzae</i>	Red. da virul.	-	-

Tabela 3 – Continuação

Proc. Celulares	Proteína	Estirpes de <i>Bradyrhizobium</i>		PHI-base		T: 21 DAI	P: 21 DAI
		CPAC 7	CPAC 15	Organismo	Fenótipo	ID	ID
	Cisteína peptidase	BJA5080_RS02040	BJS_RS02705	-	-	bll2364	-
	Ulp1 protease	BJA5080_RS00155	BJS_RS40430	-	-	-	bll8244
	ClpP/crotonase	BJA5080_RS15195	-	-	-	blr4872	-
Outros	Fasina	BJA5080_RS16605	BJS_RS07560	-	-	-	bll5155
	Proteína PEP2	BJA5080_RS39765	BJS_RS32725	<i>N. haematococca</i>	Red. da virul.	bll7793	bll7793
	Domínio RmlC	BJA5080_RS35465	BJS_RS07705	-	-	blr1513	-
	Domínio RmlC	BJA5080_RS09405	BJS_RS15525	-	-	bll3763	-
Peptídeo sinal*	Peptídeo sinal	BJA5080_RS24225	BJS_RS26870	-	-	blr6562	-
	Peptídeo sinal	BJA5080_RS03305	BJS_RS21015	-	-	blr2607	-
	Peptídeo sinal ²	BJA5080_RS36065	BJS_RS29175	-	-	bll6993	-
	Peptídeo sinal	BJA5080_RS04670	BJS_RS13135	-	-	bll2824	-
	Peptídeo sinal	BJA5080_RS05955	BJS_RS18285	-	-	blr3076	-
	Peptídeo sinal	BJA5080_RS38600	BJS_RS31410	-	-	-	blr7465
	Peptídeo sinal ²	BJA5080_RS38950	BJS_RS31770	-	-	blr7534	blr7534
	Peptídeo sinal	BJA5080_RS14345	BJS_RS09775	-	-	blr4716	-
	Peptídeo sinal	BJA5080_RS18345	BJS_RS06470	-	-	bll5501	-
	Peptídeo sinal ²	BJA5080_RS18655	BJS_RS06265	-	-	bll5535	-
	Peptídeo sinal ¹	BJA5080_RS18940	BJS_RS05980	-	-	bll5589	-
	Peptídeo sinal	BJA5080_RS36555	BJS_RS29560	-	-	bll7080	-
	Peptídeo sinal	BJA5080_RS30640	BJS_RS37355	-	-	bll0598	-
	Peptídeo sinal	BJA5080_RS30945	BJS_RS37050	-	-	bll0660	-
	Peptídeo sinal	BJA5080_RS31070	BJS_RS36960	-	-	bll0679	bll0679
	Peptídeo sinal	BJA5080_RS33465	BJS_RS34570	-	-	blr1130	-
	Peptídeo sinal	BJA5080_RS36985	BJS_RS29985	-	-	-	blr7166
	Peptídeo sinal	BJA5080_RS20025	BJS_RS21815	-	-	bll5657	-
	Peptídeo sinal ²	BJA5080_RS10860	BJS_RS13830	-	-	blr4051	-
	Peptídeo sinal	BJA5080_RS21095	BJS_RS22975	-	-	-	blr5868
	Peptídeo sinal	BJA5080_RS24305	BJS_RS26970	-	-	bll6577	bll6577
	Peptídeo sinal	BJA5080_RS11465	-	-	-	blr4174	-
	Peptídeo sinal	BJA5080_RS01030	-	-	-	bll6987	-
	Peptídeo sinal ³	BJA5080_RS01215	-	-	-	blr4262	-

*Peptídeo sinal: sequência de aminoácidos localizada na porção N-terminal das proteínas./ Localização celular predita para os peptídeos sinais: **1-** membrana plasmática; **2-** periplasma; **3-** membrana externa.

Na Tabela 3 ressaltamos também três proteínas dentre as relacionadas com o estresse oxidativo, as quais constituem 7% das proteínas agrupadas em processos celulares (Figura 18-a). Essa categoria é bastante importante, pois durante a simbiose ocorre uma explosão oxidativa na planta, que pode comprometer o processo de simbiose [Santos *et al.*, 2000; Davies & Walker, 2007]. Notamos ainda que, a proteína AhpD (BJA5080_RS28755) está presente apenas em *B. diazoefficiens* CPAC 7, tendo sido identificada como DE 21 DAI em *B. diazoefficiens* USDA 110 [Delmotte *et al.*, 2010], o que representa mais uma diferença entre as estirpes brasileiras.

Da mesma forma, identificamos várias proteases (19%), ou também chamadas de peptidases (Figura 18-a), onde algumas podem ser consideradas fatores de virulência (Apêndice D), pois quando secretadas para dentro da célula do hospedeiro, orquestram a modificação de proteínas do mesmo, proporcionando para a bactéria maior tolerância às condições adversas, como por exemplo, às respostas de defesa [Xia, 2004; Frees *et al.*, 2013]. As proteases podem ser classificadas em cinco grandes classes: serina, cisteína, aspartato, treonina e metaloprotease [Barrett, 1994; Xia, 2004]. Algumas proteases identificadas (Tabela 3), também parecem ser de grande importância nesta interação simbiótica, pois já foram identificadas nesta condição na estirpe USDA 110 [Pessi *et al.*, 2007; Delmotte *et al.*, 2010].

Em processos celulares identificamos proteínas que foram agrupadas como "Outros"(13%), os quais estão relacionados ao ciclo e divisão celular, ao estresse, entre outros (Apêndice D). Na Tabela 3 destacamos algumas proteínas, que tiveram similaridade com outras proteínas associadas à interação bactéria-hospedeiro. Entre essas, uma nos despertou particular atenção, a proteína PEP2 (do inglês, *Pea Pathogenicity Protein 2*), presente em ambas estirpes brasileiras, identificada como DE 21 DAI e que ainda possui cerca de 36% de identidade e 90% de cobertura com a proteína PEP2 de *Nectria haematococca*. A PEP2 foi caracterizada em *Nectria haematococca*, sendo considerada como determinante na patogenicidade deste fungo na planta *Pisum sativum* (ervilha), causando a podridão da raiz [Han *et al.*, 2001]. Esta proteína parece ter sido adquirida pelo fungo via transferência horizontal [Han *et al.*, 2001; Liu *et al.*, 2003], e neste estudo a apontamos como um potencial alvo de estudo posterior, a fim de entendermos sua função em *Bradyrhizobium*.

Finalmente, destacamos as proteínas que possuem peptídeo sinal em sua sequência de aminoá-

cidos, as quais correspondem a 32% do total de proteínas desta categoria (Figura 18-a). Peptídeo sinal refere-se a uma sequência de aminoácidos presente, geralmente, na região N-terminal da maioria das proteínas recém-sintetizadas que serão destinadas às vias de secreção [Martoglio & Dobberstein, 1998; Hegde & Bernstein, 2006; Filloux, 2010; Ivankov *et al.*, 2013]. Podemos observar na Tabela 3, que para alguns peptídeos sinais foi possível prever sua localização celular (membrana plasmática, periplasma ou membrana externa), mas para a maioria essa informação ainda é desconhecida. Essas proteínas que possuem peptídeos sinais nos despertaram interesse pois tiveram seus homólogos expressos em condição de simbiose na estirpe USDA 110 (Tabela 3). Entretanto, novos estudos são necessários para caracterizá-las funcionalmente.

Diante do apresentado, destacamos a grande relevância em se estudar as proteínas hipotéticas, pois atualmente as informações nos banco de dados são atualizadas de forma rápida e novos conhecimentos são acrescentados. Como podemos verificar aqui, mesmo com pouco tempo (cerca de 2 anos) entre a publicação dos genomas das estirpes brasileiras de *Bradyrhizobium* [Siqueira *et al.*, 2014] e o presente estudo, várias novas e importantes informações ao conteúdo proteico destes organismos foram incluídas, ressaltando aquelas relacionadas com a simbiose. Esse tipo de estudo tem também o papel de minerar as informações e apontar caminhos para novas pesquisas. Por exemplo, agora as proteínas, até então com função desconhecida, poderão ser pontualmente estudadas em novos estudos experimentais.

Lembrando que, mesmo com todo empenho, ainda temos proteínas hipotéticas nos genomas. Antes deste estudo, cada bactéria possuía menos de 30% de PHs em relação ao total de proteínas dos genomas, e agora diminuimos essa porcentagem para cerca de 18% de PHs restantes (CPAC 7: 1.460 e CPAC 15: 1.602).

5.3 Discriminando regiões atípicas com a máxima entropia

O presente estudo propõe o uso da máxima entropia como uma nova estratégia para a exploração individual de genomas bacterianos, com enfoque nas regiões atípicas, as quais podem ser priorizadas durante as análises, pois são as regiões com maior chance de se encontrar características exclusivas

de cada organismo. Além disso, a metodologia baseada na ME tem a proposta de oferecer rapidez na execução das análises, sem a necessidade, a priori, de arquivos de anotação, somente a sequência do genoma no formato FASTA.

Tal estratégia, consiste em escanear o genoma em tamanhos de regiões pré-determinadas, calculando a frequência de oligonucleotídeos (2, 4 e 6) de cada região. Cada região analisada recebe um valor ME, os quais são visualizados em um gráfico. A distribuição destes valores ao longo do genoma seria homogênea se a distribuição de oligonucleotídeos não tivesse variação entre as regiões. Como genomas bacterianos são muito dinâmicos [Koonin *et al.*, 2001; Darmon & Leach, 2014], tais variações podem significar informações novas sobre o conteúdo do genoma. Por isso, regiões com valores $ME \geq Q_3$ são consideradas regiões candidatas à regiões atípicas previstas pelo método. Portanto, as regiões que devem ser exploradas na busca por características específicas.

No intuito de avaliarmos quais seriam os melhores parâmetros a serem utilizados, nós aplicamos a máxima entropia em dois genomas de *Xanthomonas* (XAC e XCC) com diferentes tamanhos de regiões e de oligonucleotídeos, os resultados estão apresentados na Figura 19.

Figura 19: Avaliação dos parâmetros (tamanho das regiões e tamanho dos oligonucleotídeos) a serem utilizados na identificação de regiões atípicas de genomas bacterianos baseado na máxima entropia. **a-** ME aplicada ao genoma de *X. axonopodis* pv. *citri* 306; **b-** ME aplicada ao genoma de *X. campestris* pv. *campestris* ATCC 33913. SENS: sensibilidade; ESP: especificidade; EF: eficiência.

A identificação correta das regiões atípicas em genomas bacterianos depende naturalmente das medidas adotadas e do critério discriminante. Por essa razão, diferentes métodos vêm sendo desen-

volvidos na tentativa de melhorar as predições [Azad & Lawrence, 2011].

Como podemos observar, com os parâmetros testados na metodologia da máxima entropia, tanto para tamanho de regiões quanto para tamanho do oligonucleotídeo, o F-score e a eficiência da predição se mantiveram relativamente constantes, principalmente quando utilizamos regiões de 2 a 10kb, um importante resultado da metodologia proposta (Figura 19). Somente quando empregamos tamanhos de regiões de 50kb, observamos um pequeno decréscimo da eficiência e do F-score (Figura 19). A medida de eficiência foi obtida a partir da sensibilidade e especificidade, pela necessidade em se encontrar um equilíbrio entre elas, já que a sensibilidade e a especificidade caminham em direções opostas (Figura 19).

Por outro lado, quando consideramos a análise visual dos gráficos, aquelas baseadas em menores tamanhos de regiões perdem resolução, conforme exemplificado com o genoma de *X. campestris* pv. *campestris* ATCC 33913 no Apêndice E. Uma alternativa é realizar uma repetição de análises alternando os parâmetros para definir o que melhor se adapta ao conjunto de dados, uma vez que a ME não apresentou grande perda de eficiência (Figura 19).

A metodologia ME também pode ser considerada uma opção computacional rápida para a análise de genomas bacterianos, pois levou cerca de sete minutos, em um computador com processador Intel Core i3-540 Processador e 8GB RAM, para analisar cada um dos dois genoma de *Xanthomonas*.

Na Tabela 4, apresentamos uma descrição das características dos métodos disponíveis para identificação de regiões atípicas em comparação com a metodologia desenvolvida aqui. Duas dessas ferramentas estão disponíveis em banco de dados, onde os resultados já encontram-se processados para os genomas publicados, e os demais nos permite baixá-los para a execução das análises localmente. Além disso, cada ferramenta emprega seu critério discriminante, podendo adicionalmente utilizar um método. Além disso os arquivos de entrada e os resultados variam (Tabela 4).

Dentre as ferramentas de bioinformática analisadas, apenas a ME (desenvolvida neste estudo) e Alien Hunter realizam as análises independentemente de arquivo de anotação. Quando comparadas, a ME mostrou-se mais rápida computacionalmente na análise dos genomas de *Xanthomonas* (Tabela 4). Em relação ao IslandPath e SIGI-HMM, foram os mais rápidos, mas sua performance inferior às demais ferramentas (Figura 20).

Tabela 4: Características dos métodos disponíveis para detecção de regiões atípicas em genomas bacterianos comparados com a máxima entropia.

Ferramenta	Disponível em	Critério discriminante	Método	Entrada	Resultado	*Tempo de execução
ME	Pacote	di, tetra ou hexanucleotídeos	Máxima entropia	Formato FASTA	Coord e Proteínas	7 min
Alien Hunter	Pacote	Octanucleotídeos	IVOM + HMM	Formato FASTA	Coord	33 min
HGT-DB	Banco de dados	Uso de códon e Conteúdo G+C		-	-	-
Islander	Banco de dados	Ilha integrativa - tRNA		-	-	-
IslandPath	Pacote e Banco de dados	GC e Viés de Dinucleotídeos		Arquivos NCBI (.ppt/.faa/.ffn)	Coord	34 seg
SIGI-HMM	Pacote	Uso de códon	HMM	Formato EMBL ou GBK	GFF e Coord	36 seg

*Média do tempo de análise dos dois genomas de *Xanthomonas* (5Mb)

Uma outra vantagem é estar disponível em pacote executável, o qual pode ser aplicado a qualquer conjunto de dados.

Na Figura 20 apresentamos uma comparação do F-score e eficiência (EF), das ferramentas analisando dois genomas de *Xanthomonas* [Lima *et al.*, 2005, 2008]. Todas as análises foram realizadas no mesmo computador, sendo que para as ferramentas disponíveis em banco de dados, os resultados foram baixados e a eficiência calculada, conforme apresentado na metodologia. De forma geral, a metodologia ME desenvolvida neste estudo mostrou-se mais eficiente (EF: 82% e EF: 91%) em comparação com as outras ferramentas. Além disso, ME foi superior quando comparado seu F-score (Figura 20). A segunda ferramenta mais eficiente foi o Alien Hunter (EF: 54% e EF: 60%), a qual mais se assemelha com a nova metodologia desenvolvida (Figura 20).

Portanto, considerando a performance, os requisitos para a análise e o tempo de execução, ME pode ser considerada a opção mais eficiente para a discriminação de regiões atípicas, de forma simples e rápida computacionalmente. Além disso, para a análise é necessário apenas a sequência FASTA do genoma, por isso, tal metodologia pode ser aplicada a genomas recém-sequenciados.

Os principais resultados da metodologia proposta são um gráfico e um arquivo de coordenadas com as regiões preditas, oferecendo um guia na investigação dos genomas. Assim, regiões genômicas podem ser priorizadas durante o estudo.

Figura 20: Comparação da performance das ferramentas utilizadas para detecção de regiões atípicas em genomas bacterianos. **a-** *Xanthomonas axonopodis* pv. *citri* 306; **b-** *Xanthomonas campestris* pv. *campestris* ATCC 33913. EF: eficiência.

Para mostrarmos que ME é capaz de apontar regiões dos genomas que podem conter características exclusivas, apresentamos o perfil ME de *X. campestris* pv. *campestris* ATCC 33913 (Figura 21), onde foram selecionadas regiões com valores ME baixos, em torno da mediana e $ME \geq Q_3$. Com estas regiões, verificamos sua porcentagem de cobertura dentro de genomas de bactérias da mesma espécie (intra-espécie) e entre genomas do mesmo gênero (inter-espécie). Cada ponto no gráfico corresponde a uma região genômica de 50kb. Como podemos observar, os valores ME variam muito dentro deste genoma (Figura 21-a), o que corrobora com o fato do mesmo ser rico em elementos de transposição [da Silva *et al.*, 2002]. Pois conforme discutido por Darmon & Leach [2014], tais elementos contribuem para a instabilidade genômica. Além disso, este genoma possui 35 ilhas genômicas [Lima *et al.*, 2005, 2008], outro fator que explica tantas variações nos valores de entropia.

Conforme já esperado, a cobertura (%) das regiões selecionadas com diferentes valores ME teve maior variação dentro do gênero (Figura 21-b), do que quando comparadas dentro da espécie *Xanthomonas campestris* (Figura 21-c). Conseguimos ver ainda que, as regiões com baixo valor ME e com valores em torno da mediana são as regiões mais conservadas dessas bactérias. Por outro lado, as regiões selecionadas que apresentaram os maiores valores (38, 50, 60, 75 e 79) correspondem a regiões com maior variação e com menor cobertura dentro das espécies de *Xanthomonas campestris*, estando associadas a ilhas genômicas, elementos de transposição e bacteriófagos [Lima *et al.*, 2005, 2008].

Figura 21: Máxima de entropia aplicada ao genoma de *Xanthomonas campestris* pv. *campestris* ATCC 33913. **a-** ME aplicada ao genoma de *X. campestris* pv. *campestris* ATCC 33913; **b-** Porcentagem de cobertura (eixo Y) de regiões selecionadas do cromossomo (eixo X) contra todas as espécies selecionadas do gênero *Xanthomonas* (12 bactérias); **c-** Porcentagem de cobertura (eixo Y) de regiões selecionadas do cromossomo (eixo X) contra as espécies de *Xanthomonas campestris* (3 bactérias).

Esta análise foi interessante para confirmarmos que as regiões preditas pela ME correspondem as regiões mais variáveis e com reais possibilidades de se encontrar características não compartilhadas por todas, ou seja, exclusivas.

Como todos resultados preditivos, os mesmos devem ser checados, pois muitas vezes, regiões são apontadas como regiões atípicas pela presença de genes ribossomais (rDNA), como foi o caso das regiões 92 e 99 (Figura 21-a), não correspondendo necessariamente a regiões atípicas. Isso tem sido observado também em outros estudos, os quais destacam que rDNA possuem uma assinatura específica, diferente do restante do genoma [Dufraigne *et al.*, 2005]. Na metodologia apresentada, este problema pode ser resolvido para os genomas que possuem anotação, onde um aviso é dado para

regiões preditas que possuem rDNA, caso contrário, todas as regiões candidatas devem ser cuidadosamente analisadas.

Com isso, apresentamos uma nova metodologia para a identificação de regiões atípicas em genomas bacterianos, e mostramos que a estratégia proposta baseada na ME pode ser um ponto de partida nas análises genômicas, já que pode guiar o pesquisador para regiões específicas, requerendo somente o genoma no formato FASTA.

5.3.1 Máxima entropia aplicada aos genomas de *B. diazoefficiens* CPAC 7 e *B. japonicum* CPAC 15

Após desenvolvermos a metodologia de máxima entropia para discriminar regiões anômalas em genomas bacterianos, a aplicamos aos genomas das duas estirpes brasileiras de *Bradyrhizobium*. O intuito desta análise foi nos direcionar de forma rápida para regiões genômicas onde pudéssemos encontrar particularidades de cada genoma, com enfoque na capacidade simbiótica. Ambos genomas foram analisados com regiões genômicas de 25 kb e hexanucleotídeos. Assim, ME aplicada à CPAC 7 resultou em 364 regiões, e à CPAC 15 384 regiões genômicas (Figura 22), sendo que a predição de regiões atípicas apontou 91 regiões do genoma de *B. diazoefficiens* CPAC 7 como de possível interesse, e em *B. japonicum* CPAC 15 foram preditas 96 regiões a serem exploradas (Figura 22).

Conforme já mencionado anteriormente, as regiões consideradas atípicas são aquelas que apresentam um desvio na distribuição de oligonucleotídeos em algumas regiões, uma vez que tal distribuição no genoma bacteriano é relativamente estável [Karlin & Burge, 1995; Bentley & Parkhill, 2004], o que pode corresponder a sequências de DNA integradas no genoma, muitas vezes adquiridas via transferência horizontal de genes, ou ainda pela presença de elementos de transposição [Dobrindt *et al.*, 2004; Darmon & Leach, 2014]. A este tipo de DNA integrado chamamos de ilhas genômicas, às quais representam extrema importância para as bactérias, uma vez que podem conter genes que conferem maior adaptação ambiental em condições adversas, resistência à antibióticos, entre outros [Ochman & Moran, 2001; Finan, 2002; Dobrindt *et al.*, 2004; Juhas *et al.*, 2009]. As ilhas genômicas que conferem habilidades específicas, como patogenicidade e simbiose, são chamadas de ilhas de patogenicidade e

Figura 22: Perfil de máxima entropia aplicada aos genomas de *Bradyrhizobium diazoefficiens* CPAC 7 e *Bradyrhizobium japonicum* CPAC 15. Cada ponto no gráfico corresponde a uma região genômica de 25 kb, sendo que as regiões com valores $ME \geq Q_3$ (linha azul) representam as regiões de interesse.

ilhas simbióticas [Preston *et al.*, 1998; Dobrindt *et al.*, 2004; Gal-Mor & Finlay, 2006].

Tais regiões atípicas poderão ser exploradas em estudos posteriores, com enfoque evolutivo, a fim de compreender quanto da composição destes genomas foi adquirido via transferência horizontal. Entretanto, no presente estudo estamos interessados em destacar somente os genes exclusivos identificados nas regiões atípicas, uma vez que existe grande possibilidade de encontrarmos características distintivas entre as estirpes brasileiras de *Bradyrhizobium*, que podem contribuir para as diferenças apresentadas durante a simbiose. Pois, *B. diazoefficiens* CPAC 7 possui maior eficiência em fixar N_2 , enquanto que *B. japonicum* CPAC 15 é mais competitiva [Hungria *et al.*, 1996; Mendes *et al.*, 2004; Siqueira *et al.*, 2014].

Em CPAC 7, as regiões 70 a 87, apontadas pela ME, são regiões da ilha simbiótica, e em CPAC 15 a ilha simbiótica corresponde as regiões de 1 a 18 e de 374 a 384 (Figura 22), conforme já descrito em Siqueira *et al.* [2014]. Além da ilha simbiótica, Siqueira *et al.* [2014] também já haviam identificado várias ilhas genômicas nestes genomas, às quais corroboram com nossos resultados. Entretanto, a máxima entropia apontou outras regiões atípicas adicionais (Apêndice H). Várias dessas regiões, apesar de serem atípicas, estão presentes em ambos genomas, como é o caso da maior parte da ilha simbiótica. E como o objetivo desta análise foi apontar particularidades dos genomas, foram excluídas aquelas regiões de CPAC 7 que apresentaram uma cobertura $\geq 75\%$ no genoma de CPAC 15, e vice-versa (Apêndices F e G). Dessa forma, selecionamos 49 regiões para explorarmos de *B. diazoefficiens* CPAC 7 e 70 regiões do genoma de *B. japonicum* CPAC 15, cujas coordenadas estão apresentadas no Apêndice H.

Considerando as regiões selecionadas, foram extraídas somente as ORFs exclusivas de cada estirpe, sendo que aqui apresentamos somente aquelas que ainda não haviam sido relatadas na análise dos elementos de transposição e das proteínas hipotéticas. O total de ORFs exclusivas obtidas nesta análise foi 269 e 368 ORFs de *B. diazoefficiens* CPAC 7 e *B. japonicum* CPAC 15, respectivamente. Estes resultados estão listados no Apêndice I, os quais podem auxiliar no direcionamento de novos estudos. A maioria dos genes exclusivos, em ambas bactérias, está associada com processos metabólicos (Apêndice I). Apesar de outros estudos serem necessários para afirmar que tais regiões correspondem de fato a ilhas genômicas, já era esperado que a maioria dos genes presentes nas regiões atípicas fos-

sem genes metabólicos [Ochman *et al.*, 2000; Kanhere & Vingron, 2009]. Proteínas do metabolismo primário podem aumentar a competitividade e adaptação das bactérias em determinadas condições [Dobrindt *et al.*, 2004].

Além disso, nessas regiões, conforme já ressaltado por Siqueira *et al.* [2014], verificamos a presença de elementos de transposição, o que também é comum nestas regiões [Dobrindt *et al.*, 2004; Darmon & Leach, 2014], contudo tais elementos não são apresentados na lista final de ORFs exclusivas, pois a análise dos mesmos foi realizada na abordagem de elementos de transposição. Também identificamos genes com funções regulatórias, como por exemplo, reguladores transcricionais (Apêndice I). Nos chamou atenção a presença de reguladores transcricionais LuxR exclusivos de CPAC 15, os quais estão relacionados com o processo de *quorum sensing*, crucial para a interação da bactéria com a planta [Soto *et al.*, 2006; Nasser & Reverchon, 2007; Subramoni *et al.*, 2011]. Temos também proteínas transportadoras, proteínas secretoras, proteínas de conjugação, entre outras (Apêndice I).

Discutiremos em maiores detalhes seis genes selecionados que codificam proteínas que podem estar diretamente ligadas à capacidade simbiótica das bactérias (Tabela 5), uma vez que estes genes não foram abordados no estudo de análise comparativa de CPAC 7 e CPAC 15 [Siqueira *et al.*, 2014].

Tabela 5: Genes de destaque selecionados pela máxima entropia nos genomas das estirpes brasileiras de *Bradyrhizobium*.

Anotação	Estirpes de <i>Bradyrhizobium</i>		Região Genômica
	Bd CPAC 7	Bj CPAC 15	
Regulador transcricional LysR	BJA5080_05086	-	Região 155
FixW	-	BJS_08204	Região 7
NoeE	-	BJS_07601	Região 14
Dihidrorizobiotoxina sintase (RtxA)	-	BJS_07600	Região 14
Nex18 simbioticamente induzida	-	BJS_07054	Região 38
Proteína de ligação ao hospedeiro	-	BJS_00878	Região 227

Em CPAC 7, destacamos um regulador transcricional LysR extraído da região 155 apontada pela máxima entropia (Tabela 5). Este gene pode ser encontrado em várias espécies de rizóbios como *B. diazoefficiens*, *Sinorhizobium fredii*, *Rhizobium phaseoli*, *Rhizobium etli*, entre outras. Contudo está ausente em espécies de *B. japonicum*. Na estirpe CPAC 15, também identificamos reguladores transcricionais LysR exclusivos (Apêndice I).

Talvez estudos direcionados para esta família de reguladores transcricionais (LysR) em genomas de rizóbios possa nos trazer novas informações. Genes da família LysR podem regular diversos conjuntos de genes, incluindo metabolismo, *quorum sensing* e motilidade [Maddocks & Oyston, 2008]. Além disso, já foi demonstrado que reguladores transcricionais LysR são requeridos na nodulação de *S. meliloti* [Luo *et al.*, 2005; MacLean *et al.*, 2008].

A região 7 do genoma de *B. japonicum* CPAC 15 (Figura 22) compreende parte da ilha simbiótica. Nesta região com tamanho de 25 kb, apenas 14% está presente em CPAC 7, a maioria dos genes são exclusivos de CPAC 15 e estão anotados como hipotéticos. Contudo, identificamos um gene anotado como *fixW* (BJS_08204) que nos chamou atenção, ausente na espécie *B. diazoefficiens*. A proteína codificada pelo gene *fixW* possui 55% de identidade e 35% de cobertura com o *fixW* caracterizado em *R. leguminosarum* [Hontelez *et al.*, 1989; Martínez *et al.*, 2004]. Assim, por estar presente na ilha simbiótica de CPAC 15 e ausente em CPAC 7, talvez possa ser interessante estudar este gene funcionalmente, pois não sabemos se é um gene ativo ou não.

A região 14 também corresponde a parte da ilha simbiótica de CPAC 15, onde 55% desta região está ausente em CPAC 7. Dentre os genes exclusivos, destacamos dois, o gene *noeE* (BJS_07601-328172:328786) e dihidrorizobiotoxina sintase (BJS_07600-330515:332944).

A proteína codificada pelo gene *noeE* realiza o processo de sulfatação dos fatores Nod de *Rhizobium* sp. NGR234, a qual está associada com a especificidade de hospedeiro [Hanin *et al.*, 1997; Quesada-Vincens *et al.*, 1998]. Conforme já relatado por Göttfert *et al.* [2001], os fatores Nod de *B. diazoefficiens* USDA 110 não possuem sulfatação descrita. Mas apesar disso, destacamos este gene por estar presente em vários outros genomas de *Bradyrhizobium*, incluindo *B. japonicum* CPAC 15, *B. japonicum* E109 e *B. japonicum* USDA 6, mas ausente em *B. diazoefficiens* CPAC 7. Além disso, ressaltamos a presença de um segundo gene que também corresponde ao *noeE*, e que na anotação do genoma publicado de CPAC 15 não havia sido identificado [Siqueira *et al.*, 2014], mas está presente na coordenada 327722:328069, ao lado de BJS_07601 (328172:328786). Tais genes correspondem aos homólogos *blr2073* e *blr2074* em *B. diazoefficiens* USDA 110.

Outro gene de destaque é o que codifica a proteína dihidrorizobiotoxina sintase (RtxA). Esta proteína é uma rizobiotoxina presente em algumas espécies de *Bradyrhizobium*, como *B. japonicum* e *B.*

elkanii, a qual pode causar clorose foliar em plantas hospedeiras suscetíveis [Ruan & Peters, 1992; Ruan *et al.*, 1993; Yasuta *et al.*, 2001]. Esse gene pode explicar os sintomas de clorose em folhas de algumas cultivares de soja causado apenas por CPAC 15, na qual o gene está presente; e a ausência em CPAC 7, que não desenvolve esses sintomas [Boddey & Hungria, 1997]. Uma outra atividade metabólica desta rizobiotoxina, interessante para a simbiose, é sua capacidade em inibir a atividade da ACC sintase (Aminociclopropano-1-carboxílico ácido sintase - EC 4.4.1.14), enzima chave na biossíntese do etileno [Yu *et al.*, 1979; Yasuta *et al.*, 2001]. Uma vez que o etileno está associado com a resposta de defesa da planta, podendo regular negativamente a formação dos nódulos [Goodlass & Smith, 1979; Grennan, 2008].

Adicionalmente, a produção desta rizobiotoxina em *B. elkanii* aumenta sua competitividade e a nodulação na leguminosa *Macroptilium atropurpureum* [Yuhashi *et al.*, 2000]. Tanto *B. elkanii* quanto *B. japonicum* possuem a capacidade de estabelecer simbiose com esta leguminosa, assim como com a soja [Kuykendall *et al.*, 1992]. Entretanto, *B. elkanii* mostrou-se mais competitiva do que *B. japonicum* USDA 110, reclassificada mais tarde como *B. diazoefficiens* USDA 110 [Yuhashi *et al.*, 2000; Delamuta *et al.*, 2013]. Sabendo agora que a estirpe estudada nestas condições pertence na verdade à espécie *B. diazoefficiens*, seria interessante explorar o papel desta rizobiotoxina em estirpes de *B. diazoefficiens* e *B. japonicum*, principalmente no que diz respeito à competitividade da bactéria. Pois esta é exatamente uma das principais diferenças fenotípicas encontradas entre *B. japonicum* CPAC 15 e *B. diazoefficiens* CPAC 7, sendo que a CPAC 15 possui este gene e é a mais competitiva entre as duas, já CPAC 7 não possui este gene, uma importante diferença encontrada no conteúdo da ilha simbiótica [Siqueira *et al.*, 2014].

Outra diferença interessante entre os genomas destas duas estirpes brasileiras foi a presença do gene *nex18*, encontrado na região genômica 38 de *B. japonicum* CPAC 15 (Tabela 5), anotada até então como hipotética [Siqueira *et al.*, 2014]. A proteína codificada por este gene possui o domínio fasciclina, associada à adesão celular [Moody & Williamson, 2013]. Esta proteína pode ser encontrada em várias bactérias, sendo induzida durante a simbiose, conforme mostrado em *S. meliloti* 1021 [Oke & Long, 1999a; Rossbach *et al.*, 2008]. Oke & Long [1999b] mostraram também que este gene (*nex18*) é altamente expresso no desenvolvimento de bacteroides, cujos seus mutantes possuem uma

simbiose comprometida. Além disso, também está relacionada com a resposta de estresse dos rizóbios [Djordjevic *et al.*, 2003; Cytryn *et al.*, 2007].

Destacamos ainda, uma proteína codificada por um gene presente na região genômica 227 da estirpe CPAC 15, apontada pela máxima entropia, proteína de ligação ao hospedeiro (ou em inglês, *Host attachment protein*) (Tabela 5). Tal proteína está ausente em genomas de *B. diazoefficiens*, sendo requerida para fixação da bactéria às células do hospedeiro, o que pode representar uma importante diferença entre as estirpes estudadas, uma vez que, conforme demonstrado em *Agrobacterium tumefaciens*, mutações em genes desta família podem reduzir a habilidade de colonização das raízes [Matthysse *et al.*, 2000; Matthysse & McMahan, 2001]. Por isso, apontamos este gene como potencial alvo de estudo experimental, o que pode estar associado diretamente com as diferenças simbióticas de *B. diazoefficiens* CPAC 7 e *B. japonicum* CPAC 15.

Como foi possível verificar com esta análise, ainda há muito para ser explorado nestes genomas de interesse. Com a metodologia desenvolvida (ME), conseguimos minerar nosso conjunto de dados e apontar exclusividades interessantes das estirpes brasileiras de *Bradyrhizobium*. Além de indicar genes que poderiam ser estudados funcionalmente (Tabela 5), disponibilizamos uma lista com vários outros genes exclusivos de cada estirpe para que também possam ser melhor explorados em estudos futuros (Apêndice I). Complementarmente, detectamos novas regiões genômicas atípicas em ambos genomas de *Bradyrhizobium* (Apêndice H), as quais ainda não haviam sido apontadas na análise comparativa [Siqueira *et al.*, 2014]. Portanto, tais informações poderão ser utilizadas em possíveis estudos de caráter evolutivo.

6

Conclusões

A genômica comparativa compreende estratégias de transferência de informação entre genomas distintos que resultam em aumento do poder preditivo de funções biológicas no organismo alvo. A predição de função potencial de um dado genoma é necessária para a realização de experimentos de validação biológica e descoberta científica. O presente estudo ampliou o conhecimento de duas linhagens de *Bradyrhizobium* usadas em condições de campo para fornecer nitrogênio para a cultura da soja. Três abordagens distintas foram empregadas para comparar os genomas de CPAC 7 e CPAC 15 sendo elas, o impacto local de transposons do tipo IS, a atualização da anotação de genes considerados hipotéticos por Siqueira *et al.* [2014] e a identificação de regiões atípicas e posterior detecção de genes exclusivos presentes nessas regiões em cada genoma. A identificação de regiões atípicas foi realizada após o desenvolvimento de uma ferramenta computacional baseada em Máxima Entropia.

A anotação curada dos elementos de transposição do tipo IS presentes nestes genomas identificou 33 novas sequências de inserção completas. Além disso, o impacto dos ISs sobre genes foi analisado (interrupção, inserção no promotor, inserção no terminador e ORF modificada) e observamos que genes potencialmente associados ao processo de simbiose (*nopAN*, *nopAG*, *rhcU*, *modC*, *hypB*) são afetados diferentemente pela inserção dos elementos. Verificamos que 13 eventos de inserção são conservados em outros genomas de *Bradyrhizobium*, o que levanta a possibilidade de que tais elementos possam estar auxiliando na regulação gênica, como é o caso do ISRj1 inserido na região

promotora do gene *nopAN*, localizado na ilha simbiótica.

Um total de 1.204 genes considerados hipotéticos foram anotados e verificou-se que 297 possuem participação comprovada no processo de interação planta-batéria em outros sistemas. Dentre elas, destacam-se as proteínas envolvidas com o processo de motilidade e adesão celular, fatores de virulência, proteínas efetoras, proteases e várias proteínas que possuem peptídeos sinais. Este estudo representa um ganho de conhecimento sobre o conteúdo gênico destas estirpes economicamente importantes para o Brasil e apontam para novos candidatos a serem estudados experimentalmente. Na Figura 23 apresentamos o incremento de informações na ilha simbiótica de CPAC 7 e CPAC 15 quanto aos genes impactados por elementos ISs e genes, até então hipotéticos, que tiveram sua anotação atualizada.

Considerando também a plasticidade dos genomas bacterianos, os quais podem apresentar particularidades em regiões atípicas dos genomas, desenvolvemos uma nova metodologia baseada na máxima entropia. A metodologia ME é capaz de identificar tais regiões de forma rápida a partir da sequência de nucleotídeos desprovida de qualquer informação de anotação. Tal metodologia foi comparada com outros métodos conhecidos, superando-os em termos de eficiência e tempo de execução. A nova metodologia desenvolvida neste trabalho, possui a vantagem de independe, a priori, da anotação do genoma, podendo ser empregada para direcionar as análises em genomas recém-sequenciados.

A máxima entropia foi utilizada para apontar regiões atípicas dos genomas de CPAC 7 e CPAC 15, de onde extraímos 269 genes exclusivos de CPAC 7 e 368 exclusivos de CPAC 15. A ME foi capaz de apontar novas regiões atípicas em ambos genomas, além daquelas já descritas por Siqueira *et al.* [2014]. Ao explorarmos as regiões atípicas, destacamos genes como *fixW*, *noeE*, *rtxA* e *nex18* que são potencialmente responsáveis por parte das diferenças simbióticas, estando presentes apenas em CPAC 15. Portanto, demonstramos que a máxima entropia representa uma estratégia eficiente na seleção de características e na mineração de dados, nos direcionando a explorar regiões genômicas pontuais, que possuam maior chance de encontrarmos particularidades de cada genoma.

Figura 23: Ilha simbiótica de *B. diazoefficiens* CPAC 7 (Maior eficiência em fixar N_2) e *B. japonicum* CPAC 15 (Maior competitividade) com as principais contribuições obtidas em comparação com o conhecimento anterior ao presente estudo. *: Gene *nopAN* com possível interferência de elemento IS, e em CPAC 7 não constava na anotação publicada. IS: Sequência de Inserção. A ilha simbiótica está delimitada por recombinase e tRNA, respectivamente

Diante dos resultados apresentados, verificamos que não é possível apontar um único fator responsável pelas diferenças simbióticas de CPAC 7 e CPAC 15, e sim um conjunto de fatores somados que as distinguem, uma vez que possuem várias particularidades. Para entendermos melhor o papel dos fatores apontados neste estudo serão necessários estudos experimentais. A Figura 24 ilustra de modo resumido alguns aspectos específicos de cada bactéria encontrados durante o desenvolvimento desta tese, os quais podem estar associados a suas diferenças fenotípicas.

Figura 24: Representação de alguns aspectos genéticos específicos de cada estirpe de *Bradyrhizobium* identificados neste estudo que podem estar relacionados com suas diferenças fenotípicas. **Vermelho:** exclusividades identificadas na análise dos elementos de transposição; **Verde:** genes exclusivos identificados entre os hipotéticos; **Azul:** genes exclusivos de cada estirpe identificados nas regiões atípicas.

Apêndices

Elementos ISs completos identificados em *B. diazoefficiens* CPAC 7

	IS	Família	Coord. inicial da ORF	Coord. final da ORF	Tam. da) ORF (pb)	Tam. da proteína	Coord. inicial do IS	Coord. final do IS	Tam. do IS (pb)	Tam. da DR (pb)	Tam. da IR (pb)	Contig
1	ISBd1	ISNCY ssgr ISLbi1	1668	157	1512	503	1757	15	1743	0	12/15	Cont26
2	ISRj2	IS3 ssgr IS150	21752	22057	306	101	21664	23027	1364	3	21/26	Cont26
	ISRj2		21931	22881	951	316	21664	23027	1364	3	21/26	Cont26
3	ISBd20	IS630	29006	29194	189	62	28921	30106	1186	2	19/27	Cont26
	ISBd20		29226	30095	870	289	28921	30106	1186	2	19/27	Cont26
4	ISBd14	IS91	191841	192797	957	318	191773	192875	1103	0	4	Cont26
5	ISBd11	IS91	493151	494116	966	321	493056	494338	1283	0	8/11	Cont26
6	ISBd24	IS3 ssgr IS150	1739433	1739744	312	103	1739346	1740629	1284	3	22/26	Cont26
	ISBd24		1739741	1740568	828	275	1739346	1740629	1284	3	22/26	Cont26
7	ISBd16	IS110 ssgr IS1111	1742932	1741898	1035	344	1743050	1741641	1410	0	3/12	Cont26
8	ISRj2	IS3 ssgr IS150	1748853	1747903	951	318	1749121	1747757	1365	3	21/26	Cont26
	ISRj2		1749032	1748727	306	102	1749121	1747757	1365	3	21/26	Cont26
9	ISBd9	IS21	1849608	1848904	705	234	1851124	1848825	2300	2	23/29	Cont26
	ISBd9		1850972	1849618	1355	450	1851124	1848825	2300	2	23/29	Cont26
10	ISRj1	IS630	1874	2938	1065	354	1831	2948	1118	2	6	Cont22
11	ISBd2	IS256	103393	104551	1158	385	103252	104646	1395	4	13	Cont21
12	ISRj1	IS630	96721	97785	1065	354	96711	97828	1118	2	6	Cont21
13	ISRj2	IS3 ssgr IS150	85693	85388	306	101	85781	84418	1364	3	21/26	Cont21
	ISRj2		85514	84564	951	316	85781	84418	1364	3	21/26	Cont21
14	ISRj1	IS630	76915	77979	1065	354	76905	78022	1118	2	6	Cont21
15	ISBd2	IS256	43069	44228	1159	385	45528	44133	1396	9	13	Cont21
16	ISRj1	IS630	22327	23391	1065	354	22317	23434	1118	2	6	Cont21
17	ISBd5	IS1380	19125	20513	1389	462	18968	20589	1622	4	19	Cont21
18	ISBd23	IS5 ssgr IS427	9469	9017	453	150	9478	8645	834	4	9/15	Cont21
	ISBd23		9101	8707	395	130	9478	8645	834	4	9/15	Cont21
19	ISBd17	IS110 ssgr IS1111	5843	6874	1032	343	5670	7008	1339	3	5/12	Cont21
20	ISRj1	IS630	3795	5135	1341	446	4063	5178	1116	2	6	Cont21
21	ISBj2	IS5 ssgr IS5	1497	997	501	166	1505	1	1505	0	13/18	Cont21
	ISBj2		1001	165	837	278	1505	1	1505	0	13/18	Cont21
22	ISBd4	IS481	23594	22584	1011	336	23760	22521	1240	0	18/25	Cont25

Tabela 6 – Continuação

	IS	Família	Coord. inicial da ORF	Coord. final da ORF	Tam. da) ORF (pb)	Tam. da proteína	Coord. inicial do IS	Coord. final do IS	Tam. do IS (pb)	Tam. da DR (pb)	Tam. da IR (pb)	Contig
23	ISBd5	IS1380	26680	25804	877	291	25635	26791	1157	4	19	Cont25
24	ISBd12	IS21	2371	871	1500	499	2485	6	2480	0	23/27	Cont20
	ISBd12		871	8	864	287	2485	6	2480	0	23/27	Cont20
25	ISRj2	IS3 ssgr IS150	18713	17763	951	316	18980	17617	1364	3	21/26	Cont20
	ISRj2		18892	18587	306	101	18980	17617	1364	3	21/26	Cont20
26	ISBja8	IS66	37943	38318	376	125	37845	39429	1585	0	16/25	Cont20
	ISBja8		38338	38650	313	104	37845	39429	1585	0	16/25	Cont20
	ISBja8		38651	39317	667	221	37845	39429	1585	0	16/25	Cont20
27	ISBd23	IS5 ssgr IS427	45854	46248	395	130	45791	46625	835	4	9/15	Cont20
	ISBd23		46162	46616	455	150	45791	46625	835	4	9/15	Cont20
28	ISBd18	IS110 ssgr IS1111	88414	87446	969	322	88564	87208	1357	0	4/10	Cont20
29	ISBd6	ISL3	90150	88594	1557	518	90217	88564	1654	5	20/24	Cont20
30	ISBd23	IS5 ssgr IS427	95262	94810	453	150	95632	94801	832	4	9/15	Cont20
	ISBd23		95570	95178	393	130	95632	94801	832	4	9/15	Cont20
31	ISBd1	ISNCY ssgr ISLbi1	100387	101811	1425	474	100298	102039	1742	0	12/15	Cont20
32	ISRj1	IS630	112060	112473	414	137	112017	113133	1117	2	6	Cont20
	ISRj1		112464	113123	660	219	112017	113133	1117	2	6	Cont20
33	ISBd15	IS6	1690650	1691408	759	252	1690643	1691458	816	0	14/16	Cont24
34	ISBd23	IS5 ssgr IS427	1689573	1689121	453	150	1689582	1688751	832	4	9/15	Cont24
	ISBd23		1689205	1688813	393	130	1689582	1688751	832	4	9/15	Cont24
35	ISBd19	IS110 ssgr IS1111	1665644	1664649	996	331	1665666	1664404	1263	0	0	Cont24
36	ISBj6	IS701	1656038	1657798	1761	586	1656412	1657859	1448	4	18/19	Cont24
37	ISRj2	IS3 ssgr IS150	1648774	1649079	306	101	1647804	1649167	1364	4	21/26	Cont24
	ISRj2		1647950	1648900	951	316	1647804	1649167	1364	4	21/26	Cont24
38	ISBd13	IS21	1626856	1626578	279	92	1627144	1624900	2245	0	16/17	Cont24
	ISBd13		1626333	1625974	360	119	1627144	1624900	2245	0	16/17	Cont24
	ISBd13		1625901	1624936	966	321	1627144	1624900	2245	0	16/17	Cont24
39	ISBd23	IS5 ssgr IS427	1615369	1615761	393	130	1614993	1615823	831	4	9/15	Cont24
	ISBd23		1615002	1615453	452	149	1614993	1615823	831	4	9/15	Cont24
40	ISBd25	IS3 ssgr IS3	1601719	1602018	300	99	1601435	1602077	643	2	5	Cont24
	ISBd25		1601483	1601710	228	75	1601435	1602077	643	2	5	Cont24
41	ISBj2	IS5 ssgr IS5	1601106	1600606	501	166	1601114	1599608	1507	6	13/18	Cont24
	ISBj2		1600610	1599774	837	278	1601114	1599608	1507	6	13/18	Cont24
42	ISRj1	IS630	1595793	1596857	1065	354	1595783	1596900	1118	2	6	Cont24
43	ISBd10	IS21	1571116	1572546	1431	476	1570205	1572810	2606	0	17/18	Cont24
	ISBd10		1570310	1571119	810	269	1570205	1572810	2606	0	17/18	Cont24
44	ISBja3	IS110 ssgr IS1111	1518069	1519114	1046	347	1517797	1519204	1408	0	7/13	Cont24
45	ISBd5	IS1380	1504464	1503076	1389	462	1504621	1503000	1622	4	19	Cont24
46	ISBd12	IS21	1490104	1491402	1299	432	1489037	1491516	2480	2	23/27	Cont24
	ISBd12		1489929	1490042	114	37	1489037	1491516	2480	2	23/27	Cont24
	ISBd12		1489039	1489902	864	287	1489037	1491516	2480	2	23/27	Cont24
47	ISBd5	IS1380	1455979	1455752	228	75	1456136	1454516	1621	4	19	Cont24

Tabela 6 – Continuação

Tabela 6 - Continuação												
IS	Família	Coord. inicial da ORF	Coord. final da ORF	Tam. da) ORF (pb)	Tam. da proteína	Coord. inicial do IS	Coord. final do IS	Tam. do IS (pb)	Tam. da DR (pb)	Tam. da IR (pb)	Contig	
	ISBd5	1455677	1454592	1086	361	1456136	1454516	1621	4	12/19	Cont24	
48	ISRj2	IS3 ssgr IS150	1452200	1451250	951	316	1452346	1450984	1363	0	21/26	Cont24
	ISRj2		1451376	1451071	306	101	1452346	1450984	1363	0	21/26	Cont24
49	ISRj2	IS3 ssgr IS150	1442060	1442363	304	100	1441089	1442451	1363	0	21/26	Cont24
	ISRj2		1441235	1442184	950	315	1441089	1442451	1363	0	21/26	Cont24
50	ISBd23	IS5 ssgr IS427	1439666	1439214	453	150	1439675	1438844	832	0	9/15	Cont24
	ISBd23		1439298	1438906	393	130	1439675	1438844	832	0	9/15	Cont24
51	ISBd3	IS256	1420027	1418822	1206	401	1420058	1418756	1303	0	24/28	Cont24
52	ISBd23	IS5 ssgr IS427	1417403	1417795	393	130	1417026	1417857	832	4	9/15	Cont24
	ISBd23		1417035	1417487	453	150	1417026	1417857	832	4	9/15	Cont24
53	ISBj5	IS5 ssgr IS5	738816	739736	921	306	738381	739905	1525	6	15/16	Cont24
	ISBj5		738389	738850	462	153	738381	739905	1525	6	15/16	Cont24
54	ISBj4	IS110 ssgr IS1111	455280	454216	1065	354	455716	454141	1576	0	15	Cont24
55	ISBj4	IS110 ssgr IS1111	2835	1772	1064	353	2910	1337	1574	0	15	Cont23
56	ISBd23	IS5 ssgr IS427	466236	466628	393	130	466174	467005	832	0	9/15	Cont23
	ISBd23		466544	466996	453	150	466174	467005	832	0	9/15	Cont23
57	ISBd5	IS1380	572883	571495	1389	462	572959	571338	1622	4	19	Cont23
58	ISBd22	IS630	587151	587549	399	132	588181	587138	1044	2	19/25	Cont23
	ISBd22		587985	587707	279	92	588181	587138	1044	2	19/25	Cont23
59	ISBj4	IS110 ssgr IS1111	624403	625467	1065	354	624328	625903	1576	0	15	Cont23
60	ISBd7	IS1182	634731	636185	1455	484	634669	636365	1697	0	16/17	Cont23
61	ISBd21	IS630	639116	639502	387	128	639028	640071	1044	2	19/23	Cont23
	ISBd21		639585	640058	474	157	639028	640071	1044	2	19/23	Cont23
62	ISRj1	IS630	712694	711630	1065	354	712737	711620	1118	2	6	Cont23
63	ISBj5	IS5 ssgr IS5	871296	872642	1347	448	871127	872650	1524	6	15/16	Cont23
64	ISBd5	IS1380	2622638	2621250	1389	462	2622714	2621093	1622	4	19	Cont23
65	ISBd5	IS1380	3348671	3347283	1389	462	3348747	3347126	1622	4	19	Cont23
66	ISBja4	IS110 ssgr IS1111	3445472	3444442	1031	342	3445552	3444182	1371	0	13	Cont23
67	ISBd5	IS1380	3761179	3761580	402	133	3761070	3762690	1621	4	19	Cont23
	ISBd5		3761607	3762533	927	308	3761070	3762690	1621	4	19	Cont23
68	ISBd5	IS1380	3800523	3801866	1344	447	3800401	3802023	1623	4	19	Cont23
69	ISBd5	IS1380	3867854	3869242	1389	462	3867778	3869399	1622	4	19	Cont23
70	ISBd5	IS1380	3872583	3871195	1389	462	3872659	3871038	1622	4	19	Cont23
71	ISBd8	IS5 ssgr IS5	245696	247025	1330	442	245523	247057	1535	6	13/15	Cont27
72	ISBd5	IS1380	453279	452353	927	308	453816	452196	1621	4	19	Cont27
	ISBd5		453707	453306	402	133	453816	452196	1621	4	19	Cont27
73	ISRj1	IS630	950626	951288	663	220	950616	951735	1120	2	6	Cont27
	ISRj1		951318	951692	375	124	950616	951735	1120	2	6	Cont27
74	ISBd12	IS21	1049101	1049388	288	95	1048987	1051467	2481	0	23/27	Cont27
	ISBd12		1049357	1050010	654	217	1048987	1051467	2481	0	23/27	Cont27
	ISBd12		1049405	1050403	999	332	1048987	1051467	2481	0	23/27	Cont27
	ISBd12		1050417	1050602	186	61	1048987	1051467	2481	0	23/27	Cont27

Tabela 6 – Continuação

	IS	Família	Coord. inicial da ORF	Coord. final da ORF	Tam. da) ORF (pb)	Tam. da proteína	Coord. inicial do IS	Coord. final do IS	Tam. do IS (pb)	Tam. da DR (pb)	Tam. da IR (pb)	Contig
	ISBd12		1050602	1051465	864	287	1048987	1051467	2481	0	23/27	Cont27
75	ISBd20	IS630	1153885	1154994	1110	369	1153821	1155005	1185	2	19/27	Cont27
76	ISBd12	IS21	1155907	1155044	864	287	1157520	1155042	2479	0	23/26	Cont27
	ISBd12		1157406	1155907	1500	499	1157520	1155042	2479	0	23/26	Cont27
77	ISBd9	IS21	1167413	1168933	1521	506	1167261	1169786	2526	2	23/29	Cont27
	ISBd9		1169003	1169706	704	233	1167261	1169786	2526	2	23/29	Cont27

Elementos ISs completos identificados em *B. japonicum* CPAC 15

	IS	Família	Coord. inicial da ORF	Coord. final da ORF	Tam. da) ORF (pb)	Tam. da proteína	Coord. inicial do IS	Coord. final do IS	Tam. do IS (pb)	Tam. da DR (pb)	Tam. da IR (pb)
1	ISBd4	IS481	21845	20835	1011	336	22011	20772	1240	0	18/25
2	ISBd5	IS1380	24929	24053	877	292	25040	23886	1155	4	19
3	ISBja8	IS66	28406	28780	375	124	28308	29891	1584	0	16/25
	ISBja8		28800	29112	313	104	28308	29891	1584	0	16/25
	ISBja8		29113	29779	667	221	28308	29891	1584	0	16/25
4	ISBd23	IS5 ssgr IS427	36314	36706	393	130	36252	37083	832	4	9/15
	ISBd23		36622	37074	453	150	36252	37083	832	4	9/15
5	ISRj1	IS630	129829	130893	1065	354	129786	130903	1118	2	6
6	ISRj1	IS630	198213	199277	1065	354	198170	199287	1118	2	6/8
7	ISBd23	IS5 ssgr IS427	200552	200944	393	130	200490	201321	832	4	9/15
	ISBd23		200860	201312	453	150	200490	201321	832	4	9/15
8	ISBd4	IS481	207443	206433	1011	336	207609	206370	1240	7	18/25
9	ISBd25	IS3 ssgr IS3	215543	215842	300	99	215484	216126	643	2	5
	ISBd25		215851	216078	228	75	215484	216126	643	2	5
10	ISBj2	IS5 ssgr IS5	216940	216455	486	161	217965	216447	1519	6	13/18
	ISBj2		217799	216948	852	283	217965	216447	1519	6	13/18
11	ISBd23	IS5 ssgr IS427	223803	223352	452	150	224173	223343	831	0	9/15
	ISBd23		224111	223719	393	130	224173	223343	831	0	9/15
12	ISBd23	IS5 ssgr IS427	226274	226666	393	130	226212	227043	832	4	9/15
	ISBd23		226582	227034	453	150	226212	227043	832	4	9/15
13	ISRj1	IS630	228121	227057	1065	354	228164	227047	1118	2	6
14	ISBja3	IS110 ssgr IS1111	300950	301993	1044	347	300860	302265	1406	0	7/13
15	ISRj2	IS3 ssgr IS150	378200	377263	938	311	378466	377116	1351	3	21/26
	ISRj2		378378	378073	306	101	378466	377116	1351	3	21/26
16	ISBja6	IS3 ssgr IS407	381874	380948	927	308	382196	380930	1267	4	17/19
	ISBja6		382133	381868	266	88	382196	380930	1267	4	17/19
17	ISRj2	IS3 ssgr IS150	384749	383799	951	316	385016	383653	1364	4	21/26
	ISRj2		384928	384623	306	101	385016	383653	1364	4	21/26
18	ISBd3	IS256	423594	422389	1206	401	423660	422358	1303	0	24/28
19	ISBd23	IS5 ssgr IS427	765696	766088	393	130	765634	766465	832	4	9/15
	ISBd23		766004	766456	453	150	765634	766465	832	4	9/15

Tabela 7 – Continuação

	IS	Família	Coord. inicial da ORF	Coord. final da ORF	Tam. da) ORF (pb)	Tam. da proteína	Coord. inicial do IS	Coord. final do IS	Tam. do IS (pb)	Tam. da DR (pb)	Tam. da IR (pb)
20	ISBd6	ISL3	961505	963059	1555	517	961438	963089	1652	0	21/24
21	ISBd23	IS5 ssgr IS427	993800	993348	453	150	994170	993339	832	4	9/15
	ISBd23		994108	993716	393	130	994170	993339	832	4	9/15
22	ISBd6	ISL3	1010425	1011318	894	297	1010358	1011991	1634	0	21/24
	ISBd6		1011598	1011963	366	121	1010358	1011991	1634	0	21/24
23	ISBja5	IS110 ssgr IS1111	1102412	1101403	1010	335	1102493	1101142	1352	0	13
24	ISBd23	IS5 ssgr IS427	1210949	1210497	453	150	1211319	1210488	832	4	9/15
	ISBd23		1211257	1210865	393	130	1211319	1210488	832	4	9/15
25	ISBd4	IS481	1234586	1235596	1011	336	1234420	1235659	1240	7	18/25
26	ISBd16	IS110 ssgr IS1111	1248639	1249673	1035	344	1248521	1249929	1409	0	3/12
27	ISBja4	IS110 ssgr IS1111	1660004	1661035	1032	343	1659924	1661293	1370	0	13
28	ISBja6	IS3 ssgr IS407	1911424	1911690	267	88	1911361	1912628	1268	4	17/19
	ISBja6		1911684	1912610	927	308	1911361	1912628	1268	4	17/19
29	ISRj1	IS630	2249935	2248871	1065	354	2249978	2248861	1118	2	6
30	ISBj4	IS110 ssgr IS1111	2263784	2262720	1065	354	2263859	2262282	1578	0	15
31	ISBja4	IS110 ssgr IS1111	2292653	2293684	1032	343	2292572	2293946	1375	0	13
32	ISBj4	IS110 ssgr IS1111	2964282	2963218	1065	354	2964357	2962780	1578	0	15
33	ISBja4	IS110 ssgr IS1111	3396742	3397773	1032	343	3396662	3398031	1370	0	13
34	ISBja2	IS1182	4204450	4203011	1440	479	4204513	4202950	1564	2	15/18
35	ISBd23	IS5 ssgr IS427	4420082	4420473	392	130	4420020	4420850	831	4	9/15
	ISBd23		4420390	4420841	452	150	4420020	4420850	831	4	9/15
36	ISBja2	IS1182	4447379	4445940	1440	479	4447442	4445879	1564	2	15/18
37	ISBja3	IS110 ssgr IS1111	5307179	5306136	1044	347	5307269	5305864	1406	0	7/13
38	ISBja2	IS1182	5448554	5447115	1440	479	5448617	5447054	1564	0	15/18
39	ISBd23	IS5 ssgr IS427	6362300	6361848	453	150	6362670	6361839	832	4	9/15
	ISBd23		6362608	6362216	393	130	6362670	6361839	832	4	9/15
40	ISBja1	IS481	7245765	7244958	808	267	7246055	7244955	1101	4	15/18
	ISBja1		7245966	7245766	201	67	7246055	7244955	1101	4	15/18
41	ISBj4	IS110 ssgr IS1111	8352308	8353372	1065	354	8352233	8353810	1578	0	15
42	ISBd23	IS5 ssgr IS427	8594478	8594026	453	150	8594848	8594017	832	0	9/15
	ISBd23		8594786	8594394	393	130	8594848	8594017	832	0	9/15
43	ISBd20	IS630	8640427	8639318	1110	369	8640491	8639307	1185	2	19/27
44	ISRj2	IS3 ssgr IS150	8647481	8646531	951	316	8647748	8646385	1364	3	21/26
	ISRj2		8647660	8647355	306	101	8647748	8646385	1364	3	21/26
45	ISBd1	ISNCY ssgr ISLbi1	8667744	8669255	1512	503	8667655	8669397	1743	2	12/15
46	ISBj5	IS5 ssgr IS5	8672277	8673623	1347	448	8672108	8673631	1524	6	15/16
47	ISBd2	IS256	8676822	8675665	1158	385	8677028	8675634	1395	9	13
48	ISBja7	IS110 ssgr IS1111	8684999	8684034	966	321	8685008	8683855	1154	0	4/14
49	ISRj2	IS3 ssgr IS150	8722642	8721692	951	316	8722909	8721546	1364	3	21/26
	ISRj2		8722821	8722516	306	101	8722909	8721546	1364	3	21/26
50	ISRj2	IS3 ssgr IS150	8745901	8744951	951	316	8746167	8744805	1363	0	21/26
	ISRj2		8745954	8745775	180	59	8746167	8744805	1363	0	21/26

Tabela 7 – Continuação

IS	Família	Coord. inicial da ORF	Coord. final da ORF	Tam. da) ORF (pb)	Tam. da proteína	Coord. inicial do IS	Coord. final do IS	Tam. do IS (pb)	Tam. da DR (pb)	Tam. da IR (pb)	
		8746079	8745954	126	41	8746167	8744805	1363	0	21/26	
51	ISRj2										
	ISRj1	IS630	8747548	8747396	153	50	8748485	8747371	1115	2	6
	ISRj1		8747687	8747547	141	46	8748485	8747371	1115	2	6
	ISRj1		8748442	8747714	729	242	8748485	8747371	1115	2	6
52	ISBd9	IS21	8757716	8756955	762	253	8759407	8756876	2532	2	38/50
	ISBd9		8759276	8757735	1542	513	8759407	8756876	2532	2	38/50
53	ISBd12	IS21	8769261	8770760	1500	499	8769147	8771625	2479	0	23/26
	ISBd12		8770760	8771623	864	287	8769147	8771625	2479	0	23/26
54	ISBj4	IS110 ssgr IS1111	8791486	8792550	1065	354	8791411	8792988	1578	0	15
55	ISBd20	IS630	8794993	8793884	1110	369	8795057	8793873	1185	2	19/27
56	ISBja4	IS110 ssgr IS1111	8837288	8836257	1032	343	8837368	8835999	1370	0	13
57	ISBj5	IS5 ssgr IS5	8875644	8876990	1347	448	8875475	8876998	1524	6	15/16
58	ISBja1	IS481	9029108	9027930	1179	392	9029197	9027927	1271	0	15/18
59	ISBja1	IS481	9268738	9269916	1179	392	9268649	9269919	1271	5	15/18
60	ISBd9	IS21	9394463	9393754	710	236	9395979	9393675	2305	2	23/29
	ISBd9		9395833	9394622	1212	404	9395979	9393675	2305	2	23/29
61	ISBd23	IS5 ssgr IS427	9398826	9399218	393	130	9398764	9399594	831	4	9/15
	ISBd23		9399134	9399585	452	150	9398764	9399594	831	4	9/15
62	ISBd15	IS6	9407578	9408333	756	251	9407527	9408340	814	0	13/16
63	ISBd15	IS6	9429139	9428375	765	254	9429146	9428325	822	0	14/16
64	ISBd15	IS6	9433505	9432739	767	254	9433555	9432732	824	0	14/16
65	ISBd2	IS256	9438040	9436883	1158	385	9438246	9436852	1395	9	13
66	ISBd23	IS5 ssgr IS427	9439014	9439405	392	130	9438952	9439782	831	4	9/15
	ISBd23		9439322	9439773	452	150	9438952	9439782	831	4	9/15
67	ISBd15	IS6	9441608	9440850	759	252	9441658	9440843	816	0	14/16
68	ISRj1	IS630	9460137	9461201	1065	354	9460094	9461211	1118	2	6
69	ISBd2	IS256	9472997	9474155	1159	385	9472791	9474186	1396	4	13
70	ISRj1	IS630	9479652	9480716	1065	354	9479609	9480726	1118	2	6
71	ISBd23	IS5 ssgr IS427	9480792	9481184	393	130	9480730	9481561	832	4	9/15
	ISBd23		9481100	9481552	453	150	9480730	9481561	832	4	9/15
72	ISRj2	IS3 ssgr IS150	9493676	9493980	305	100	9493588	9494950	1363	3	21/26
	ISRj2		9493855	9494804	950	315	9493588	9494950	1363	3	21/26
73	ISRj1	IS630	9501388	9502452	1065	354	9501345	9502462	1118	2	6
74	ISRj2	IS3 ssgr IS150	9503486	9502626	861	286	9503753	9502480	1274	0	21/26
	ISRj2		9503665	9503360	306	101	9503753	9502480	1274	0	21/26
75	ISBd2	IS256	9534999	9533842	1158	385	9535205	9533811	1395	9	13
76	ISRj1	IS630	9555947	9556234	288	95	9555904	9557017	1114	2	6
	ISRj1		9556360	9557007	648	215	9555904	9557017	1114	2	6
77	ISBd23	IS5 ssgr IS427	9568691	9568239	453	150	9569061	9568230	832	4	9/15
	ISBd23		9568999	9568607	393	130	9569061	9568230	832	4	9/15
78	ISBd17	IS110 ssgr IS1111	9570832	9571863	1032	343	9570698	9572036	1339	3	5/12

Tabela 7 – Continuação

	IS	Família	Coord. inicial da ORF	Coord. final da ORF	Tam. da) ORF (pb)	Tam. da proteína	Coord. inicial do IS	Coord. final do IS	Tam. do IS (pb)	Tam. da DR (pb)	Tam. da IR (pb)
79	ISRj1	IS630	9572571	9573635	1065	354	9572528	9573645	1118	2	6
80	ISBj2	IS5 ssg IS5	9577545	9576211	1335	444	9577711	9576203	1509	6	13/18
81	ISBd9	IS21	9578167	9579379	1213	404	9578021	9580326	2306	2	37/50
	ISBd9		9579538	9580247	710	236	9578021	9580326	2306	2	37/50

Gene *modC* compartilhando sequência com o ISBd2. a- Estrutura do ISBd2; b- Sequência de aminoácido do gene *modC*, destacando a região codificada pelo ISBd2.

a- Estrutura do ISBd2

IR
Transposase

— Sequência compartilhada

>1395 nts

GAGCTATGCGGGA AACTGGGTGATGACGGAATGAGATAGGCGGCGTATCGAGGCGGGTGTGAGCCTGCC
AGAACCTCAAGGAGAGCGATACGCCATGAACGAGCATAGCAACATTGTCCACTGCGTCAGCCCGATGAGA
TCGACGATCCACTGACGAATATTTGCGATCTGGTGCTCGGCAGCTGCTTGCAGAGGCTGTGAGATGGAA
GCCGAGGCGTTTCTCGCCGCGATGAAGGGCTTGAAGCGTCCCGATGCGCGGACCGCCTCGTGCGGCAC
GGCCATGGTCCAGTGCGGACGATCCAGACGGGGATCGGCGCCGTGGAAGTCGCCCCGGGTAAAGATTGCGG
ATCGCGCGGTGACCAAGCATGGCGAGCGGATCCGCTTCACTCGGCGATCCTGCCGTTGTGGGCACGGC
GCACGAAGAGCTTGGATGCACTTTGCCGTTCTGTACCTGCGAGGCATCTCGACGGGCGACTTCCAGGAG
GCGCTGGCGGCGCTCTTGGGCAAGGATGCGCCGAATCTTCTCCGGGGGTTGTTTCCAGACTGACGACGG
AGTGGCAGCTCGAGTACGAGCGTTGGCAGAAGCGCGATCTGTGCGCGCGCCGGTACGTATACGTGTGGGC
GGACGGCGTCTTCTGCAAGGCTCGCATGGAAGACCACAGCGAATGCATGCTGGTGCTGATCGGCGCGACG
CCGGAAGGCAAGAAGGAACATCGGCTTCCAGGTCGGCGTGCGCGAGAGCACGCAGAGCTGGCACGAA
CTGCTCGTCGAGGCGAAAACCCGTGGGCTGAAGATCGCCCCGGAATCGCCGTGACGGCGCGCTC
GGCTTCTGGAAGGCGCTCGACGAGGTCTTCCCCGCCACGCGACATCAGCGGTGCTGGGTGCACAAAACCG
CGAATATCTTGAACAAAGTGCAGTGTCCGTACAGGCCAGCATGAAGAAGGATCTGCGCGAGGTCTATTGG
CGTCCAACCGAGCTTGGCCGAAGTGGCGATCGATGCTTTGCCGAGAAATACGGAGCGAAGTACGACAAG
GCGGTGAGTGCCTGACGAAAGATCGCGACGCAATGCTTGCCTTCTACGAATCCCCGCCGAGCATTGGGA
CCACTTGCGGACGACGAATCCATCGAAAGCGTGTTCGCGACGGTCCGGCACAGAACGGTGCGCACGAAA
GGTTGTTATCGTCAACGACTGCCAAGTTGATGGTGTTCAGGCTGCTCTGCGCCGCATCAAAGACCTGGCG
GCGGCTGAAAGGCACAAATCAGTTGCCGAAGGTGAGCGCAGGTGTGAGATTGAAAACGGCATCGAGGTC
ATCCAAGTGCCGAAAACACGCGCCCTGATCGCCTCGTCACCCAAAATCCCGCATAGCTC

b- Sequência codificada pelo gene *modC*

MGGFTKEAKKGYIEVAFNGSLASILLDTQFSVPAKGVTAIFGPPGCGKTTLARCIAGVQRLPNGFCAI
DGEIWQDETTFRPPHLRRVAYVFQLPILSSHLSVRRTLLYNASNSEPTLIDFDGVVELLGLAPLLERTP
SHLSKTERQRLALGSALLGQPRLLLLDDPLIVRDRSAKCEILPFLERILQSLALPMIYISHDITDIERLAD
HLVMMKDGTVTAAGPLNVTQSDPALASRSEAAICLDTMVGYYDGRYGLVKLRKLSARFLIPAVPLRP
GARLRLRIAAGDVSIACEPPRASSILNVFRARITASLPHGDAEVTLVLTLETGHSVPLLARITRRSFD
ALRLSIGAEVFAEELCGKLGDDGMR

Codificado pelo ISBd2

Proteínas hipotéticas que tiveram novas informações adicionadas

Proteína	Estirpes de <i>Bradyrhizobium</i>		PHI-base	Fenótipo	T: 21 DAI	P: 21 DAI	T: 28 DAI	VL
	CPAC 7	CPAC 15	ID		ID	ID	ID	ID
Biossíntese geral								
ent-caureno sintase	BJA5080_RS26465	BJS_RS01535	PHI:3360 X. oryzae	Red. da virul.	blr2150	blr2150	blr2150	-
ent-copalil difosfato sintase	BJA5080_RS26470	BJS_RS01530	-	-	blr2149	blr2149	blr2149	-
corismato mutase	BJA5080_RS34850	BJS_RS32955	-	-	blI1398	blI1398	-	-
2-oxoglutarato-Fe (II) oxigenase	BJA5080_RS17975	-	-	-	blr5430	blr5430	-	-
biossíntese de carboidrato	BJA5080_RS17980	-	-	-	blr5431	blr5431	-	-
acetolactato sintase	BJA5080_RS39045	-	-	-	blI7553	-	-	-
3-hidroxiclcanoato sintetase	BJA5080_RS07870	BJS_RS10465	-	-	-	blr3456	-	-
amp-dependente sintetase	-	BJS_RS16535	-	-	-	-	-	-
mono-oxigenase biossíntese de antibióticos	BJA5080_RS05665	-	-	-	-	-	-	-
mono-oxigenase biossíntese de antibióticos	BJA5080_RS33500	-	-	-	-	-	-	-
mono-oxigenase biossíntese de antibióticos	BJA5080_RS07610	-	-	-	-	blr3404	-	-
mono-oxigenase biossíntese de antibióticos	BJA5080_RS38990	BJS_RS31810	-	-	-	-	-	-
aspartato racemase	BJA5080_RS25500	BJS_RS28210	-	-	blI6811	-	-	-
biossíntese do sistema suf	BJA5080_RS12230	BJS_RS12450	-	-	blr4343	-	-	-
atp sintase	BJA5080_RS13040	-	-	-	-	-	-	-
diamide sintase	-	BJS_RS07485	-	-	-	-	-	-
d-aspartato ligase	BJA5080_RS16795	-	-	-	-	-	-	-
dihidrodipicolinato redutase	BJA5080_RS30540	BJS_RS37475	-	-	-	-	-	-
dt dp-4-dehydroramnose -epimerase	BJA5080_RS27490	BJS_RS43650	-	-	-	-	-	-
epimerase	-	BJS_RS25200	-	-	-	-	-	-
biossíntese de exopolissacarídeo	BJA5080_RS23185	-	-	-	-	-	-	-
biossíntese de exopolissacarídeo	-	BJS_RS11460	-	-	-	-	-	-
biossíntese de exopolissacarídeo	-	BJS_RS18025	-	-	-	-	-	-
biossíntese de exopolissacarídeo	-	BJS_RS16605	-	-	-	-	-	-
gdp-manose 6-dehidrogenase	BJA5080_RS08985	-	-	-	-	-	-	-
gmp sintase	-	BJS_RS07525	-	-	-	-	-	-
biossíntese cluster ferro-enxofre	BJA5080_RS12235	BJS_RS12445	-	-	-	-	-	-
biossíntese de lipopolissacarídeo	BJA5080_RS31290	BJS_RS36665	-	-	-	-	-	-
biossíntese de lipopolissacarídeo	-	BJS_RS11465	-	-	-	-	-	-
biossíntese do cluster metal-enxofre	BJA5080_RS08935	BJS_RS15710	-	-	blr3682	-	-	-
biossíntese da ubiquinona	-	BJS_RS23615	PHI:3364 M. marinum	Red. da virul.	-	-	-	-

Tabela 8 – Continuação

Proteína	Estirpes de <i>Bradyrhizobium</i>		PHI-base		T: 21 DAI	P: 21 DAI	T: 28 DAI	VL
	CPAC 7	CPAC 15	ID	Fenótipo	ID	ID	ID	ID
mur-ligase	BJA5080_RS38890	BJS_RS31695	-	-	-	-	-	-
peptídeo sintase não-ribossomal	BJA5080_RS27410	-	-	-	-	-	-	-
pantoato - beta-alanina-ligase	BJA5080_RS35900	BJS_RS43335	-	-	-	-	-	-
pantoato - beta-alanina-ligase	-	BJS_RS43340	-	-	-	-	-	-
peptídeo sintetase	BJA5080_RS09735	-	-	-	-	-	-	-
fitoquelatina sintase	BJA5080_RS19625	-	-	-	blI7663	-	-	-
fitoeno sintase	BJA5080_RS05585	BJS_RS18655	-	-	blr3001	-	-	-
policetídeo ciclase	-	BJS_RS21255	-	-	-	-	-	-
policetídeo desidrase ciclase	BJA5080_RS31445	-	-	-	-	-	-	-
policetídeo desidrase ciclase	BJA5080_RS14300	BJS_RS09805	-	-	-	-	-	blI4707
policetídeo ciclase	BJA5080_RS38050	BJS_RS30865	-	-	-	-	-	-
policetídeo ciclase	BJA5080_RS09440	BJS_RS15470	-	-	-	-	-	-
policetídeo ciclase	BJA5080_RS33605	BJS_RS41105	-	-	-	-	-	-
policetídeo ciclase	BJA5080_RS30695	BJS_RS37290	-	-	-	-	-	-
policetídeo sintase	-	BJS_RS25175	PHI:3090 P. larvae	Red. da virul.	-	-	-	-
biossíntese de polissacarídeo	BJA5080_RS39145	BJS_RS31970	-	-	-	-	-	-
biossíntese de polissacarídeo	BJA5080_RS02070	BJS_RS02735	-	-	-	-	-	-
biossíntese de polissacarídeo	BJA5080_RS02095	BJS_RS02760	-	-	blI2375	-	-	-
biossíntese de polissacarídeo	-	BJS_RS23690	-	-	-	-	-	-
biossíntese de polissacarídeo	-	BJS_RS23570	-	-	-	-	-	-
biossíntese de polissacarídeo	-	BJS_RS11495	-	-	-	-	-	-
éster ciclase	-	BJS_RS03305	-	-	-	-	-	-
éster ciclase	-	BJS_RS09895	-	-	-	-	-	-
biossíntese queosina	-	BJS_RS43530	-	-	-	-	-	-
tiamina-fosfato sintase	BJA5080_RS33785	-	-	-	-	blr1192	-	-
espermidina sintase	BJA5080_RS23550	-	-	-	-	-	-	-
terpenóide ciclase	BJA5080_RS23240	-	-	-	-	-	-	-
biossíntese de ubiquinona metiltransferase	BJA5080_RS03270	BJS_RS21045	-	-	-	-	-	-
biossíntese de ubiquinona menaquinona	-	BJS_RS23565	PHI:3364 M. marinum	Red. da virul.	-	-	-	-
4-hidroxi-3-metilbut-2-enilo difosfato redutase	BJA5080_RS38085	-	-	-	bsl7372	-	-	-
acetil-carboxilase	BJA5080_RS08715	-	-	-	-	-	-	-
glutacionil-espermidina sintase	BJA5080_RS26210	BJS_RS01760	-	-	-	-	-	-
oxigenase	-	BJS_RS40690	-	-	-	-	-	-
cobalamina-B12	-	BJS_RS34865	-	-	-	-	-	-
Processos Metabólicos								
glicosil hidrolase	BJA5080_RS02085	BJS_RS02750	-	-	-	blI2373	-	-
glicosil hidrolase	BJA5080_RS02110	BJS_RS02775	-	-	-	blI2378	-	-
glicosil hidrolase	-	BJS_RS11455	-	-	-	-	-	-
glicosil hidrolase 79	BJA5080_RS25450	BJS_RS28165	-	-	-	-	-	-
glicosil hidrolase	BJA5080_RS27110	BJS_RS00770	-	-	blr1964	blr1964	blr1964	-
glicosil hidrolase	BJA5080_RS27310	BJS_RS43455	-	-	blr1964	blr1964	blr1964	-
glicosil hidrolase	BJA5080_RS33185	-	-	-	-	-	-	-
glicosil hidrolase	BJA5080_RS07390	BJS_RS17015	-	-	blI3360	-	-	-

Tabela 8 – Continuação

Proteína	Estirpes de <i>Bradyrhizobium</i>		PHI-base		T: 21 DAI	P: 21 DAI	T: 28 DAI	VL
	CPAC 7	CPAC 15	ID	Fenótipo	ID	ID	ID	ID
glicosil-transferase	BJA5080_RS08770	-	-	-	-	-	-	-
glicosil-transferase	BJA5080_RS21630	-	PHI:3448 K. pneumoniae	Red. da virul.	-	-	-	-
glicosil-transferase	BJA5080_RS21685	-	PHI:2481 X. citri	Red. da virul.	-	-	-	-
glicosil-transferase	-	BJS_RS11485	-	-	-	-	-	-
glicosil-transferase	-	BJS_RS11480	-	-	-	-	-	-
glicosil transferase 2	-	BJS_RS20420	-	-	-	-	-	-
glicosil transferase 2	-	BJS_RS23555	-	-	-	-	-	-
glicosil transferase 2	BJA5080_RS21690	-	PHI:2481 Xanthomonas citri	Red. da virul.	-	-	-	-
glicosiltransferase	-	BJS_RS11490	-	-	-	-	-	-
glicosil transferase 2	BJA5080_RS26900	BJS_RS00990	PHI:31 C. albicans	Red. da virul.	-	-	-	-
glicosiltransferase AgID	BJA5080_RS17920	-	-	-	blI5419	-	-	-
glicosil transferase 1	BJA5080_RS23255	-	PHI:2938 X. fastidiosa	Perda da patogen.	-	-	-	-
glicosiltransferase 87	BJA5080_RS05885	BJS_RS18345	-	-	blr3063	-	-	-
glicosiltransferase 87	-	BJS_RS33155	-	-	-	-	-	-
glicosil bnr	-	BJS_RS21615	-	-	-	-	-	-
di-hidroxi-2-eno- aldolase	BJA5080_RS00070	BJS_RS40315	-	-	-	-	-	-
2og-fe oxigenase	BJA5080_RS26840	BJS_RS01050	-	-	blr2042	blr2042	-	-
3-carboxymuconate ciclase	-	BJS_RS42025	-	-	-	-	-	-
3-desidrogenase hidroxi-isobutirato	BJA5080_RS35665	-	-	-	-	-	-	-
3-desidrogenase hidroxi-isobutirato	-	BJS_RS00325	-	-	-	-	-	-
3-ceto-5-aminohexanoato	BJA5080_RS32510	BJS_RS35315	-	-	-	-	-	-
3-oxoacil-acp redutase	BJA5080_RS27735	BJS_RS40005	-	-	-	blI0021	-	-
4-oxalocrotonato tautomerase	BJA5080_RS07095	-	-	-	-	-	-	-
4-oxalocrotonato tautomerase	BJA5080_RS10705	-	-	-	-	-	-	-
6-aminohexanoate-cíclico-hidrolase	BJA5080_RS35855	-	-	-	-	-	-	-
acetoacetato decarboxilase	BJA5080_RS27765	BJS_RS39975	PHI:250 C. heterostrophus	Red. da virul.	-	blr0028	-	-
Acetoacetyl-redutase	-	BJS_RS31070	-	-	-	-	-	-
acetiltransferase	BJA5080_RS00445	BJS_RS00715	-	-	blI1906	blI1906	blI1906	-
acetiltransferase	-	BJS_RS20370	PHI:2529 A. fumigatus	Letal	-	-	-	-
acetiltransferase	BJA5080_RS14205	-	-	-	blI4688	-	-	-
acetiltransferase	BJA5080_RS07620	-	-	-	-	blr3406	-	-
Acil-desidrogenase	BJA5080_RS09320	BJS_RS22735	-	-	-	-	-	-
N-acil- aciltransferase	BJA5080_RS21325	-	-	-	-	-	-	-
Acil-sintetase	-	BJS_RS04065	PHI:3090 P. larvae	Red. da virul.	-	-	-	-
Acil-CoA-aciltransferase	BJA5080_RS16485	BJS_RS07645	-	-	-	-	-	-
aciltransferase	BJA5080_RS21695	-	-	-	-	-	-	-
aciltransferase	-	BJS_RS11505	-	-	-	-	-	-
aciltransferase	-	BJS_RS04715	-	-	-	-	-	-
aciltransferase 3	BJA5080_RS10080	-	PHI:3655 S. marcescens	Aumento da virul.	blI3890	-	-	-
aciltransferase 3	-	BJS_RS31815	-	-	-	-	-	-
aciltransferase 3	-	BJS_RS11470	-	-	-	-	-	-
aciltransferase 3	-	BJS_RS11475	-	-	-	-	-	-
aciltransferase	BJA5080_RS36310	BJS_RS29360	-	-	-	-	-	-

Tabela 8 – Continuação

Proteína	Estirpes de <i>Bradyrhizobium</i>		PHI-base	Fenótipo	T: 21 DAI	P: 21 DAI	T: 28 DAI	VL
	CPAC 7	CPAC 15	ID		ID	ID	ID	ID
álcool desidrogenase	BJA5080_RS06020	-	-	-	-	-	-	-
Aldeído de ativação	BJA5080_RS26845	BJS_RS01045	-	-	blr2041	blr2041	-	-
Aldeído de ativação	BJA5080_RS15570	-	-	-	-	-	-	-
Aldeído de ativação	-	BJS_RS15800	-	-	-	-	-	-
Aldolase-tipo de domínio TIM	BJA5080_RS04755	BJS_RS19420	-	-	-	-	-	-
Aldolase-tipo de domínio TIM	BJA5080_RS00285	BJS_RS00210	-	-	-	-	-	-
Aldolase-tipo de domínio TIM	BJA5080_RS00295	BJS_RS00225	-	-	-	-	-	-
fosfatase alcalina	BJA5080_RS36830	BJS_RS29815	-	-	-	-	-	-
alcanossulfonato mono-oxigenase	BJA5080_RS36875	-	-	-	blI7145	-	-	-
alofanato hidrolase	BJA5080_RS08720	-	PHI:2654 C. albicans	Red. da virul.	-	-	-	-
alfa beta hidrolase	-	BJS_RS29090	-	-	-	-	-	-
alfa beta hidrolase	-	BJS_RS11090	-	-	-	-	-	-
alfa beta hidrolase	-	BJS_RS16810	-	-	-	-	-	-
alfa beta hidrolase	-	BJS_RS31935	-	-	-	-	-	-
Alfa-cetoglutarato descarboxilase	BJA5080_RS24875	BJS_RS27560	-	-	-	-	-	-
Alfa-cetoglutarato descarboxilase	BJA5080_RS25655	BJS_RS28420	-	-	blI6844	-	-	-
Alfa-cetoglutarato descarboxilase	BJA5080_RS11070	BJS_RS13610	-	-	-	-	-	-
Alfa-cetoglutarato descarboxilase	BJA5080_RS14010	BJS_RS10110	-	-	blr4609	-	-	-
Alfa-cetoglutarato descarboxilase	BJA5080_RS17530	BJS_RS07030	-	-	-	-	-	-
Alfa-cetoglutarato descarboxilase	BJA5080_RS31570	BJS_RS36325	-	-	blI0775	-	-	-
Alfa-cetoglutarato descarboxilase	BJA5080_RS34515	BJS_RS33350	-	-	-	-	-	-
Alfa-cetoglutarato descarboxilase	-	BJS_RS30310	-	-	-	-	-	-
Domínio hidrolase	BJA5080_RS01820	BJS_RS02495	-	-	-	-	-	-
Domínio hidrolase	BJA5080_RS40710	BJS_RS20305	-	-	-	-	-	-
Domínio hidrolase	BJA5080_RS28255	BJS_RS39440	-	-	blI0132	-	-	-
Domínio hidrolase	BJA5080_RS37255	-	-	-	-	-	-	-
Domínio hidrolase	BJA5080_RS33210	-	-	-	-	-	-	-
Domínio hidrolase	BJA5080_RS02435	-	-	-	-	-	-	-
Domínio hidrolase	BJA5080_RS16090	-	-	-	-	-	-	-
Domínio hidrolase	-	BJS_RS25790	-	-	-	-	-	-
Domínio hidrolase	-	BJS_RS40985	-	-	-	-	-	-
aminoacrilato perácido redutase	BJA5080_RS28845	-	-	-	-	-	-	-
aminoglicósido 3-N-acetiltransferase	BJA5080_RS23245	-	-	-	-	-	-	-
aminoglicósido 3-N-acetiltransferase	BJA5080_RS23285	-	-	-	-	-	-	-
atpase archaea	BJA5080_RS12815	-	-	-	-	-	-	-
atpase archaea	BJA5080_RS12805	-	-	-	-	-	-	-
arilsulfato sulfotransferase	BJA5080_RS38735	BJS_RS31555	-	-	-	-	-	-
ATP-grasp	BJA5080_RS21000	BJS_RS22885	-	-	blI5849	-	-	-
ATP-grasp	BJA5080_RS26830	BJS_RS01060	-	-	blI2045	blI2045	-	-
ATPase	-	BJS_RS25225	-	-	-	-	-	-
ATPase	-	BJS_RS03865	-	-	-	-	-	-
ATPase	-	BJS_RS00375	-	-	-	-	-	-
ATPase	-	BJS_RS33675	-	-	-	-	-	-

Tabela 8 – Continuação

Proteína	Estirpes de <i>Bradyrhizobium</i>		PHI-base		T: 21 DAI	P: 21 DAI	T: 28 DAI	VL
	CPAC 7	CPAC 15	ID	Fenótipo	ID	ID	ID	ID
ATPase aaa	BJA5080_RS19350	-	-	-	-	-	-	-
ATPase aaa	BJA5080_RS33310	-	-	-	-	-	-	-
ATPase aaa	-	BJS_RS24945	-	-	-	-	-	-
ATPase aaa	BJA5080_RS38195	-	-	-	-	-	-	-
ATPase	-	BJS_RS10830	-	-	-	-	-	-
ATPase	BJA5080_RS33325	-	-	-	-	-	-	-
ATPase parcial	BJA5080_RS03605	-	-	-	blr5173	-	-	-
ATPase	-	BJS_RS10825	-	-	-	-	-	-
benzoilformato descarboxilase	BJA5080_RS11475	-	-	-	-	-	-	-
carboxilase biotina	-	BJS_RS01200	-	-	-	-	-	-
aminotransferase cadeia ramificada	BJA5080_RS39035	-	-	-	blI7551	blI7551	-	blI7551
Bromoperoxidase	BJA5080_RS36125	-	-	-	-	-	-	-
anidrase carbônica	-	BJS_RS32215	-	-	-	-	-	-
Carboximuconolactona descarboxilase	BJA5080_RS01515	-	-	-	-	blI6901	-	-
Carboximuconolactona descarboxilase	BJA5080_RS07850	-	-	-	blr3452	-	-	-
Flavoproteína colina desidrogenase	BJA5080_RS19245	-	-	-	-	-	-	-
Cisteína hidrolase	-	BJS_RS05290	-	-	-	-	-	-
citocromo c	BJA5080_RS00230	BJS_RS00150	-	-	-	-	-	-
citocromo c	-	BJS_RS15660	-	-	-	-	-	-
citocromo c oxidase de tipo cbb3	BJA5080_RS36430	-	-	-	blI7057	blI7057	-	-
citocromo C peroxidase	-	BJS_RS04120	-	-	-	-	-	-
citocromo c	-	BJS_RS15140	-	-	-	-	-	-
citocromo c	BJA5080_RS03990	BJS_RS04030	-	-	-	-	-	-
citocromo c	BJA5080_RS10975	BJS_RS13690	-	-	blr4073	-	-	-
citocromo c	BJA5080_RS11755	BJS_RS12895	-	-	-	-	-	-
citocromo c	BJA5080_RS18255	BJS_RS32710	PHI:2601 A. fumigatus	Misto	-	-	-	-
citocromo P450	BJA5080_RS26505	-	-	-	blr2143	-	-	-
citocromo P450	BJA5080_RS16020	-	-	-	-	-	-	-
deubiquitinase	-	BJS_RS00455	-	-	-	-	-	-
DTDP-4-redutase dehidroramnose	BJA5080_RS23180	-	-	-	-	-	-	-
DTDP-glicose -dehidratase	BJA5080_RS01985	-	-	-	-	-	-	-
Elongator 3 / domínio NifB	BJA5080_RS40680	BJS_RS34005	-	-	-	blr7973	-	-
enamidase	BJA5080_RS09765	-	-	-	-	blI3830	-	-
Endopeptidase, domínio P60	BJA5080_RS34870	-	-	-	blr1402	-	-	-
Enolase C-terminal	BJA5080_RS04745	BJS_RS19430	-	-	-	-	-	-
Enoil-hidratase	BJA5080_RS07780	-	-	-	-	-	-	-
esterase	-	BJS_RS25795	-	-	-	-	-	-
ferredoxina	BJA5080_RS00715	BJS_RS44000	-	-	blr1765	blr1765	blr1765	-
succinato desidrogenase	BJA5080_RS13970	BJS_RS10170	-	-	-	-	-	-
beta desidrogenase	-	BJS_RS40660	-	-	-	-	-	-
Formil-transferase	-	BJS_RS26895	-	-	-	-	-	-
Formil-transferase	-	BJS_RS25190	-	-	-	-	-	-
Frutose-bisfosfatase	BJA5080_RS15505	-	-	-	-	-	-	-

Tabela 8 – Continuação

Proteína	Estirpes de <i>Bradyrhizobium</i>		PHI-base		T: 21 DAI	P: 21 DAI	T: 28 DAI	VL
	CPAC 7	CPAC 15	ID	Fenótipo	ID	ID	ID	ID
Gama-carboximuconolactona descarboxilase	BJA5080_RS20090	BJS_RS21875	-	-	-	-	-	-
Gama-glutamilciclotransferase, domínio AIG2	BJA5080_RS37980	BJS_RS30805	-	-	-	-	-	-
GDP-manose pirofosfatase	-	BJS_RS03795	-	-	blI2567	-	-	-
lipase acilhidrolase	BJA5080_RS10585	-	-	-	-	-	-	-
Gluconato 2-desidrogenase	-	BJS_RS07765	-	-	-	-	-	-
glucosamina-6-fosf. desaminase	-	BJS_RS25525	PHI:206 C. albicans	Red. da virul.	-	-	-	-
glutaminase	-	BJS_RS06075	-	-	-	-	-	-
glutamina ciclotransferase	BJA5080_RS24630	BJS_RS27295	-	-	-	-	-	-
glutaciona-dependente formaldeído	BJA5080_RS18770	BJS_RS09115	-	-	-	-	-	-
glutelina	-	BJS_RS39480	-	-	-	-	-	-
glicosaminoglicanos	BJA5080_RS40880	-	-	-	-	-	-	-
glioxalase	BJA5080_RS13650	-	-	-	-	-	-	-
Domínio glioxalase	-	BJS_RS20830	-	-	-	-	-	-
glioxalase	BJA5080_RS01485	BJS_RS20260	-	-	-	blr6907	-	-
acetiltransferase	BJA5080_RS22270	-	-	-	-	-	-	-
GTPase	BJA5080_RS27935	BJS_RS39805	-	-	-	-	-	-
haloácido desalogenase	-	BJS_RS21645	-	-	-	-	-	-
Domínio HD/ PDEase	BJA5080_RS11570	BJS_RS13080	-	-	-	-	-	-
hidrolase	BJA5080_RS25255	-	-	-	-	-	-	-
hidrolase	-	BJS_RS13140	-	-	-	-	-	-
glioxilase hidrolase	BJA5080_RS32620	-	-	-	-	blI1007	-	-
glioxilase hidrolase	-	BJS_RS21620	-	-	-	-	-	-
hydroxiacilglutaciona hidrolase	BJA5080_RS30535	-	-	-	blr0577	-	-	-
cetoesteroide isomerase	BJA5080_RS36975	-	-	-	-	-	-	-
cetoesteroide isomerase	-	BJS_RS05005	-	-	-	-	-	-
cetoesteroide isomerase	BJA5080_RS21465	BJS_RS23385	-	-	blI5943	-	-	-
cinesina	BJA5080_RS41165	BJS_RS41030	-	-	-	-	-	-
L-2-amino-tiazolina-4- ác. carboxhidrolase	BJA5080_RS15270	BJS_RS08815	-	-	-	-	-	-
L-sorbose-na-desidrogenase	-	BJS_RS06565	-	-	-	-	-	-
lactato dedesidrogenase	BJA5080_RS21835	-	-	-	-	-	-	-
ATPase	BJA5080_RS08360	-	-	-	-	-	-	-
lipase	-	BJS_RS20540	-	-	-	-	-	-
Luciferase mono-oxigenase	BJA5080_RS07790	-	-	-	-	-	-	-
lisina descarboxilase	-	BJS_RS15265	-	-	-	blI3794	-	blI3794
lisina descarboxilase	-	BJS_RS28890	-	-	-	-	-	-
malato desidrogenase	BJA5080_RS05450	-	-	-	-	-	-	-
metalo-hidrolase	BJA5080_RS12370	BJS_RS12285	PHI:2340 B. glumae	Misto	blI4367	blI4367	-	blI4367
mono-oxigenase	BJA5080_RS32065	-	-	-	-	-	-	-
mono-oxigenase	BJA5080_RS08925	-	-	-	blr3680	-	-	-
Fad mono-oxigenase	-	BJS_RS11410	-	-	-	-	-	-
oxidase	-	BJS_RS04035	-	-	-	-	-	-
NADH-quinona oxidorreductase	BJA5080_RS29015	BJS_RS16365	-	-	-	-	-	-
Nitrilotriacetato mono-oxigenase	BJA5080_RS07250	-	-	-	-	-	-	-

Tabela 8 – Continuação

Proteína	Estirpes de <i>Bradyrhizobium</i>		PHI-base		T: 21 DAI	P: 21 DAI	T: 28 DAI	VL
	CPAC 7	CPAC 15	ID	Fenótipo	ID	ID	ID	ID
nitrito redutase	BJA5080_RS11080	BJS_RS13590	-	-	-	-	-	-
nitro-redutase	-	BJS_RS05430	-	-	-	-	-	-
nitro-redutase	-	BJS_RS05435	-	-	-	-	-	-
Oleato hidratase	BJA5080_RS34290	-	-	-	blr1289	blr1289	-	-
oxidorreductase	-	BJS_RS23610	-	-	-	-	-	-
patatina	-	BJS_RS00515	-	-	-	-	-	-
patatina fosfolipase	BJA5080_RS04035	-	-	-	-	-	-	-
ácido fosfatídico fosfatase	BJA5080_RS19470	-	-	-	-	-	-	-
ácido fosfatídico fosfatase	BJA5080_RS06150	-	-	-	-	-	-	-
fosfatidiletanolamina N-metiltransferase	BJA5080_RS33315	-	PHI:3364 M. marinum	Red. da virul.	-	-	-	-
Fosfoenolpiruvato fosfomutase	-	BJS_RS00615	-	-	-	-	-	-
fosfoesterase	-	BJS_RS06150	-	-	-	-	-	-
fosfohidrolase	BJA5080_RS19705	-	-	-	-	-	-	-
Fosfolipase	-	BJS_RS20565	-	-	-	-	-	-
metabolismo fosfonato	BJA5080_RS12270	BJS_RS12410	-	-	-	-	-	-
metabolismo fosfonato	BJA5080_RS21825,	BJS_RS11520	-	-	-	-	-	-
	BJA5080_RS04285							
fosfotransferase	BJA5080_RS17945	-	-	-	blr5424	-	-	-
fotossistema subunidade h	-	BJS_RS04985	-	-	-	-	-	-
fotossistema subunidade h	-	BJS_RS20475	-	-	-	-	-	-
fotossistema subunidade h parcial	-	BJS_RS05280	-	-	-	-	-	-
ACAD	BJA5080_RS19075	BJS_RS05815	-	-	-	-	-	-
hidrolase	-	BJS_RS25180	-	-	-	-	-	-
polissacarídeo liase	BJA5080_RS23230	-	-	-	-	-	-	-
Fmn-dependente	BJA5080_RS24815	BJS_RS27505	-	-	blr6679	-	-	-
proteína de fixação de nitrogênio putativa	BJA5080_RS00785	BJS_RS43935	-	-	-	-	-	-
piridoxal fosfato transferase	BJA5080_RS21005	BJS_RS22890	-	-	blh15850	-	-	-
piridoxal fosfato transferase	-	BJS_RS22545	-	-	-	-	-	-
piridoxal fosfato transferase	-	BJS_RS05095	-	-	-	-	-	-
pirrolo-quinolino quinona	-	BJS_RS33335	-	-	-	-	-	-
piruvato desidrogenase	-	BJS_RS05085	-	-	-	-	-	-
glucose desidrogenase	BJA5080_RS20585	-	-	-	-	-	-	-
ribulose-5-fosfato 3-epimerase	BJA5080_RS39475	BJS_RS32325	-	-	-	-	-	-
Domínio rossmann	BJA5080_RS21060	-	-	-	-	-	-	-
esterase tipo hidrolase	-	BJS_RS10245	-	-	-	-	-	-
Fad- dependente desaturase	BJA5080_RS05595	BJS_RS18645	-	-	blr3003	blr3003	-	-
Esqualeno-ciclase	-	BJS_RS26345	-	-	-	-	-	-
oxidação de enxofre	BJA5080_RS03695	-	-	-	-	-	-	-
oxidação de enxofre	-	BJS_RS35180	-	-	blh1013	-	-	-
pirofosfato de tiamina	BJA5080_RS06455	BJS_RS17795	-	-	-	-	-	-
dioxigenase	BJA5080_RS07495	BJS_RS16695	-	-	-	-	-	-
transglutaminase	-	BJS_RS05080	-	-	-	-	-	-
tripsina peptidase	-	BJS_RS00365	-	-	-	-	-	-

Tabela 8 – Continuação

Proteína	Estirpes de <i>Bradyrhizobium</i>		PHI-base	Fenótipo	T: 21 DAI	P: 21 DAI	T: 28 DAI	VL
	CPAC 7	CPAC 15			ID	ID	ID	ID
ubiquinol-citocromo c redutase	BJA5080_RS30495	-	-	-	-	-	-	-
UDP-glucose 4-epimerase	BJA5080_RS23335	-	-	-	-	-	-	-
uroporfirina III-c-metiltransferase	BJA5080_RS29355	BJS_RS21245	-	-	-	-	-	-
uroporfirina III-c-metiltransferase	-	BJS_RS23680	-	-	blI6009	-	-	-
xilose isomerase	BJA5080_RS08870	-	-	-	-	-	-	-
xilose isomerase	-	BJS_RS02200	-	-	-	-	-	-
xiluloquinase	-	BJS_RS21655	-	-	-	-	-	-
Hidroxietiltiazole quinase	BJA5080_RS15605	BJS_RS08485	-	-	blI4947	-	-	-
peptidase zn	BJA5080_RS27945	BJS_RS39795	-	-	-	-	-	-
Acetoacetyl-sintetase	BJA5080_RS07665	BJS_RS16085	-	-	-	-	-	-
acetil-coenzima A-sintetase	-	BJS_RS19405	-	-	-	-	-	-
Demetilmenaquinone metiltransferase	BJA5080_RS01680	BJS_RS22535	-	-	-	-	-	-
descarboxilase	BJA5080_RS09835	BJS_RS15000	-	-	blr3843	-	-	-
Dimetilanolilina mono-oxigenase	BJA5080_RS16790	-	-	-	-	-	-	-
Fad oxidoreductase	BJA5080_RS36415	-	-	-	-	-	-	-
Homoserina acetiltransferase	BJA5080_RS24805	BJS_RS27480	PHI:2597 A. fumigatus	Misto	blr6677	-	-	-
Homoserina acetiltransferase	BJA5080_RS40120	BJS_RS42190			-	blI7862	-	-
Homoserina acetiltransferase	-	BJS_RS42075			-	-	-	-
Nad-ependente de deacetilase	BJA5080_RS08800	-	-	-	-	-	-	-
Nad-dependente de epimerase	BJA5080_RS10210	-	-	-	blr3921	blr3921	-	blr3921
NADH desidrogenase ubiquinona Fe-S	BJA5080_RS16945	-	-	-	-	-	-	-
NADH oxidoreductase-ubiquinona	-	BJS_RS04775	-	-	-	-	-	-
NADH-ubiquinona oxidoreductase	BJA5080_RS37795	-	-	-	-	-	-	-
reação fotossintética	BJA5080_RS20020	BJS_RS41985	-	-	blI5656	-	-	-
Dissulfureto isomerase	-	BJS_RS03405	-	-	blI2497	-	-	blI2497
lumazina	BJA5080_RS24600	-	-	-	-	-	-	-
Serina acetiltransferase	BJA5080_RS12790	-	-	-	-	-	-	-
Serina desidratase	BJA5080_RS23380	-	-	-	-	-	-	-
transglutaminase	BJA5080_RS15180	-	-	-	-	-	-	-
S-adenosil-L-metionina metiltransferase	BJA5080_RS24340	BJS_RS27005	-	-	blI6585	-	-	-
S-adenosil-L-metionina metiltransferase	-	BJS_RS02345	-	-	-	-	-	-
S-adenosil-L-metionina metiltransferase	BJA5080_RS02080	BJS_RS02745	-	-	-	blr2372	-	blr2372
S-adenosil-L-metionina metiltransferase	BJA5080_RS21480	BJS_RS23400	-	-	blI5946	blI5946	-	-
S-adenosil-L-metionina metiltransferase	BJA5080_RS03630	-	-	-	blr0865	-	-	-
metiltransferase sam-dependente	BJA5080_RS04290	-	-	-	-	-	-	-
metiltransferase sam-dependente	-	BJS_RS32945	-	-	-	-	-	-
fibronectina de tipo III	BJA5080_RS02015	BJS_RS02680	-	-	-	blr2359	-	-
Alfa-glucano	BJA5080_RS19610	-	-	-	-	-	-	-
beta-glicanase	-	BJS_RS11500	-	-	-	-	-	-
Glicosiltransferase 1	BJA5080_RS23205	-	-	-	-	-	-	-
aldolase	BJA5080_RS09830	BJS_RS15005	-	-	-	-	-	-
Alquil sulfatase	BJA5080_RS04320	BJS_RS19915	-	-	-	blr2751	-	-
NUDIX hidrolase	BJA5080_RS39050	-	-	-	-	-	-	-

Tabela 8 – Continuação

Proteína	Estirpes de <i>Bradyrhizobium</i>		PHI-base		T: 21 DAI	P: 21 DAI	T: 28 DAI	VL
	CPAC 7	CPAC 15	ID	Fenótipo	ID	ID	ID	ID
NUDIX hidrolase	BJA5080_RS23800	BJS_RS26480	-	-	-	-	-	-
domínio NYN	BJA5080_RS01885	BJS_RS02545	-	-	-	blr5067	-	blr5067
metiltransferase 11	BJA5080_RS15840	-	-	-	-	-	-	-
metiltransferase 11	BJA5080_RS23215	-	-	-	-	-	-	-
metiltransferase 11	-	BJS_RS12715	-	-	-	-	-	-
metiltransferase 11	-	BJS_RS21350	PHI:3364 M. marinum	Red. da virul.	-	-	-	-
metiltransferase 11	-	BJS_RS03290	-	-	-	-	-	-
Funções Regulatórias								
adenilato-ciclase	BJA5080_RS04820	-	-	-	-	-	-	-
proteína de cache	BJA5080_RS21965	-	-	-	blr6052	-	-	-
adenilato-ciclase	-	BJS_RS29300	-	-	-	-	-	-
adenilato-ciclase	-	BJS_RS25185	-	-	-	-	-	-
adenilato-ciclase	-	BJS_RS25825	-	-	-	-	-	-
regulador transcricional	-	BJS_RS39745	-	-	-	-	-	-
adenilato-ciclase	-	BJS_RS15135	-	-	-	-	-	-
proteína reguladora	-	BJS_RS23915	-	-	-	-	-	-
regulador transcricional	-	BJS_RS20410	-	-	-	-	-	-
adenilato-ciclase	BJA5080_RS27390	-	-	-	blr1671	blr1671	-	-
adenilato-ciclase	BJA5080_RS25165	BJS_RS27845	-	-	blI6746	blI6746	-	-
adenilato-ciclase	BJA5080_RS05740	BJS_RS18490	-	-	-	-	-	-
proteína adenilação aminoácido	-	BJS_RS21450	-	-	-	-	-	-
efh sensor de cálcio	BJA5080_RS20140	-	-	-	-	-	-	-
crp FNR família regulador transcricional	-	BJS_RS17660	-	-	blI3193	-	-	-
família ciclase	-	BJS_RS08785	-	-	-	-	-	-
quinase	BJA5080_RS33375	-	-	-	-	-	-	-
Diguanilato-ciclase	BJA5080_RS35865	-	-	-	-	-	-	-
Diguanilato-ciclase	-	BJS_RS26780	-	-	-	-	-	-
Diguanilato-ciclase parcial	BJA5080_RS35860	-	-	-	-	-	-	-
regulador transcricional	BJA5080_RS40095	BJS_RS42215	-	-	-	-	-	-
regulador transcricional parcial	BJA5080_RS26450	BJS_RS01550	-	-	blr2153	-	-	-
regulador transcricional parcial	-	BJS_RS02185	-	-	-	-	-	-
regulador transcricional	BJA5080_RS18055	BJS_RS21355	-	-	-	-	-	-
regulador transcricional	BJA5080_RS06030	-	-	-	-	-	-	-
regulador transcricional	BJA5080_RS37815	-	-	-	-	-	-	-
regulador transcricional	BJA5080_RS13570	-	-	-	-	-	-	-
regulador transcricional	BJA5080_RS03450	-	-	-	-	-	-	-
regulador transcricional	BJA5080_RS34085	-	-	-	-	-	-	-
regulador transcricional	-	BJS_RS34200	-	-	-	-	-	-
regulador transcricional	-	BJS_RS26955	-	-	-	-	-	-
regulador transcricional	-	BJS_RS22315	-	-	-	-	-	-
regulador transcricional	-	BJS_RS26415	-	-	-	-	-	-
regulador transcricional	-	BJS_RS18595	-	-	-	-	-	-
regulador transcricional	-	BJS_RS19680	PHI:4176 D. solani	Aumento da virul.	-	-	-	-

Tabela 8 – Continuação

Proteína	Estirpes de <i>Bradyrhizobium</i>		PHI-base		T: 21 DAI	P: 21 DAI	T: 28 DAI	VL
	CPAC 7	CPAC 15	ID	Fenótipo	ID	ID	ID	ID
regulador transcricional	-	BJS_RS28730	-	-	-	-	-	-
regulador transcricional	-	BJS_RS23835	-	-	-	-	-	-
regulador transcricional	-	BJS_RS42465	-	-	-	-	-	-
regulador transcricional	-	BJS_RS12710	-	-	-	-	-	-
histidina	-	BJS_RS16380	-	-	-	-	-	-
histidina quinase	BJA5080_RS00610	BJS_RS44115	-	-	-	-	-	-
histidina quinase	BJA5080_RS28545	BJS_RS39185	-	-	-	-	-	-
histidina quinase	BJA5080_RS32050	BJS_RS35825	-	-	-	-	-	-
histidina quinase	BJA5080_RS03600	-	-	-	-	-	-	-
histidina quinase	BJA5080_RS36035	-	-	-	-	-	-	-
histidina quinase	-	BJS_RS04090	PHI:3072 X. oryzae	Red. da virul.	-	-	-	-
histidina quinase	-	BJS_RS04865	-	-	-	-	-	-
histidina quinase	-	BJS_RS41955	PHI:3072 X. oryzae	Red. da virul.	-	-	-	-
histidina quinase	-	BJS_RS05385	-	-	-	-	-	-
histidina quinase	-	BJS_RS38810	PHI:3440 P. aeruginosa	Misto	-	-	-	-
histidina quinase	-	BJS_RS04465	-	-	-	-	-	-
histidina quinase	-	BJS_RS23015	-	-	-	-	-	-
histidina parcial	-	BJS_RS05020	-	-	-	-	-	-
histidina parcial	-	BJS_RS41945	-	-	-	-	-	-
regulador negativo da expressão de beta-lactamase	-	BJS_RS24300	-	-	-	-	-	-
regulação de degradação do ácido fenilacético	BJA5080_RS18915	BJS_RS06005	-	-	-	-	-	-
quinase fosfato	-	BJS_RS37155	-	-	blr0637	blr0637	-	-
quinase	-	BJS_RS04135	-	-	-	-	-	-
receptor regulador de resposta	BJA5080_RS35695	-	-	-	blr1564	-	-	-
sensor de histidina quinase	-	BJS_RS35890	-	-	-	-	-	-
Transdução de sinal histidina quinase	-	BJS_RS17565	PHI:2729 X. oryzae	Red. da virul.	-	-	-	-
transdutor de sinal	BJA5080_RS22080	BJS_RS24145	-	-	blr6074	blr6074	-	-
Tetraciclina regulador transcricional, TetR	BJA5080_RS27890	BJS_RS39850	-	-	-	-	-	-
ativador de transcrição	BJA5080_RS38185	-	-	-	bsl7393	-	-	-
ativador de transcrição	-	BJS_RS32645	-	-	-	-	-	-
regulador transcricional LuxR	-	BJS_RS04440	-	-	-	-	-	-
ativador transcricional	-	BJS_RS30985	-	-	-	-	-	-
ativador transcricional	BJA5080_RS37580	-	-	-	-	-	-	-
regulador transcricional	BJA5080_RS30005	-	-	-	-	-	-	-
regulador transcricional	BJA5080_RS01830	BJS_RS02505	-	-	-	-	-	-
regulador transcricional	BJA5080_RS41350	BJS_RS40755	-	-	-	-	-	-
regulador transcricional	BJA5080_RS35365	BJS_RS42855	-	-	blr1492	blr1492	-	-
regulador transcricional	BJA5080_RS38215	-	-	-	-	-	-	-
regulador transcricional	BJA5080_RS21850	-	-	-	-	-	-	-
regulador transcricional	BJA5080_RS40860	-	-	-	-	-	-	-
regulador transcricional	BJA5080_RS19775	-	PHI:2684 S. enterica serovar	Red. da virul.	-	-	blt2109	-
regulador transcricional	BJA5080_RS04295	-	-	-	-	-	-	-
regulador transcricional	BJA5080_RS31185	-	-	-	bsr0701	-	-	-

Tabela 8 – Continuação

Proteína	Estirpes de <i>Bradyrhizobium</i>		PHI-base		T: 21 DAI	P: 21 DAI	T: 28 DAI	VL
	CPAC 7	CPAC 15	ID	Fenótipo	ID	ID	ID	ID
regulador transcricional	-	BJS_RS00410	-	-	-	-	-	-
regulador transcricional	-	BJS_RS10600	-	-	-	-	-	-
regulador transcricional	-	BJS_RS04375	-	-	-	-	-	-
regulador transcricional	-	BJS_RS04960	-	-	-	-	-	-
regulador transcricional	-	BJS_RS04095	-	-	-	-	-	-
regulador transcricional	-	BJS_RS04900	-	-	-	-	-	-
regulador transcricional	-	BJS_RS05355	PHI:3117 P. syringae	Misto	-	-	-	-
regulador transcricional	-	BJS_RS00475	-	-	-	-	-	-
regulador transcricional	-	BJS_RS05155	-	-	-	-	-	-
regulador transcricional	-	BJS_RS26440	-	-	-	-	-	-
regulador de transcrição, AbiEi	-	BJS_RS00530	-	-	-	-	-	-
regulador de transcrição, AbiEi	-	BJS_RS40670	-	-	-	-	-	-
regulador de transcrição, parcial	BJA5080_RS08960	BJS_RS15685	-	-	-	-	-	-
regulador transcricional	BJA5080_RS19680	BJS_RS36910	-	-	blI7670	-	-	-
sensor histidina quinase	-	BJS_RS05615	PHI:2439 B. glumae	Red. da virul.	-	-	-	-
família XRE regulador transcricional	BJA5080_RS23500	-	-	-	blr6338	-	-	-
família XRE regulador transcricional	BJA5080_RS03585	-	-	-	-	-	-	-
família XRE regulador transcricional	BJA5080_RS23545	-	-	-	-	-	-	-
proteína contendo o domínio CHRD	BJA5080_RS21295	BJS_RS23190	-	-	blr5909	-	-	-
domínio FIST, N-terminal	BJA5080_RS01695	-	-	-	-	-	-	-
domínio FIST, N-terminal	BJA5080_RS29250	-	-	-	-	-	-	-
polifosfato quinase	BJA5080_RS04615	BJS_RS19560	PHI:3634 M. tuberculosis	Red. da virul.	-	-	-	-
polifosfato quinase	BJA5080_RS23770	BJS_RS26450	PHI:3634 M. tuberculosis	Red. da virul.	-	blI6474	-	-
Proteínas de ligação								
proteína ligação (2Fe-2s)	BJA5080_RS00770	-	-	-	blr1753	-	-	-
proteína de ligação ao cálcio	BJA5080_RS15880	BJS_RS25265	-	-	-	-	-	-
proteína de ligação de DNA	BJA5080_RS33290	-	-	-	-	-	-	-
proteína de ligação de DNA	BJA5080_RS13625	-	-	-	-	-	-	-
proteína de ligação de DNA	-	BJS_RS12565	-	-	-	-	-	-
proteína de ligação de DNA	-	BJS_RS16655	-	-	-	-	-	-
proteína de ligação de DNA	-	BJS_RS10685	-	-	-	-	-	-
proteína de ligação de DNA	-	BJS_RS05320	-	-	-	-	-	-
proteína de ligação de DNA	BJA5080_RS24625	BJS_RS27290	-	-	-	-	-	-
proteína de ligação de soluto extracelular	BJA5080_RS10945	BJS_RS13715	-	-	-	-	-	-
proteína de ligação- fibronectina parcial	BJA5080_RS27235	-	-	-	-	blI1926	-	-
globina	BJA5080_RS32965	-	-	-	-	-	-	-
domínio de ligação de metal	BJA5080_RS37420	-	-	-	blr7247	blr7247	-	-
proteína de ligação de penicilina	-	BJS_RS28400	-	-	-	-	-	-
Repetição de pentapeptídeo	BJA5080_RS22745	BJS_RS24725	PHI:3420 B. pseudomallei	Red. da virul.	-	-	-	-
proteína de ligação de peptídeos	BJA5080_RS08365	-	-	-	-	-	-	-
proteína de ligação de peptidoglicano	BJA5080_RS28010	BJS_RS39705	-	-	-	-	-	-
proteína de ligação de peptidoglicano	BJA5080_RS16830	-	-	-	-	-	-	-
proteína fosfatase fosfotirosina	BJA5080_RS19260	-	-	-	-	-	-	-

Tabela 8 – Continuação

Proteína	Estirpes de <i>Bradyrhizobium</i>		PHI-base		T: 21 DAI	P: 21 DAI	T: 28 DAI	VL
	CPAC 7	CPAC 15	ID	Fenótipo	ID	ID	ID	ID
reação fotossintética	-	BJS_RS22525	-	-	-	-	-	-
proteína de ligação de DNA	-	BJS_RS02195	-	-	-	-	-	-
repetição sel1	BJA5080_RS11455	BJS_RS13160	-	-	-	-	-	-
repetição sel1	BJA5080_RS33595	BJS_RS41120	-	-	bl11155	bl11155	-	-
repetição sel1	BJA5080_RS35485	-	-	-	-	-	-	-
proteína de ligação de DNA	BJA5080_RS12185	-	-	-	bl14334	-	-	-
Tetratricoceptídeo	BJA5080_RS16220	-	-	-	-	-	-	-
Tetratricoceptídeo	-	BJS_RS16050	-	-	-	-	-	-
Tetratricoceptídeo	-	BJS_RS30120	-	-	-	-	-	-
Tetratricoceptídeo	-	BJS_RS22245	-	-	-	-	-	-
Tetratricoceptídeo	BJA5080_RS02475	BJS_RS03150	-	-	-	-	-	-
Tetratricoceptídeo	BJA5080_RS14360	BJS_RS09760	-	-	-	-	-	-
Tetratricoceptídeo	BJA5080_RS18540	BJS_RS06390	-	-	-	-	-	-
Tetratricoceptídeo	BJA5080_RS36350	BJS_RS29400	-	-	bl17042	-	-	-
Tetratricoceptídeo	BJA5080_RS31325	BJS_RS36625	-	-	-	-	-	-
Tetratricoceptídeo	BJA5080_RS36745	BJS_RS29735	-	-	-	-	-	-
Tetratricoceptídeo	BJA5080_RS20845	BJS_RS22750	-	-	-	-	-	-
Tetratricoceptídeo	BJA5080_RS13860	-	-	-	-	-	-	-
Tetratricoceptídeo	BJA5080_RS34385	BJS_RS33470	-	-	blr1308	-	-	-
Tetratricoceptídeo	BJA5080_RS03830	-	-	-	-	-	-	-
Tetratricoceptídeo	BJA5080_RS30285	BJS_RS37700	-	-	-	-	-	-
Tetratricoceptídeo	BJA5080_RS34315	BJS_RS33545	-	-	-	bl11294	-	-
Tetratricoceptídeo parcial	-	BJS_RS02270	-	-	-	-	-	-
Domínio WD40/YVTN	BJA5080_RS23825	BJS_RS26505	-	-	-	-	-	-
Domínio WD40/YVTN	BJA5080_RS08965	BJS_RS15680	-	-	-	-	-	-
domínio Winged- ligação de DNA de ligação	BJA5080_RS01550	BJS_RS28645	-	-	blr6894	-	-	-
domínio Winged- ligação de DNA de ligação	BJA5080_RS40840	-	-	-	-	-	-	-
domínio Winged- ligação de DNA de ligação	-	BJS_RS11670	-	-	-	-	-	-
Repetição de anquirina	BJA5080_RS35770	-	-	-	-	-	-	-
domínio GYF	BJA5080_RS30675	BJS_RS37310	-	-	-	-	-	-
domínio GYF	-	BJS_RS11445	-	-	-	-	-	-
Homeodomínio	BJA5080_RS12900	-	-	-	-	-	-	-
Domínio OST-HTH/LOTUS	BJA5080_RS06420	BJS_RS17830	-	-	blr3164	blr3164	-	-
Metabolismo de DNA								
5- nucleotidase	BJA5080_RS19360	-	-	-	-	-	-	-
Aminoacil-tRNA sintetase	BJA5080_RS23130	-	-	-	bl10957	-	-	-
Domínio arc	-	BJS_RS10900	-	-	-	-	-	-
aspartil-tRNA glutamil amidotransferase	-	BJS_RS06120	-	-	-	-	-	-
Aspartil-tRNA sintetase	BJA5080_RS31530	-	-	-	blr0767	-	-	-
ATPase envolvida no reparo do DNA	BJA5080_RS00455	BJS_RS00725	-	-	-	-	-	-
ATPase envolvida no reparo do DNA	BJA5080_RS37265	-	-	-	-	-	-	-
DNA polimerase bifuncional primase	BJA5080_RS18380	-	-	-	-	-	-	-
cruzamento endodeoxyribonuclease junção	BJA5080_RS35270	BJS_RS05310	-	-	bsl1473	-	-	-

Tabela 8 – Continuação

Proteína	Estirpes de <i>Bradyrhizobium</i>		PHI-base		T: 21 DAI	P: 21 DAI	T: 28 DAI	VL
	CPAC 7	CPAC 15	ID	Fenótipo	ID	ID	ID	ID
Cisteinil-tRNA sintetase	-	BJS_RS15830	-	-	-	-	-	-
helicase	BJA5080_RS26280	BJS_RS01695	-	-	-	-	-	-
helicase	BJA5080_RS26290	-	-	-	-	-	-	-
helicase	BJA5080_RS26430	-	-	-	-	-	-	-
helicase	BJA5080_RS03640	-	-	-	-	-	-	-
helicase	-	BJS_RS40675	-	-	-	-	-	-
DNA-ligase	BJA5080_RS15850	-	-	-	-	-	-	-
DNA-ligase	BJA5080_RS19810	-	-	-	-	-	-	-
DNA polimerase	-	BJS_RS02880	-	-	-	-	-	-
DNA polimerase III	BJA5080_RS08975	-	-	-	-	-	-	-
DNA primase	BJA5080_RS00095	BJS_RS40340	-	-	-	-	-	-
DNA primase	BJA5080_RS13575	-	-	-	-	-	-	-
DNA primase	BJA5080_RS18375	-	-	-	-	-	-	-
DNA primase	-	BJS_RS05465	-	-	-	-	-	-
Reparo de DNA	BJA5080_RS12575	-	-	-	-	-	-	-
Reparo de DNA	-	BJS_RS25450	-	-	-	-	-	-
Polimerase v - reparo de DNA	BJA5080_RS41410	-	-	-	-	-	-	-
DnaB	BJA5080_RS33215	-	-	-	-	-	-	-
dna topoisomerase subunidade iv b	BJA5080_RS29335	BJS_RS38685	-	-	-	-	-	-
DNA-3-metiladenina glicosilase	-	BJS_RS25780	-	-	-	-	-	-
endoribonuclease	BJA5080_RS14145	-	-	-	blI4678	blI4678	-	-
h-ns família histona	BJA5080_RS12565	-	-	-	-	-	-	-
helicase	BJA5080_RS03625	-	-	-	-	-	-	-
helicase	BJA5080_RS03635	-	-	-	-	-	-	-
endonuclease hnh	BJA5080_RS38235	-	-	-	-	-	-	-
endonuclease hnh	BJA5080_RS13620	-	-	-	-	-	-	-
endonuclease hnh	BJA5080_RS12570	-	-	-	-	-	-	-
endonuclease hnh	BJA5080_RS03575	-	-	-	-	-	-	-
endonuclease hnh	-	BJS_RS03860	-	-	-	-	-	-
Metionil-tRNA formiltransferase	BJA5080_RS08980	-	-	-	-	-	-	-
nuclease	BJA5080_RS19745	-	-	-	-	-	-	-
nuclease	BJA5080_RS27245	-	-	-	blI1924	-	-	-
nucleóide	-	BJS_RS34350	-	-	-	-	-	-
nuclease	BJA5080_RS18430	-	-	-	-	-	-	-
polimerase	-	BJS_RS16595	-	-	-	-	-	-
polimerase	-	BJS_RS43690	-	-	-	-	-	-
polimerase	-	BJS_RS00295	-	-	-	-	-	-
exonuclease SbcCD	BJA5080_RS27365	-	-	-	-	-	-	-
Piridina	-	BJS_RS40930	-	-	-	-	-	-
proteína de replicação	-	BJS_RS20590	-	-	-	-	-	-
endonuclease de restrição	-	BJS_RS14170	-	-	-	-	-	-
maturase transcriptase reversa	BJA5080_RS41395	BJS_RS40595	-	-	blr1915	-	-	-
maturase transcriptase reversa	BJA5080_RS41400	BJS_RS40590	-	-	blr1915	-	-	-

Tabela 8 – Continuação

Proteína	Estirpes de <i>Bradyrhizobium</i>		PHI-base		T: 21 DAI	P: 21 DAI	T: 28 DAI	VL
	CPAC 7	CPAC 15	ID	Fenótipo	ID	ID	ID	ID
domínio CopG	-	BJS_RS44275	-	-	-	-	-	-
ribonuclease	BJA5080_RS34725	BJS_RS33080	-	-	-	-	-	-
ribonuclease	-	BJS_RS05510	-	-	-	-	-	-
ribonuclease	BJA5080_RS33285	-	-	-	-	-	-	-
ribonuclease	-	BJS_RS34260	-	-	-	-	-	-
RNA helicase	BJA5080_RS40900	-	-	-	-	-	-	-
metiltransferase rna	BJA5080_RS39280	BJS_RS32085	-	-	-	-	-	-
serina parcial	BJA5080_RS38225	-	-	-	-	-	-	-
serina recombinase	BJA5080_RS27695	BJS_RS40045	-	-	-	-	-	-
serina recombinase	-	BJS_RS34230	-	-	-	-	-	-
serina recombinase	-	BJS_RS34240	-	-	-	-	-	-
serina recombinase	-	BJS_RS34235	-	-	-	-	-	-
serina recombinase	-	BJS_RS34245	-	-	-	-	-	-
invertase	-	BJS_RS05115	-	-	-	-	-	-
recombinase	BJA5080_RS38240	-	-	-	-	-	-	-
dna invertase parcial	BJA5080_RS30980	-	-	-	-	-	-	-
topoisomerase II	BJA5080_RS26115	BJS_RS01855	-	-	-	-	-	-
fator de iniciação de tradução 2	BJA5080_RS16585	-	-	-	blr5152	-	-	-
inibidor de iniciação da tradução	-	BJS_RS02235	-	-	-	-	-	-
proteína da maquinaria de tradução	BJA5080_RS12670	-	-	-	-	blI4412	-	-
tRNA endonuclease	BJA5080_RS27970	BJS_RS39770	-	-	-	-	-	-
fator de elongação	BJA5080_RS16765	-	-	-	-	-	-	-
timidilato-sintase	BJA5080_RS15785	BJS_RS08305	-	-	-	-	-	-
Metalo-dependente fosfatase	-	BJS_RS07800	-	-	-	-	-	-
Metalo-dependente fosfatase	BJA5080_RS27260	-	-	-	-	-	-	-
metiltransferase	BJA5080_RS23640	-	-	-	blI6449	-	-	-
metiltransferase	BJA5080_RS18000	-	-	-	blr5435	-	-	-
metiltransferase	BJA5080_RS27335	BJS_RS43490	-	-	-	-	-	-
metiltransferase	BJA5080_RS12875	-	-	-	-	-	-	-
nucleosídeo fosforilase	BJA5080_RS27980	BJS_RS39750	-	-	-	-	-	-
nucleosídeo fosforilase	BJA5080_RS31145	BJS_RS36855	-	-	blI0693	-	-	-
Nucleotideo-difosfato-transferase	BJA5080_RS30120	BJS_RS37885	-	-	blr0496	-	-	-
nucleosídeo trifosfato hidrolase	BJA5080_RS02665	BJS_RS03380	-	-	blI1791	blI2492	-	-
nucleosídeo trifosfato hidrolase	BJA5080_RS00640	BJS_RS44080	-	-	blI1791	blI1791	-	-
nucleosídeo trifosfato hidrolase	BJA5080_RS00645	BJS_RS44075	-	-	blI1790	-	-	-
nucleosídeo trifosfato hidrolase	BJA5080_RS33355	-	-	-	-	-	-	-
nucleosídeo trifosfato hidrolase	BJA5080_RS33295	-	-	-	-	-	-	-
nucleosídeo trifosfato hidrolase	BJA5080_RS33240	-	-	-	-	-	-	-
nucleosídeo trifosfato hidrolase	BJA5080_RS41340	-	-	-	-	blI8177	-	-
nucleosídeo trifosfato hidrolase	-	BJS_RS43505	-	-	-	-	-	-
nucleotidiltransferase	BJA5080_RS33660	BJS_RS41165	-	-	-	-	-	-
domínio CopG	-	BJS_RS43255	-	-	-	-	-	-

Tabela 8 – Continuação

Proteína	Estirpes de <i>Bradyrhizobium</i>		PHI-base	Fenótipo	T: 21 DAI	P: 21 DAI	T: 28 DAI	VL
	CPAC 7	CPAC 15			ID	ID	ID	ID
Integ., transposases e fagos								
infecção abortiva de fago Lactococcus II	-	BJS_RS20635	-	-	-	-	-	-
resistência a infecção de fago	BJA5080_RS30970	-	-	-	-	-	-	-
infecção abortiva de fago, família AbiV	BJA5080_RS16680	-	-	-	-	-	-	-
proteína de bacteriófago	BJA5080_RS12870	-	-	-	-	-	-	-
proteína de bacteriófago	BJA5080_RS12960	-	-	-	-	-	-	-
integrase	BJA5080_RS18360	-	-	-	-	-	-	-
DNA de fago	-	BJS_RS04435	-	-	-	-	-	-
integrase	BJA5080_RS34060	-	-	-	-	-	-	-
integrase	-	BJS_RS03855	-	-	-	-	-	-
integrase	-	BJS_RS40210	-	-	-	-	-	-
integrase	-	BJS_RS23825	-	-	-	-	-	-
integrase	-	BJS_RS40200	-	-	-	-	-	-
integrase	BJA5080_RS27630	BJS_RS40105	-	-	-	-	-	-
integrase	BJA5080_RS19635	-	-	-	-	-	-	-
integrase	-	BJS_RS24025	-	-	-	-	-	-
integrase	-	BJS_RS20730	-	-	-	-	-	-
integrase	-	BJS_RS40475	-	-	blr8234	-	-	-
cauda fago- lambda	BJA5080_RS13010	-	-	-	-	-	-	-
Micobacteriófago D29	BJA5080_RS27915	BJS_RS39825	-	-	-	-	-	-
proteína da cápside do fago	BJA5080_RS12950	-	-	-	-	-	-	-
terminase de fago	BJA5080_RS17025	-	-	-	-	-	-	-
integrase de fago	BJA5080_RS21840	-	-	-	-	-	-	-
integrase de fago	-	BJS_RS03820	-	-	-	-	-	-
integrase de fago	BJA5080_RS35935	BJS_RS43375	-	-	blr1635	-	-	-
integrase de fago	BJA5080_RS40825	-	-	-	-	-	-	-
fago parcial	-	BJS_RS28900	-	-	-	-	-	-
DNA de fago	BJA5080_RS00100	-	-	-	-	-	-	-
fago hk97	BJA5080_RS12940	-	-	-	-	-	-	-
proteína de fago	BJA5080_RS12880	-	-	-	-	-	-	-
proteína de fago	-	BJS_RS04645	-	-	-	-	-	-
terminase de fago parcial	BJA5080_RS22205	BJS_RS24265	-	-	-	-	-	-
profago cp4-57	-	BJS_RS25235	-	-	-	-	-	-
profago cp4-57	BJA5080_RS17050	-	-	-	-	-	-	-
resolvase	BJA5080_RS27775	-	-	-	-	-	-	-
resolvase	BJA5080_RS40905	-	-	-	-	-	-	-
resolvase	-	BJS_RS21285	-	-	-	-	-	-
resolvase	-	BJS_RS01490	-	-	-	-	-	-
resolvase	BJA5080_RS12930	-	-	-	-	-	-	-
integrase	-	BJS_RS04565	-	-	-	-	-	-
Proteína Tn7	BJA5080_RS19215	-	-	-	-	-	-	-
transposase	BJA5080_RS26460	BJS_RS01540	-	-	-	-	-	-
transposase	BJA5080_RS27095	BJS_RS00785	-	-	-	-	-	-

Tabela 8 – Continuação

Proteína	Estirpes de <i>Bradyrhizobium</i>		PHI-base		T: 21 DAI	P: 21 DAI	T: 28 DAI	VL
	CPAC 7	CPAC 15	ID	Fenótipo	ID	ID	ID	ID
transposase	BJA5080_RS00175	BJS_RS40455	-	-	-	-	-	-
transposase	BJA5080_RS00405	BJS_RS00675	-	-	-	-	-	-
transposase	BJA5080_RS00450	BJS_RS00720	-	-	bll1685	-	-	-
transposase	BJA5080_RS41475	BJS_RS40520	-	-	-	-	-	-
transposase	BJA5080_RS27680	BJS_RS40055	-	-	-	-	-	-
transposase	BJA5080_RS26420	-	-	-	-	-	-	-
transposase	BJA5080_RS41470	-	-	-	-	-	-	-
transposase	BJA5080_RS27430	-	-	-	bsr1677	-	-	-
transposase	BJA5080_RS27715	-	-	-	-	-	-	-
transposase	BJA5080_RS27285	-	-	-	-	-	-	-
transposase	BJA5080_RS23535	-	-	-	-	-	-	-
transposase	BJA5080_RS40850	-	-	-	-	-	-	-
transposase	BJA5080_RS41510	-	-	-	-	-	-	-
transposase	-	BJS_RS43400	-	-	-	-	-	-
transposase	-	BJS_RS40650	-	-	-	-	-	-
transposase	-	BJS_RS40490	-	-	bll4232	blr8233	-	-
transposase	-	BJS_RS28960	-	-	-	-	-	-
transposase	-	BJS_RS03950	-	-	-	-	-	-
transposase	-	BJS_RS01560	-	-	-	-	-	-
transposase	-	BJS_RS42860	-	-	-	-	-	-
transposase	BJA5080_RS25460	-	-	-	-	-	-	-
proteína de fago	BJA5080_RS39330	BJS_RS32135	-	-	-	-	-	-
proteína de fago	-	BJS_RS25000	-	-	-	-	-	-
Bacteriófago phiKZ	BJA5080_RS40070	BJS_RS42240	-	-	bll7854	-	-	-
proteína de fago, GP25	-	BJS_RS24995	-	-	-	-	-	-
proteína de fago	BJA5080_RS35975	BJS_RS43415	-	-	-	-	-	-
terminase de fago	BJA5080_RS34090	-	-	-	-	-	-	-
Transporte e prot. de membrana								
abc transportador	BJA5080_RS12885	-	-	-	-	-	-	-
abc transportador	-	BJS_RS00965	-	-	bll2018	-	bll2018	-
ABC transportador substrato de ligação	-	BJS_RS02115	-	-	-	-	-	-
ABC transportador substrato de ligação	BJA5080_RS09190	-	-	-	-	-	-	-
ABC transportador substrato de ligação	BJA5080_RS06775	-	-	-	-	-	-	-
ABC transportador substrato de ligação	BJA5080_RS22785	-	-	-	-	-	-	-
ABC transportador substrato de ligação	-	BJS_RS25205	-	-	-	-	-	-
ABC transportador substrato de ligação	-	BJS_RS20800	-	-	-	-	-	-
ABC transportador substrato de ligação	-	BJS_RS15985	-	-	-	-	-	-
ABC transportador substrato de ligação	-	BJS_RS21295	-	-	-	-	-	-
ABC transportador substrato de ligação	-	BJS_RS38660	-	-	-	-	-	-
ABC transportador substrato de ligação	-	BJS_RS17420	-	-	-	-	-	-
ABC transportador substrato de ligação	-	BJS_RS22170	-	-	-	-	-	-
ABC transportador substrato de ligação	BJA5080_RS05975	BJS_RS15030	-	-	-	-	-	-
abc transportador	-	BJS_RS15750	-	-	-	-	-	-

Tabela 8 – Continuação

Proteína	Estirpes de <i>Bradyrhizobium</i>		PHI-base		T: 21 DAI	P: 21 DAI	T: 28 DAI	VL
	CPAC 7	CPAC 15	ID	Fenótipo	ID	ID	ID	ID
ABC transportador substrato de ligação	-	BJS_RS15600	-	-	-	-	-	-
transporte de bicarbonatosulfonato	-	BJS_RS14955	-	-	-	-	-	-
Transporte periplasmático	-	BJS_RS40980	-	-	-	-	-	-
ABC transportador – permease	BJA5080_RS27440	-	-	-	blr1679	-	-	-
Anti-sigma fator K	BJA5080_RS26175	BJS_RS01795	-	-	-	-	-	-
Autotransportador domínio-beta	BJA5080_RS02270	BJS_RS02955	-	-	-	-	-	-
Autotransportador	BJA5080_RS39760	BJS_RS32720	-	-	-	-	-	-
transportador de biopolímero	BJA5080_RS20180	-	-	-	-	-	-	-
ABC transportador – permease	BJA5080_RS01420	BJS_RS02305	-	-	-	-	-	-
ABC transportador – permease	BJA5080_RS14055	-	-	-	-	-	-	-
Beta-caroteno	BJA5080_RS26950	BJS_RS00940	-	-	blI2009	-	-	-
proteína BUG	BJA5080_RS14640	BJS_RS09440	-	-	blI4776	-	-	-
proteína BUG	BJA5080_RS28235	BJS_RS39460	-	-	-	-	-	-
proteína BUG	BJA5080_RS29750	BJS_RS38245	-	-	-	-	-	-
proteína BUG	BJA5080_RS32350	BJS_RS35490	-	-	-	-	-	-
proteína BUG	BJA5080_RS22790	BJS_RS24775	-	-	-	-	-	-
proteína BUG	BJA5080_RS23005	BJS_RS25275	-	-	-	-	-	-
proteína BUG	-	BJS_RS13500	-	-	-	-	-	-
proteína BUG	-	BJS_RS16930	-	-	-	-	-	-
proteína BUG	BJA5080_RS23965	BJS_RS26635	-	-	blI6513	-	-	-
proteína BUG	BJA5080_RS04065	BJS_RS20155	-	-	-	-	-	-
proteína BUG	BJA5080_RS23750	BJS_RS26425	-	-	-	-	-	-
cálcio: transporte de prótons	-	BJS_RS19970	-	-	-	-	-	-
transportador de cromato	BJA5080_RS18180	-	-	-	-	-	-	-
ACEA- família de transportador	-	BJS_RS08960	-	-	-	-	-	-
proteína extracelular	-	BJS_RS31860	-	-	-	-	-	-
f-type h + -Transportor de atpase	BJA5080_RS36090	BJS_RS29200	-	-	-	-	-	-
f-type h + -Transportor de atpase	-	BJS_RS11585	-	-	-	-	-	-
f-type h + -Transportor de atpase	-	BJS_RS04105	-	-	-	-	-	-
f-type h + -Transportor de atpase	-	BJS_RS25140	-	-	-	-	-	-
sistema de transporte	BJA5080_RS36370	BJS_RS29420	-	-	-	-	-	-
lipoproteína	-	BJS_RS11385	-	-	-	-	-	-
glutaciona permease	BJA5080_RS03660	-	-	-	-	-	-	-
glicina betaína permease	-	BJS_RS02325	-	-	-	-	-	-
transportador de níquel	-	BJS_RS21605	-	-	-	-	-	-
transportador RND	-	BJS_RS38780	-	-	-	-	-	-
proteína de membrana interna	BJA5080_RS23865	BJS_RS26545	-	-	blr6493	-	-	-
proteína de membrana interna	-	BJS_RS41880	-	-	-	-	-	-
proteína de membrana interna	-	BJS_RS41920	-	-	-	-	-	-
proteína de membrana interna	-	BJS_RS41950	-	-	-	-	-	-
abc transportador de ferro	-	BJS_RS22550	-	-	-	-	-	-
transportador de ferro	BJA5080_RS36015	-	-	-	-	-	-	-
proteína de ligação a leucina	-	BJS_RS41550	-	-	-	-	-	-

Tabela 8 – Continuação

Proteína	Estirpes de <i>Bradyrhizobium</i>		PHI-base		T: 21 DAI	P: 21 DAI	T: 28 DAI	VL
	CPAC 7	CPAC 15	ID	Fenótipo	ID	ID	ID	ID
lipoproteína	BJA5080_RS37520	BJS_RS30535	-	-	-	blI7267	-	-
lipoproteína	BJA5080_RS29835	-	-	-	blr0444	-	-	-
lipoproteína	BJA5080_RS37260	-	-	-	-	-	-	-
precursor de lipoproteína	BJA5080_RS22365	-	-	-	-	-	-	-
Proteína associada a membrana	BJA5080_RS10530	BJS_RS14210	-	-	-	-	-	-
membrana parcial	-	BJS_RS42080	-	-	-	-	-	-
proteína de membrana	BJA5080_RS23695	BJS_RS26340	-	-	bsr6460	-	-	-
proteína de membrana	BJA5080_RS02185	BJS_RS02860	-	-	blr2393	-	-	-
proteína de membrana	BJA5080_RS03850	BJS_RS12995	-	-	-	-	-	-
proteína de membrana	BJA5080_RS15925	BJS_RS08215	-	-	blr5024	blr5024	-	-
proteína de membrana	BJA5080_RS16000	BJS_RS08140	-	-	-	-	-	-
proteína de membrana	BJA5080_RS30880	BJS_RS37105	-	-	-	-	-	-
proteína de membrana	BJA5080_RS33710	BJS_RS41220	-	-	-	-	-	-
proteína de membrana	BJA5080_RS20365	BJS_RS22185	-	-	blI5722	-	-	-
proteína de membrana	BJA5080_RS16950	-	-	-	-	-	-	-
proteína de membrana	BJA5080_RS37455	-	-	-	-	blr7254	-	-
proteína de membrana	BJA5080_RS10330	-	-	-	-	-	-	-
proteína de membrana	BJA5080_RS07860	-	-	-	blr3454	-	-	-
proteína de membrana	BJA5080_RS30825	-	-	-	blI0636	-	-	-
proteína de membrana	-	BJS_RS42910	-	-	-	-	-	-
proteína de membrana	-	BJS_RS35880	-	-	-	-	-	-
proteína de membrana	-	BJS_RS11715	-	-	blr4470	-	-	-
proteína de membrana	-	BJS_RS27875	-	-	bsr6752	-	-	-
proteína de membrana	-	BJS_RS05450	-	-	-	-	-	-
proteína de membrana	-	BJS_RS12385	-	-	-	-	-	-
proteína de membrana	-	BJS_RS38825	-	-	-	-	-	-
proteína de membrana	-	BJS_RS38325	-	-	-	-	-	-
proteína de membrana	-	BJS_RS29070	-	-	-	-	-	-
proteína de membrana	-	BJS_RS42990	-	-	-	-	-	-
proteína de membrana	-	BJS_RS05955	-	-	-	-	-	-
proteína de membrana	-	BJS_RS38830	-	-	-	-	-	-
proteína de membrana	-	BJS_RS14285	-	-	-	-	-	-
proteína de membrana para captação de colicina	BJA5080_RS25945	BJS_RS02105	-	-	blr2248	-	-	-
proteína de membrana	-	BJS_RS23620	-	-	-	-	-	-
proteína de membrana	BJA5080_RS29635	BJS_RS38405	-	-	-	-	-	-
proteína de membrana- metaloendopeptidase	BJA5080_RS14475	BJS_RS09620	-	-	blI4743	-	-	-
proteína de membrana- metaloendopeptidase	-	BJS_RS07915	-	-	-	-	-	-
MFS transportador	-	BJS_RS29325	-	-	-	-	-	-
MFS transportador	BJA5080_RS26325	BJS_RS01650	-	-	-	-	-	-
MFS transportador	BJA5080_RS08215	BJS_RS15905	-	-	-	-	-	-
MFS transportador	BJA5080_RS29645	-	-	-	-	-	-	-
MFS transportador	-	BJS_RS11100	-	-	-	-	-	-
MFS transportador	-	BJS_RS04190	-	-	-	-	-	-

Tabela 8 – Continuação

Proteína	Estirpes de <i>Bradyrhizobium</i>		PHI-base		T: 21 DAI	P: 21 DAI	T: 28 DAI	VL
	CPAC 7	CPAC 15	ID	Fenótipo	ID	ID	ID	ID
MFS transportador	-	BJS_RS15405	-	-	-	-	-	-
MFS transportador	BJA5080_RS18895	-	-	-	-	-	-	-
permease	BJA5080_RS04050	-	-	-	-	-	-	-
multidroga MFS transportador	-	BJS_RS25915	PHI:2731 V. dahliae	Perda da patogen.	-	-	-	-
lipoproteína da membrana externa	BJA5080_RS24385	-	-	-	-	-	-	-
proteína de membrana externa	BJA5080_RS34665	-	-	-	-	-	-	-
proteína de membrana externa	BJA5080_RS38295	BJS_RS31050	-	-	blI7405	-	-	-
proteína de membrana externa	BJA5080_RS14730	-	-	-	-	-	-	-
proteína da membrana externa, OmpA	-	BJS_RS00380	-	-	-	-	-	-
proteína da membrana externa, OmpA	BJA5080_RS08605	-	-	-	blr3613	-	-	-
proteína da membrana externa, OmpA	-	BJS_RS11685	-	-	-	-	-	-
proteína da membrana externa, PagP	-	BJS_RS00660	-	-	blI2085	-	-	-
proteína da membrana externa, OmpA	-	BJS_RS31055	-	-	blI7406	-	-	-
Transporte de ácido para-hidroxibenzóico	-	BJS_RS41805	-	-	-	-	-	-
abc transportador	BJA5080_RS12910	-	-	-	-	-	-	-
abc transportador deníquel	BJA5080_RS07725	BJS_RS15960	-	-	-	-	-	-
proteína periplasmática	BJA5080_RS18595	BJS_RS06320	-	-	-	-	-	-
proteína periplasmática	BJA5080_RS40445	BJS_RS41575	-	-	blr7925	-	-	-
proteína de membrana externa polimórfica	BJA5080_RS14335	-	-	-	blr4714	-	-	-
porina	-	BJS_RS26040	-	-	-	-	-	-
exportador de proteína rica em glutamina	BJA5080_RS19505	-	-	-	-	-	-	-
lipoproteína rara A	BJA5080_RS23740	-	-	-	-	-	-	-
lipoproteína rara A	-	BJS_RS25195	-	-	-	-	-	-
transportador de açúcar	BJA5080_RS31215	BJS_RS36805	-	-	-	-	-	-
Domínio TolB	BJA5080_RS40160	BJS_RS42125	-	-	-	-	-	-
Domínio TolB	BJA5080_RS34890	BJS_RS32905	-	-	blI1406	-	-	-
Domínio TolB	BJA5080_RS08755	-	-	-	-	blI1368	-	-
Domínio TolB	-	BJS_RS25125	-	-	-	-	-	-
Domínio TolB	-	BJS_RS16520	-	-	-	-	-	-
sódio: transporte de prótons	-	BJS_RS25590	-	-	-	-	-	-
transportador de açúcar	BJA5080_RS21715	-	PHI:2067 M. oryzae	Perda da patogen.	-	-	-	-
transportador de açúcar	-	BJS_RS16580	-	-	-	-	-	-
permease de fosfato de açúcar	BJA5080_RS06025	-	-	-	-	-	-	-
proteína de efluxo treonina	BJA5080_RS01555	-	-	-	-	-	-	-
Domínio TolB	BJA5080_RS39770	BJS_RS32730	-	-	blr7794	-	-	-
proteína transmembranar	BJA5080_RS03845	BJS_RS18040	-	-	-	-	-	-
proteína transmembranar 189	BJA5080_RS02360	-	-	-	-	-	-	-
proteína transmembranar TauE	BJA5080_RS02580	BJS_RS03275	-	-	blr2470	-	-	-
transportador	BJA5080_RS07625	-	-	-	-	-	-	-
transportador	-	BJS_RS35500	-	-	blI6514	-	-	-
transportador	-	BJS_RS22390	-	-	-	-	-	-
transportador	-	BJS_RS12580	-	-	-	-	-	-
transportador tricarboxilate	BJA5080_RS21775	-	-	-	blI6034	-	-	-

Tabela 8 – Continuação

Proteína	Estirpes de <i>Bradyrhizobium</i>		PHI-base		T: 21 DAI	P: 21 DAI	T: 28 DAI	VL
	CPAC 7	CPAC 15	ID	Fenótipo	ID	ID	ID	ID
transportador tricarboxilate	BJA5080_RS07810	-	-	-	-	-	-	-
translocação twin-arginina	BJA5080_RS17565	-	-	-	-	-	-	-
translocação twin-arginina	BJA5080_RS38725	-	-	-	blr7490	-	-	-
translocação twin-arginina	BJA5080_RS18955	-	-	-	blI5592	-	-	-
translocação twin-arginina	-	BJS_RS28735	-	-	-	-	-	-
MFS transportador	BJA5080_RS16660	BJS_RS07505	-	-	-	-	-	-
proteína transmembranar	BJA5080_RS41490	BJS_RS40505	-	-	-	-	-	-
Repetição de cisteína	BJA5080_RS30445	BJS_RS37560	-	-	-	-	-	-
Repetição de cisteína	BJA5080_RS34520	BJS_RS33330	-	-	-	-	-	-
Repetição de cisteína	BJA5080_RS16735	-	-	-	-	-	-	-
Repetição de cisteína	-	BJS_RS10630	-	-	-	-	-	-
domínio ACEA	-	BJS_RS14065	-	-	-	-	-	-
domínio ACEA	-	BJS_RS01195	-	-	-	-	-	-
Superfamília NTF2	-	BJS_RS40945	-	-	-	-	-	-
proteína transmembranar	BJA5080_RS34795	-	-	-	-	-	-	-
Processos Celulares								
Adesina	-	BJS_RS26170	PHI:3354 V. vulnificus	Red. da virul.	-	-	-	-
Adesina	BJA5080_RS18675	-	PHI:3450 B. cenocepacia	Red. da virul.	-	-	-	-
proteína de quimiotaixia	BJA5080_RS18495	BJS_RS06430	-	-	-	-	-	-
proteína de quimiotaixia	BJA5080_RS34790	-	-	-	-	blr1386	-	-
proteína flagelar	BJA5080_RS14670	BJS_RS09410	-	-	blI4781	-	-	-
proteína flagelar	BJA5080_RS20870	BJS_RS22765	-	-	blI5823	-	-	-
proteína flagelar FlaG	BJA5080_RS20960	BJS_RS22855	-	-	-	-	-	-
proteína de quimiotaixia	BJA5080_RS28945	BJS_RS16300	-	-	-	-	-	-
proteína de quimiotaixia	-	BJS_RS11435	-	-	-	-	-	-
lectina	BJA5080_RS20760	BJS_RS22655	-	-	-	-	-	-
Proteína de conjugação parcial	BJA5080_RS00970	BJS_RS43720	-	-	-	-	-	-
proteína de conjugação	BJA5080_RS27170	-	-	-	-	-	-	-
proteína de conjugação	BJA5080_RS00015	-	-	-	-	-	-	-
proteína de conjugação	BJA5080_RS21910	-	-	-	-	-	-	-
proteína de conjugação	-	BJS_RS01430	-	-	-	-	-	-
proteína de conjugação	-	BJS_RS34225	-	-	-	-	-	-
proteína de montagem pilus	BJA5080_RS03275	BJS_RS21040	-	-	-	-	-	-
proteína de montagem pilus	-	BJS_RS21550	-	-	-	-	-	-
proteína de montagem pilus, CPAD	BJA5080_RS27540	BJS_RS00015	-	-	blI1841	-	-	-
proteína de montagem pilus, Flp	BJA5080_RS10310	BJS_RS14455	-	-	-	-	-	-
proteína de secreção rica em cisteína	BJA5080_RS35335	BJS_RS42825	-	-	blr1486	-	-	-
proteína de secreção	BJA5080_RS01065	-	-	-	-	-	-	-
proteína de secreção, HlyD	BJA5080_RS09890	BJS_RS38320	-	-	-	-	-	-
proteína de secreção	BJA5080_RS21955	-	-	-	-	-	-	-
proteína de secreção	BJA5080_RS31555	-	-	-	-	-	-	-
proteína de secreção	-	BJS_RS23720	PHI:3662 X. fastidiosa	Red. da virul.	-	-	-	-
proteína de secreção	-	BJS_RS21210	-	-	-	-	-	-

Tabela 8 – Continuação

Proteína	Estirpes de <i>Bradyrhizobium</i>		PHI-base	Fenótipo	T: 21 DAI	P: 21 DAI	T: 28 DAI	VL
	CPAC 7	CPAC 15	ID		ID	ID	ID	ID
Proteína de secreção	BJA5080_RS18820	BJS_RS06110	-	-	blr5567	-	-	-
Proteína de secreção I	-	BJS_RS21215	-	-	-	-	-	-
Proteína de secreção I	-	BJS_RS20395	-	-	-	-	-	-
Proteína de secreção I	-	BJS_RS20405	-	-	-	-	-	-
Proteína de secreção IV	BJA5080_RS38220	-	-	-	-	-	-	-
Proteína de secreção IV	-	BJS_RS34220	-	-	-	-	-	-
Proteína de secreção IV	-	BJS_RS20585	-	-	-	-	-	-
Proteína de secreção VI	BJA5080_RS03685	-	-	-	-	-	-	-
Proteína de secreção IV	BJA5080_RS27455	BJS_RS43615	-	-	-	-	-	-
família BA14k - fator de virulência	BJA5080_RS36865	BJS_RS07750	-	-	-	-	-	-
família BA14k - fator de virulência	-	BJS_RS25240	-	-	-	-	-	-
família BA14k - fator de virulência	BJA5080_RS11910	BJS_RS12740	-	-	-	-	-	-
BrnA antitoxina	BJA5080_RS14165	BJS_RS09945	-	-	bsl4681	-	-	-
Antitoxina	BJA5080_RS16685	-	-	-	-	-	-	-
AbiEii toxina, tipo de sistema IV TA	BJA5080_RS40855	-	-	-	-	-	-	-
AbiEii toxina, tipo de sistema IV TA	BJA5080_RS27230	-	-	-	blI1927	blI1927	-	-
AbiEii toxina, tipo de sistema IV TA	-	BJS_RS43165	-	-	-	-	-	-
AbiEii toxina, tipo de sistema IV TA	-	BJS_RS00535	-	-	-	-	-	-
Toxina endonuclease, YhaV	BJA5080_RS27495	BJS_RS43655	-	-	-	-	-	-
Antitoxina	-	BJS_RS05350	-	-	-	-	-	-
Antitoxina	-	BJS_RS30995	-	-	-	-	-	-
fator de virulência	BJA5080_RS18060	-	-	-	-	-	-	-
família BrkB- fator de virulência	BJA5080_RS19330	BJS_RS11675	-	-	-	-	-	-
Fator de virulência E	BJA5080_RS34075	-	-	-	-	-	-	-
Fator de virulência E	BJA5080_RS40845	-	-	-	-	-	-	-
Proteína VGR	BJA5080_RS08460	BJS_RS25730	-	-	-	-	-	-
Proteína VirK	-	BJS_RS42070	-	-	-	-	-	-
pectina liase	-	BJS_RS23765	-	-	-	-	-	-
Antitoxina	BJA5080_RS26590	BJS_RS01335	-	-	-	-	-	-
Antitoxina	-	BJS_RS23705	-	-	-	-	-	-
Efeto tipo III	BJA5080_RS26815	BJS_RS01075	-	-	bsl2050	-	-	-
Efeto tipo III	BJA5080_RS00530	BJS_RS44185	-	-	-	-	-	-
Efeto tipo III	BJA5080_RS00600	BJS_RS44120	-	-	-	-	-	-
Efeto tipo III	BJA5080_RS00190	BJS_RS00110	-	-	blr1869	-	-	-
Efeto tipo III	BJA5080_RS27535	BJS_RS00010	-	-	blI1840	-	-	-
Efeto tipo III	BJA5080_RS27565	BJS_RS00040	-	-	blI1848	-	-	-
Efeto tipo III	BJA5080_RS27615	BJS_RS40610	-	-	blI1862	-	-	-
Efeto tipo III	BJA5080_RS30960	-	-	-	blI1840	-	-	-
Efeto tipo III	BJA5080_RS00605	-	-	-	-	-	-	-
Efeto tipo III	BJA5080_RS00585	BJS_RS44135	-	-	-	-	-	-
Efeto tipo III – LRR	BJA5080_RS27425	BJS_RS43585	PHI:3367 Y. pestis	Avirulência	blr1676	blr1676	-	-
Efeto tipo III	BJA5080_RS27290	BJS_RS43440	-	-	-	-	-	-
proteína secretada	BJA5080_RS24460	BJS_RS27130	-	-	blI6609	-	-	-

Tabela 8 – Continuação

Proteína	Estirpes de <i>Bradyrhizobium</i>		PHI-base		T: 21 DAI	P: 21 DAI	T: 28 DAI	VL
	CPAC 7	CPAC 15	ID	Fenótipo	ID	ID	ID	ID
proteína secretada	BJA5080_RS15340	BJS_RS08715	-	-	bsl4900	-	-	-
proteína secretada	BJA5080_RS30785	BJS_RS37205	-	-	-	blr0628	-	-
proteína secretada	BJA5080_RS22720	BJS_RS24700	-	-	-	-	-	-
proteína secretada	-	BJS_RS10895	-	-	-	-	-	-
Heme peroxidase	-	BJS_RS26190	-	-	-	-	-	-
glutaciona sintase	-	BJS_RS11560	-	-	-	-	-	-
glutaciona sintase	-	BJS_RS21595	-	-	-	-	-	-
glutaciona-S-transferase	-	BJS_RS16065	-	-	-	-	-	-
glutaciona sintase	-	BJS_RS05560	-	-	-	-	-	-
Inositol monofosfatase	-	BJS_RS21665	-	-	-	-	-	-
Domínio-AhpD	BJA5080_RS25910	BJS_RS02165	-	-	-	-	-	-
Domínio-AhpD	BJA5080_RS04110	BJS_RS20110	-	-	-	-	-	-
Domínio-AhpD	BJA5080_RS28755	-	-	-	-	blI0235	-	-
Domínio-AhpD	-	BJS_RS06940	-	-	-	-	-	-
alquil hidroperóxido reductase C	BJA5080_RS10270	BJS_RS14505	-	-	-	blr3933	-	-
catalase parcial	-	BJS_RS19950	-	-	-	-	-	-
metionina-S-óxido de redutase	-	BJS_RS04890	-	-	-	-	-	-
peroxidase	BJA5080_RS13455	BJS_RS11120	-	-	blr4509	-	-	-
tiorredoxina	BJA5080_RS31920	-	-	-	-	-	-	-
proteína osmoticamente induzida parcial	-	BJS_RS25880	-	-	-	-	-	-
selenoproteína	BJA5080_RS33030	BJS_RS34795	-	-	-	-	-	-
domínio caspase	BJA5080_RS08580	BJS_RS25610	-	-	-	-	-	-
domínio caspase	BJA5080_RS29370	BJS_RS38645	-	-	-	blr0352	-	-
Endopeptidase, NLPC	BJA5080_RS08575	BJS_RS25615	-	-	-	-	-	-
FtsH protease	BJA5080_RS16770	-	-	-	-	-	-	-
Glutamil protease	-	BJS_RS00355	-	-	-	-	-	-
metaloprotease	-	BJS_RS23175	-	-	-	-	-	-
metaloprotease	-	BJS_RS19955	-	-	-	-	-	-
Metalopeptidase	BJA5080_RS02485	BJS_RS03160	-	-	blI2452	-	-	-
peptidase	BJA5080_RS29460	BJS_RS38575	-	-	-	-	-	-
peptidase	BJA5080_RS07695	-	-	-	-	-	-	-
peptidase	BJA5080_RS00125	-	-	-	-	-	-	-
peptidase A26	-	BJS_RS01210	-	-	-	-	-	-
peptidase A26	-	BJS_RS01215	-	-	blI2085	-	-	-
Peptidase C13	-	BJS_RS08775	-	-	-	-	-	-
Peptidase C14	BJA5080_RS30775	-	-	-	blI0626	-	-	-
Peptidase C14	-	BJS_RS14700	-	-	-	-	-	-
Peptidase C14	-	BJS_RS29330	-	-	-	-	-	-
Peptidase C14	-	BJS_RS00395	-	-	-	-	-	-
Peptidase C14	BJA5080_RS30130	BJS_RS35585	-	-	blI3611	-	-	-
Peptidase C14	BJA5080_RS21255	BJS_RS23145	-	-	-	-	-	-
Peptidase M15	BJA5080_RS08590	BJS_RS25600	-	-	-	-	-	-
Peptidase M15	BJA5080_RS11790	BJS_RS12860	-	-	-	-	-	-

Tabela 8 – Continuação

Proteína	Estirpes de <i>Bradyrhizobium</i>		PHI-base		T: 21 DAI	P: 21 DAI	T: 28 DAI	VL
	CPAC 7	CPAC 15	ID	Fenótipo	ID	ID	ID	ID
Peptidase M20	-	BJS_RS26060	-	-	-	-	-	-
Peptidase M23	BJA5080_RS19675	BJS_RS36905	-	-	-	-	-	-
Peptidase S1	-	BJS_RS00400	-	-	-	-	-	-
Peptidase S1	-	BJS_RS42950	-	-	-	-	-	-
peptidase S49	BJA5080_RS12945	-	-	-	-	-	-	-
Peptidase S8	-	BJS_RS04955	-	-	-	-	-	-
peptidase U49	-	BJS_RS00420	-	-	-	-	-	-
peptidase	BJA5080_RS00120	-	-	-	-	-	-	-
protease	-	BJS_RS27500	PHI:2117 M. oryzae	Red. da virul.	-	-	-	-
protease	-	BJS_RS07905	-	-	-	-	-	-
peptidase S54	BJA5080_RS05355	BJS_RS18880	-	-	blr2953	-	-	-
serina protease	-	BJS_RS00370	-	-	-	-	-	-
cisteína peptidase	BJA5080_RS02040	BJS_RS02705	-	-	blI2364	-	-	-
transglutaminase	BJA5080_RS04655	BJS_RS19525	-	-	-	-	-	-
serina proteases, PDZ	BJA5080_RS26555	BJS_RS01400	PHI:183 C. albicans	Red. da virul.	-	-	-	-
Ulp1 protease	BJA5080_RS00155	BJS_RS43680, BJS_RS40430	-	-	-	blI8244	-	-
Ulp1 protease	BJA5080_RS30955	-	-	-	-	-	-	-
Ulp1 protease	BJA5080_RS00990	BJS_RS43745	-	-	blr1705	-	-	-
SOS peptidase	BJA5080_RS03045	BJS_RS03780	-	-	-	-	-	-
SOS peptidase	BJA5080_RS19925	-	-	-	-	-	-	-
SOS peptidase	BJA5080_RS22250	-	-	-	-	-	-	-
SOS peptidase	-	BJS_RS21340	-	-	-	-	-	-
ClpP, sítio ativo	-	BJS_RS23905	-	-	-	-	-	-
ClpP/crotonase	BJA5080_RS15195	-	-	-	blr4872	-	-	-
ClpP/crotonase	-	BJS_RS25120	-	-	-	-	-	-
regulador do ciclo celular	BJA5080_RS33080	BJS_RS34750	-	-	-	-	-	-
divisão celular	BJA5080_RS38970	BJS_RS31790	-	-	blI7538	blI7538	-	-
factor de coagulação 5/8 domínio C-terminal	-	BJS_RS14320	-	-	-	-	-	-
GCRA, regulador do ciclo celular	BJA5080_RS22380	BJS_RS24380	-	-	-	-	-	-
GCRA, regulador do ciclo celular	-	BJS_RS08300	-	-	-	-	-	-
proteína PEP2	BJA5080_RS39765	BJS_RS32725	PHI:224 N. haematococca	Red. da virul.	blI7793	blI7793	-	-
peptídeo parcial	-	BJS_RS43565	-	-	blr2108	blr2108	-	-
anquirina	BJA5080_RS35700	-	-	-	blI1566	-	-	-
proteína sec-C	BJA5080_RS41360	BJS_RS40745	-	-	-	-	-	-
proteína de germinação de esporos	-	BJS_RS15380	PHI:2620 P. aeruginosa	Red. da virul.	-	-	-	-
domínio relacionados com esporulação	BJA5080_RS14535	-	-	-	-	-	-	-
T3SS regulador negativo, GrlR	BJA5080_RS01965	BJS_RS02630	-	-	-	-	-	-
septo iniciador formação	BJA5080_RS14695	BJS_RS09380	-	-	blI4786	-	-	-
regulador de transcrição induzida pelo estresse	-	BJS_RS09235	-	-	-	-	-	-
Domínio RmlC	BJA5080_RS29000	BJS_RS16355	-	-	-	-	-	-
Domínio RmlC	BJA5080_RS35465	BJS_RS07705	-	-	blr1513	-	-	-
Domínio RmlC	BJA5080_RS09405	BJS_RS15525	-	-	blI3763	-	-	-

Tabela 8 – Continuação

Proteína	Estirpes de <i>Bradyrhizobium</i>		PHI-base		T: 21 DAI	P: 21 DAI	T: 28 DAI	VL
	CPAC 7	CPAC 15	ID	Fenótipo	ID	ID	ID	ID
Domínio RmlC	-	BJS_RS10570	-	-	-	-	-	-
Domínio RmlC	-	BJS_RS08790	-	-	-	-	-	-
Fasina	BJA5080_RS22355	BJS_RS24360	-	-	-	-	-	-
Fasina	BJA5080_RS21450	-	-	-	-	-	-	-
Fasina	BJA5080_RS16605	BJS_RS07560	-	-	-	blI5155	-	-
Fasina	-	BJS_RS21485	-	-	-	-	-	-
Fasina	-	BJS_RS12770	-	-	-	-	-	-
proteína de resistência ao cobre	-	BJS_RS08870	-	-	-	-	-	-
CsbD, proteína resposta ao estresse	BJA5080_RS02250	BJS_RS02935	-	-	-	-	-	-
proteína heme-degradantes	BJA5080_RS11505	-	-	-	-	-	-	-
hemolisina d	BJA5080_RS24960	-	-	-	-	-	-	-
heparinase	-	BJS_RS23560	-	-	-	-	-	-
heparinase	-	BJS_RS23575	-	-	-	-	-	-
invasão associada locus b	BJA5080_RS13905	BJS_RS10200	-	-	blI4589	-	-	blI4589
invasão associada locus b	BJA5080_RS30270	-	-	-	-	-	-	-
imunoglobulina	-	BJS_RS15775	-	-	-	-	-	-
peptídeo sinal	BJA5080_RS23400	BJS_RS16630	-	-	-	-	-	-
peptídeo sinal	BJA5080_RS23580	BJS_RS26180	-	-	-	-	-	-
peptídeo sinal	BJA5080_RS11465	-	-	-	blr4174	-	-	-
peptídeo sinal	BJA5080_RS08135	-	-	-	-	-	-	-
peptídeo sinal	BJA5080_RS24225	BJS_RS26870	-	-	blr6562	-	-	-
peptídeo sinal	BJA5080_RS24320	BJS_RS26985	-	-	-	-	-	-
peptídeo sinal	BJA5080_RS26160	BJS_RS01810	-	-	-	-	-	-
peptídeo sinal	BJA5080_RS03305	BJS_RS21015	-	-	blr2607	-	-	-
peptídeo sinal	BJA5080_RS03320	BJS_RS21000	-	-	-	-	-	-
peptídeo sinal	BJA5080_RS36065	BJS_RS29175	-	-	blI6993	-	-	-
peptídeo sinal	BJA5080_RS03400	BJS_RS20910	-	-	-	-	-	-
peptídeo sinal	BJA5080_RS04670	BJS_RS13135	-	-	blI2824	-	-	-
peptídeo sinal	BJA5080_RS05715	BJS_RS18515	-	-	-	-	-	-
peptídeo sinal	BJA5080_RS05955	BJS_RS18285	-	-	blr3076	-	-	-
peptídeo sinal	BJA5080_RS06660	BJS_RS17555	-	-	-	-	-	-
peptídeo sinal	BJA5080_RS38455	BJS_RS31220	-	-	-	-	-	-
peptídeo sinal	BJA5080_RS38515	BJS_RS31285	-	-	-	-	-	-
peptídeo sinal	BJA5080_RS38595	BJS_RS31405	-	-	-	-	-	-
peptídeo sinal	BJA5080_RS38600	BJS_RS31410	-	-	-	blr7465	-	-
peptídeo sinal	BJA5080_RS38690	BJS_RS31510	-	-	-	-	-	-
peptídeo sinal	BJA5080_RS10835	BJS_RS13850	-	-	-	-	-	-
peptídeo sinal	BJA5080_RS10860	BJS_RS13830	-	-	blr4051	-	-	-
peptídeo sinal	BJA5080_RS36135	BJS_RS29235	-	-	-	-	-	-
peptídeo sinal	BJA5080_RS11945	BJS_RS12690	-	-	-	-	-	-
peptídeo sinal	BJA5080_RS12110	BJS_RS12535	-	-	-	-	-	-
peptídeo sinal	BJA5080_RS38950	BJS_RS31770	-	-	blr7534	blr7534	-	blr7534
peptídeo sinal	BJA5080_RS13920	BJS_RS10185	PHI:3269 S. agalactiae	Red. da virul.	blr4592	-	-	-

Tabela 8 – Continuação

Proteína	Estirpes de <i>Bradyrhizobium</i>		PHI-base		T: 21 DAI	P: 21 DAI	T: 28 DAI	VL
	CPAC 7	CPAC 15	ID	Fenótipo	ID	ID	ID	ID
peptíeo sinal	BJA5080_RS14345	BJS_RS09775	-	-	blr4716	-	-	-
peptíeo sinal	BJA5080_RS15470	BJS_RS08590	-	-	-	-	-	-
peptíeo sinal	BJA5080_RS17410	BJS_RS07150	-	-	-	-	-	-
peptíeo sinal	BJA5080_RS17415	BJS_RS07145	-	-	-	-	-	-
peptíeo sinal	BJA5080_RS18345	BJS_RS06470	-	-	blI5501	-	-	-
peptíeo sinal	BJA5080_RS18530	BJS_RS06400	-	-	-	-	-	-
peptíeo sinal	BJA5080_RS18635	BJS_RS06285	-	-	-	-	-	-
peptíeo sinal	BJA5080_RS18655	BJS_RS06265	-	-	blI5535	-	-	-
peptíeo sinal	BJA5080_RS18725	BJS_RS06210	-	-	-	-	-	-
peptíeo sinal	BJA5080_RS39615	BJS_RS32470	-	-	-	-	-	-
peptíeo sinal	BJA5080_RS18940	BJS_RS05980	-	-	blI5589	-	-	-
peptíeo sinal	BJA5080_RS19490	BJS_RS04515	-	-	-	-	-	-
peptíeo sinal	BJA5080_RS40080	BJS_RS42230	-	-	-	-	-	-
peptíeo sinal	BJA5080_RS36555	BJS_RS29560	-	-	blI7080	-	-	-
peptíeo sinal	BJA5080_RS30450	BJS_RS37555	-	-	blI0561	-	-	-
peptíeo sinal	BJA5080_RS30640	BJS_RS37355	-	-	blI0598	-	-	-
peptíeo sinal	BJA5080_RS30945	BJS_RS37050	-	-	blI0660	-	-	-
peptíeo sinal	BJA5080_RS31070	BJS_RS36960	-	-	blI0679	blI0679	-	-
peptíeo sinal	BJA5080_RS31840	BJS_RS36030	-	-	-	-	-	-
peptíeo sinal	BJA5080_RS33465	BJS_RS34570	-	-	blr1130	-	-	-
peptíeo sinal	BJA5080_RS36985	BJS_RS29985	-	-	-	blr7166	-	-
peptíeo sinal	BJA5080_RS34300	BJS_RS33550	-	-	-	-	-	-
peptíeo sinal	BJA5080_RS34395	BJS_RS33460	-	-	-	-	-	-
peptíeo sinal	BJA5080_RS34510	BJS_RS33355	-	-	-	-	-	-
peptíeo sinal	BJA5080_RS34575	BJS_RS33270	-	-	-	-	-	-
peptíeo sinal	BJA5080_RS34795	BJS_RS33010	-	-	-	-	-	-
peptíeo sinal	BJA5080_RS35235	BJS_RS42735	-	-	-	-	-	-
peptíeo sinal	BJA5080_RS20025	BJS_RS21815	-	-	blI5657	-	-	-
peptíeo sinal	BJA5080_RS20170	BJS_RS21955	-	-	-	-	-	-
peptíeo sinal	BJA5080_RS37310	BJS_RS30320	-	-	-	-	-	-
peptíeo sinal	BJA5080_RS20730	BJS_RS22615	-	-	bsr5794	-	-	-
peptíeo sinal	BJA5080_RS21095	BJS_RS22975	-	-	-	blr5868	-	-
peptíeo sinal	BJA5080_RS21165	BJS_RS23040	-	-	-	-	-	-
peptíeo sinal	BJA5080_RS21245	BJS_RS23135	-	-	-	-	-	-
peptíeo sinal	BJA5080_RS37500	BJS_RS30515	-	-	-	-	-	-
peptíeo sinal	BJA5080_RS22490	BJS_RS24475	-	-	-	-	-	-
peptíeo sinal	BJA5080_RS22700	BJS_RS24680	-	-	-	-	-	-
peptíeo sinal	BJA5080_RS28000	-	-	-	-	-	-	-
peptíeo sinal	BJA5080_RS01030	-	-	-	blI6987	-	-	-
peptíeo sinal	BJA5080_RS01215	-	-	-	blr4262	-	-	-
peptíeo sinal	BJA5080_RS01045	-	-	-	-	-	-	-
peptíeo sinal	BJA5080_RS31970	-	-	-	-	-	-	-
peptíeo sinal	BJA5080_RS13775	-	-	-	-	-	-	-

Tabela 8 – Continuação

Proteína	Estirpes de <i>Bradyrhizobium</i>		PHI-base		T: 21 DAI	P: 21 DAI	T: 28 DAI	VL
	CPAC 7	CPAC 15	ID	Fenótipo	ID	ID	ID	ID
peptídeo sinal	-	BJS_RS12990	-	-	-	-	-	-
peptídeo sinal	-	BJS_RS05380	-	-	-	-	-	-
peptídeo sinal	-	BJS_RS20445	-	-	-	-	-	-
peptídeo sinal	-	BJS_RS05405	-	-	-	-	-	-
peptídeo sinal	-	BJS_RS02320	-	-	-	-	-	-
peptídeo sinal	-	BJS_RS18195	-	-	-	-	-	-
peptídeo sinal	-	BJS_RS24970	-	-	-	-	-	-
peptídeo sinal	-	BJS_RS25760	-	-	-	-	-	-
peptídeo sinal	-	BJS_RS22015	-	-	-	-	-	-
peptídeo sinal	BJA5080_RS23810	BJS_RS26490	-	-	-	-	-	-
peptídeo sinal	BJA5080_RS23830	BJS_RS26510	-	-	blI6486	-	-	-
peptídeo sinal	BJA5080_RS24305	BJS_RS26970	-	-	blI6577	blI6577	-	-
Outros								
nitrogenase proteína alfa molibdênio-ferro	-	BJS_RS31000	-	-	-	-	-	-
domínio com homologia NodB	BJA5080_RS08765	-	-	-	blI3646	-	-	-
proteína nodulação	BJA5080_RS00595, BJA5080_RS35985	BJS_RS44125, BJS_RS43425	-	-	-	-	-	-
domínio alfa-helicoidal	BJA5080_RS19065	BJS_RS05825	-	-	-	blr5613	-	-
proteína relacionada com Aerotolerância BATD	-	BJS_RS11350	-	-	-	-	-	-
domínio AGR-C-984p	BJA5080_RS14310	BJS_RS09795	-	-	blr4709	-	-	-
proteína de resistência a antibiótico	BJA5080_RS06450	-	-	-	-	-	-	-
proteína de resistência a antibiótico	-	BJS_RS04795	-	-	-	-	-	-
Domínio apple	BJA5080_RS33655	BJS_RS41055	-	-	-	-	-	-
proteína de processamento	-	BJS_RS22095	-	-	-	-	-	-
proteína ATC1	BJA5080_RS27910	BJS_RS39830	-	-	-	-	-	-
domínio C1q	BJA5080_RS13035	-	-	-	-	-	-	-
domínio CHAD	BJA5080_RS22070	BJS_RS24140	-	-	-	-	-	-
domínio CHAD	-	BJS_RS06090	-	-	-	-	-	-
efluxo de metal pesado	BJA5080_RS15515	BJS_RS18235	-	-	-	-	-	-
cupin	BJA5080_RS33490	-	-	-	-	blI1134	-	-
cupin 2	-	BJS_RS38870	-	-	-	-	-	-
cupin 2	BJA5080_RS38090	BJS_RS30900	-	-	-	-	-	-
cupin 2	BJA5080_RS12460	BJS_RS12210	-	-	blr4383	-	-	-
cupin 2	BJA5080_RS19710	BJS_RS04110	-	-	-	-	-	-
cupin 2	-	BJS_RS01455	-	-	blr2134	-	blr2134	-
ciclofinina	BJA5080_RS22545	BJS_RS24530	-	-	blI6165	blI6165	-	-
proteína citoplasmática	BJA5080_RS34410	BJS_RS33455	-	-	-	blI1313	-	-
proteína citosólica	BJA5080_RS11860	BJS_RS12780	-	-	-	-	-	-
Domínio alfa-beta dimérico	BJA5080_RS29350	BJS_RS08910	-	-	bsI0348	-	-	-
Domínio alfa-beta dimérico	BJA5080_RS24110	-	-	-	blr6541	-	-	-
Domínio alfa-beta dimérico	-	BJS_RS10715	-	-	-	-	-	-
Domínio alfa-beta dimérico	BJA5080_RS02450	BJS_RS03120	-	-	bsI1363	blI2446	-	-
flavoproteína	BJA5080_RS34675	BJS_RS33165	-	-	bsI1363	bsI1363	-	-

Tabela 8 – Continuação

Proteína	Estirpes de <i>Bradyrhizobium</i>		PHI-base		T: 21 DAI	P: 21 DAI	T: 28 DAI	VL
	CPAC 7	CPAC 15	ID	Fenótipo	ID	ID	ID	ID
Domínio DsrE	BJA5080_RS23405	BJS_RS16615	-	-	-	-	-	-
Domínio DR1885	BJA5080_RS36595	-	-	-	blr7088	-	-	-
Domínio DR1885	BJA5080_RS15235	BJS_RS08865	-	-	-	-	-	-
Domínio DsrEFH-like	BJA5080_RS28205	-	-	-	-	-	-	-
proteína EthD	-	BJS_RS33660	-	-	-	blr1249	-	-
domínio FECR	BJA5080_RS02275	BJS_RS02960	-	-	blI2411	-	-	-
domínio ferritina	BJA5080_RS22585	BJS_RS24580	-	-	-	-	-	-
resistência a ácido fusárico	-	BJS_RS24770	-	-	-	-	-	-
proteína rica em glicina	BJA5080_RS30070	BJS_RS37920	-	-	-	-	-	-
resistência a bleomicina	-	BJS_RS40790	-	-	-	-	-	-
resistência a bleomicina	-	BJS_RS16055	-	-	-	-	-	-
resistência a bleomicina	-	BJS_RS08805	-	-	-	-	-	-
domínio GYD	BJA5080_RS36120	-	-	-	-	-	-	-
domínio GYD	BJA5080_RS28230	-	-	-	-	-	-	-
Domínio HAD	BJA5080_RS08990	-	-	-	-	-	-	-
família haeI	BJA5080_RS36900	BJS_RS29900	-	-	blI7150	-	-	-
proteína de calor	BJA5080_RS06590	BJS_RS17635	-	-	-	-	-	-
domínio HotDog	-	BJS_RS20050	-	-	-	-	-	-
domínio HTH	BJA5080_RS27375	-	-	-	-	-	-	-
cluster de hidreto	BJA5080_RS36580	BJS_RS29590	-	-	-	-	-	-
proteína induzida por hipoxia	BJA5080_RS34730	BJS_RS33075	-	-	-	-	-	-
precursor da proteína imunogênica	BJA5080_RS01720	-	-	-	-	-	-	-
domínioKAP	BJA5080_RS33320	-	-	-	-	-	-	-
transpeptidase	BJA5080_RS30750	-	-	-	blr0621	-	-	-
Domínio L,D-transpeptidase	BJA5080_RS24325	BJS_RS26990	-	-	-	-	-	-
Domínio L,D-transpeptidase	BJA5080_RS19765	-	-	-	-	-	-	-
Domínio L,D-transpeptidase	-	BJS_RS40715	-	-	-	-	-	-
Domínio L,D-transpeptidase	BJA5080_RS03380	BJS_RS20940	-	-	-	blr4219	-	-
Domínio L,D-transpeptidase	BJA5080_RS11340	BJS_RS13285	-	-	blI4150	-	-	-
Domínio L,D-transpeptidase	BJA5080_RS11720	BJS_RS12935	-	-	-	blr4219	-	-
Domínio L,D-transpeptidase	BJA5080_RS30520	BJS_RS37485	-	-	blr0574	blr0574	-	-
Domínio L,D-transpeptidase	BJA5080_RS32705	BJS_RS35125	-	-	-	blI1024	-	-
Domínio L,D-transpeptidase	BJA5080_RS35250	BJS_RS42750	-	-	blr1468	-	-	-
Domínio L,D-transpeptidase	BJA5080_RS37330	BJS_RS29335	-	-	blI7229	-	-	-
Domínio L,D-transpeptidase	BJA5080_RS37345	BJS_RS30340	-	-	blr7232	blr7232	-	-
Domínio L,D-transpeptidase	-	BJS_RS41125	-	-	blr0192	-	-	-
proteína LABA	BJA5080_RS16170	BJS_RS08015	-	-	-	blr5067	-	blr5067
proteína LABA	BJA5080_RS33250	-	-	-	-	blr5067	-	blr5067
Lamina	-	BJS_RS11195	-	-	-	-	-	-
proteína lipo	BJA5080_RS31190	-	-	-	-	-	-	-
enzima lipolítica	-	BJS_RS34720	-	-	-	-	-	-
Domínio LTXXQ	BJA5080_RS24200	BJS_RS26840	-	-	-	-	-	-
Domínio LTXXQ	BJA5080_RS29410	BJS_RS38605	-	-	-	-	-	-

Tabela 8 – Continuação

Proteína	Estirpes de <i>Bradyrhizobium</i>		PHI-base		T: 21 DAI	P: 21 DAI	T: 28 DAI	VL
	CPAC 7	CPAC 15	ID	Fenótipo	ID	ID	ID	ID
Domínio LTXXQ	BJA5080_RS40770	-	-	-	-	-	-	-
família MmgE / PRPD	-	BJS_RS16880	-	-	-	-	-	-
família MmgE / PRPD	-	BJS_RS16870	-	-	-	-	-	-
proteína do domínio nacht	-	BJS_RS04235	-	-	-	-	-	-
repetição NHL	BJA5080_RS23435	-	-	-	-	-	-	-
repetição NHL	BJA5080_RS10885	BJS_RS13810	-	-	-	-	-	-
Família nipsnap	BJA5080_RS17545	BJS_RS07015	-	-	-	-	-	-
Família nipsnap	BJA5080_RS14770	BJS_RS09295	-	-	blI4801	-	-	-
Família nipsnap	BJA5080_RS08890	-	-	-	-	-	-	-
Família nipsnap	BJA5080_RS08820	-	-	-	-	-	-	-
fator gle1	BJA5080_RS15230	BJS_RS08875	-	-	-	-	-	-
domínio PAS	BJA5080_RS38360	BJS_RS31130	-	-	-	-	-	-
fosfotransferase fosfatase	BJA5080_RS38915	BJS_RS31735	-	-	-	-	-	-
proteína particionamento plasmídeo	BJA5080_RS30975	-	-	-	-	-	-	-
proteína particionamento plasmídeo	BJA5080_RS33245	-	-	-	-	-	-	-
glioxilato	BJA5080_RS28840	-	-	-	blr0253	-	-	-
Superfamília Holin-X	BJA5080_RS22045	BJS_RS24120	-	-	-	-	-	-
Superfamília Holin-X	-	BJS_RS05135	-	-	-	-	-	-
Superfamília Holin-X	-	BJS_RS04545	-	-	-	-	-	-
domínio ferro-quelante	BJA5080_RS06095	BJS_RS18155	-	-	-	-	-	-
domínio RES	-	BJS_RS28860	-	-	-	-	-	-
rodanase	-	BJS_RS24685	-	-	-	-	-	-
família Nif11 - parcial	-	BJS_RS00450	-	-	-	-	-	-
Domínio RLPA	BJA5080_RS05540	BJS_RS33695	-	-	-	-	-	-
domínio sialidases	-	BJS_RS14120	-	-	-	-	-	-
domínio SIR2	BJA5080_RS35820	-	-	-	-	-	-	-
sir2-como proteína de domínio	BJA5080_RS18445	-	-	-	-	-	-	-
pequena proteína com domínio helicoidal	BJA5080_RS36705	BJS_RS29690	-	-	-	-	-	-
Domínio SMI1 / KNR4	BJA5080_RS11615	-	-	-	-	-	-	-
Domínio SMI1 / KNR4	-	BJS_RS36100	-	-	-	-	-	-
Repetição SPW	BJA5080_RS14355	BJS_RS09765	-	-	blI4718	-	-	-
Repetição SPW	BJA5080_RS14950	BJS_RS09150	-	-	blI4833	-	-	-
domínio ppGpp	BJA5080_RS34110	-	-	-	-	-	-	-
proteína B	BJA5080_RS23990	BJS_RS26660	-	-	-	-	-	-
proteína de oxidação de enxofre	BJA5080_RS03690	-	-	-	-	-	-	-
proteína DsrE	BJA5080_RS25180	BJS_RS27860	-	-	blr6749	-	-	-
proteína DsrE	BJA5080_RS08190	-	-	-	-	-	-	-
Domínio SurA	-	BJS_RS23125	-	-	-	-	-	-
domínio rico em triptofano	BJA5080_RS21735	-	-	-	-	-	-	-
Domínio TusA	-	BJS_RS02440	-	-	-	-	-	-
Domínio com dupla hélice	BJA5080_RS39325	BJS_RS32130	-	-	-	-	-	-
fator de von Willebrand	BJA5080_RS25210	BJS_RS27890	-	-	blI6755	blI6755	-	-
fator de von Willebrand	BJA5080_RS03480	BJS_RS20845	-	-	-	-	-	-

Tabela 8 – Continuação

Proteína	Estirpes de <i>Bradyrhizobium</i>		PHI-base		T: 21 DAI	P: 21 DAI	T: 28 DAI	VL
	CPAC 7	CPAC 15	ID	Fenótipo	ID	ID	ID	ID
fator de von Willebrand	BJA5080_RS16625	BJS_RS07545	-	-	-	-	-	-
fator de von Willebrand	BJA5080_RS22410	-	-	-	-	-	-	-
família wd-repeat	BJA5080_RS39630	-	-	-	blI7767	blI7767	-	-
domínio XapX	BJA5080_RS01195	-	-	-	-	-	-	-
domínio XdhC	BJA5080_RS20045	-	PHI:2159 M. oryzae	Red. da virul.	-	-	-	-
YcaO-like de domínio	BJA5080_RS13655	BJS_RS10440	-	-	-	-	-	-
Domínio YCII	BJA5080_RS28020	-	-	-	-	-	-	-
Domínio YCII	BJA5080_RS03895	BJS_RS17095	-	-	-	-	-	-
Domínio YCII	BJA5080_RS36965	BJS_RS29960	-	-	-	-	-	-
Domínio YCII	BJA5080_RS24480	BJS_RS27150	-	-	-	-	-	-
Domínio YidB	BJA5080_RS28275	BJS_RS39425	-	-	-	-	-	-
Domínio YidB	-	BJS_RS04560	-	-	-	-	-	-
família yvtn	BJA5080_RS14295	BJS_RS09810	-	-	blI4706	-	-	-
domínio CGNR	BJA5080_RS24295	-	-	-	-	-	-	-
domínio PHD-tipo	BJA5080_RS34100	-	-	-	-	-	-	-
domínio PHD-tipo	BJA5080_RS18945	BJS_RS05975	-	-	bsr5590	-	-	-
proteína zinc-finger	BJA5080_RS16390	BJS_RS07810	-	-	blr5113	-	-	-
DomínioZinc-ribbon	-	BJS_RS34185	-	-	-	-	-	-
DomínioZinc-ribbon	BJA5080_RS38525	BJS_RS31295	-	-	-	-	-	-
domínio PepSY	BJA5080_RS19370	-	-	-	-	-	-	-
domínio PepSY	-	BJS_RS15765	-	-	-	-	-	-
domínio PepSY	-	BJS_RS15770	-	-	-	-	-	-
proteína de degradação de fenol	BJA5080_RS03870	-	-	-	blr2668	-	-	-
proteína de degradação do ácido fenilacético	-	BJS_RS13680	-	-	-	-	-	-

T: transcrito, P: proteoma, VL: vida-livre,.

**Análise gráfica do resultado da máxima entropia aplicada a um genoma de *Xanthomonas*,
utilizando diferentes tamanhos de regiões genômicas (2, 5, 10 e 50kb)**

Porcentagem de cobertura das regiões genômicas apontadas pela máxima entropia de CPAC 7 no genoma de CPAC 15

Porcentagem de cobertura das regiões genômicas apontadas pela máxima entropia de CPAC 15 no genoma de CPAC 7

**Regiões genômicas de *B. diazoefficiens* CPAC 7 e de *B. japonicum* CPAC 15 apontadas
pela máxima entropia a serem exploradas**

<i>Bradyrhizobium diazoefficiens</i> CPAC 7				<i>Bradyrhizobium japonicum</i> CPAC 15		
Regiões	Contig	Coordenadas	Referência	Regiões	Coordenadas	Referência
49 a 51	Cont26	1200001..1275000	Siqueira et al. 2014	4 a 8	75001..200000	Siqueira et al. 2014
57	Cont26	1400001..1425000	Siqueira et al. 2014	14	325001..350000	Siqueira et al. 2014
71 a 73	Cont26	1750001..1825000	Siqueira et al. 2014	16	375001..400000	Siqueira et al. 2014
86 e 87	Cont23	4737..54736	Siqueira et al. 2014	37 a 51	900001..1275000	Siqueira et al. 2014
91	Cont23	129737..154736	Presente estudo	67	1650001..1675000	Presente estudo
108 a 111	Cont23	554737..654736	Siqueira et al. 2014	91 a 93	2250001..2275000	Siqueira et al. 2014
146	Cont23	1504737..1529736	Presente estudo	98 a 100	2425001..2500000	Siqueira et al. 2014
150	Cont23	1604737..1629736	Presente estudo	134	3325001..3350000	Siqueira et al. 2014
152	Cont23	1654737..1679736	Presente estudo	141 e 142	3500001..3550000	Siqueira et al. 2014
155 e 156	Cont23	1729737..1779736	Presente estudo	177	4400001..4425000	Siqueira et al. 2014
158	Cont23	1804737..1829736	Presente estudo	179 e 180	4450001..4500000	Siqueira et al. 2014
189	Cont23	2579737..2604736	Siqueira et al. 2014	185 a 188	4600001..4700000	Siqueira et al. 2014
197	Cont23	2779737..2804736	Presente estudo	205 a 209	5100001..5225000	Siqueira et al. 2014
223 a 225	Cont23	3429737..3504736	Siqueira et al. 2014	218 a 220	5425001..5500000	Siqueira et al. 2014
234	Cont23	3704737..3729736	Presente estudo	225 a 228	5550001..5700000	Presente estudo
236 e 237	Cont23	3754737..3804736	Siqueira et al. 2014	251 a 253	6250001..6325000	Siqueira et al. 2014
244 e 245	Cont23	3954737..4004736	Presente estudo	255	6350001..6375000	Siqueira et al. 2014
246 a 248	Cont23	4004737..4075193	Siqueira et al. 2014	263	6550001..6575000	Presente estudo
	Cont24	1..4543		283	7050001..7075000	Presente estudo
263 e 264	Cont24	354544..404543	Siqueira et al. 2014	285	7100001..7125000	Presente estudo
265 e 266	Cont24	404544..454543	Presente estudo	295	7350001..7375000	Siqueira et al. 2014
277 a 280	Cont24	704544..804543	Siqueira et al. 2014	336	8375001..8400000	Presente estudo
314 e 315	Cont24	1629544..1679543	Presente estudo	348 e 349	8675001..8725000	Siqueira et al. 2014
318	Cont27	32586..57585	Presente estudo	352	8775001..8800000	Siqueira et al. 2014
326	Cont27	232586..257585	Siqueira et al. 2014	354	8825001..8850000	Siqueira et al. 2014
331	Cont27	357586..382585	Presente estudo	374 e 375	9325001..9375000	Siqueira et al. 2014
334 e 335	Cont27	432586..482585	Siqueira et al. 2014	377 a 379	9400001..9475000	Siqueira et al. 2014
358	Cont27	1032586..1057585	Siqueira et al. 2014	384	9575001..9583027	Siqueira et al. 2014

**Proteínas exclusivas de cada estirpe brasileira de *Bradyrhizobium* extraídas das regiões
apontadas pela máxima entropia**

Anotação	ID	Coordenada	Contig
<i>Bradyrhizobium diazoefficiens</i> CPAC 7			
regulador transcricional, família LysR	BJA5080_05086	1742588..1743526	Cont23
regulador transcricional, família LysR	BJA5080_01345	133782..134807	Cont23
(2Fe-2S) ferredoxina	BJA5080_06601	3459784..3461328	Cont23
2-pirona-4,6-dicarboxilato de hidrolase	BJA5080_01343	135895..136755	Cont23
aldolase ácido 2,4-di-hidroxi-2-eno-1,7-diólico	BJA5080_05082	1748278..1748979	Cont23
3-oxoacil-ACP redutase	BJA5080_05083	1747458..1748261	Cont23
ABC transportador ATP de ligação	BJA5080_05089	1739666..1740436	Cont23
ABC transportador ATP de ligação	BJA5080_05088	1740429..1741169	Cont23
ABC transportador permease	BJA5080_05091	1737767..1738639	Cont23
ABC transportador permease	BJA5080_05090	1738656..1739669	Cont23
ABC transportador permease	BJA5080_05077	1753136..1753939	Cont23
ABC transportador	BJA5080_00015	1654204..1655271	Cont23
ABC transportador	BJA5080_05092	1736570..1737754	Cont23
ABC transportador	BJA5080_05076	1753968..1755032	Cont23
alofanato hidrolase	BJA5080_05074	1756253..1756972	Cont23
alofanato hidrolase	BJA5080_05073	1756969..1757946	Cont23
alofanato hidrolase	BJA5080_05069	1760132..1760794	Cont23
alfa/beta hidrolase	BJA5080_01447	23228..24103	Cont23
alfa/beta hidrolase	BJA5080_00142	1526961..1527848	Cont23
amidohidrolase	BJA5080_01431	42417..44303	Cont23
amidohidrolase	BJA5080_01346	133127..133732	Cont23
biossíntese de antibiótico	BJA5080_00153	1514950..1515606	Cont23
domínio beta-lactamase	BJA5080_00931	612118..613122	Cont23
biotina	BJA5080_05071	1758742..1759167	Cont23
cupin	BJA5080_06603	3458307..3458840	Cont23
D-ribose transportador	BJA5080_01349	128940..130466	Cont23
dioxigenase	BJA5080_01437	36625..37548	Cont23
Proteína flagelar FlbY	BJA5080_05006	1829159..1829632	Cont23
proteína flagelar FlgK	BJA5080_05005	1829654..1831402	Cont23
glucosil transferase	BJA5080_05060	1769130..1770971	Cont23
glucosil transferase	BJA5080_05059	1770968..1772716	Cont23
glioxalase	BJA5080_01429	44784..45701	Cont23
acetiltransferase família GNAT	BJA5080_00002	1670178..1670675	Cont23
família GntR regulador transcricional	BJA5080_00011	1659774..1660496	Cont23
proteína de secreção da hemolisina D	BJA5080_06336	3716211..3717389	Cont23
HNH nuclease	BJA5080_00957	560027..560650	Cont23
hidrolase	BJA5080_07787	4049823..4051163	Cont23
Ig-like de domínio SoxY	BJA5080_00935	608466..608717	Cont23
ligase	BJA5080_05053	1776994..1778349	Cont23
transglicosilase lítica MltB	BJA5080_06599	3461805..3462791	Cont23
proteína de membrana	BJA5080_01427	47106..47813	Cont23
MFS transportador	BJA5080_00161	1505641..1507281	Cont23
MFS transportador	BJA5080_06593	3469949..3470926	Cont23
N, N-dimetilformamidase	BJA5080_00013	1655662..1657923	Cont23
nitrilase	BJA5080_00160	1507342..1508307	Cont23
O-metiltransferase	BJA5080_07806	4025108..4025773	Cont23

Tabela 10 – Continuação

Anotação	ID	Coordenada	Contig
peptidase M23	BJA5080_07876	3964329..3965225	Cont23
desidrogenase fosfogluconato	BJA5080_01340	139318..140232	Cont23
Fitoquelatina sintase	BJA5080_07810	4019553..4019903	Cont23
porina	BJA5080_01347	131698..133005	Cont23
putativa 2-pirona-4,6-dicarbaxilato hidrolase	BJA5080_01338	141586..142479	Cont23
Putativo ABC transportador	BJA5080_01455	15023..15718	Cont23
putativo ABC transportador permease	BJA5080_05078	1752204..1753136	Cont23
acetiltransferase	BJA5080_05058	1772860..1773438	Cont23
Adenilato-ciclase	BJA5080_01331	149274..150857	Cont23
adenilato ciclase 3	BJA5080_01463	5256..7208	Cont23
adenilato ciclase	BJA5080_01426	48074..49333	Cont23
agropina reductase	BJA5080_05081	1749057..1749749	Cont23
alquil hidroperóxido redutase	BJA5080_01424	49716..49931	Cont23
arilesterase	BJA5080_01434	40323..41342	Cont23
proteína modificação rod	BJA5080_05009	1826932..1827600	Cont23
acetil-CoA carboxilase	BJA5080_05067	1762146..1762592	Cont23
transporte de cromato	BJA5080_07879	3961916..3963103	Cont23
citocromo B561	BJA5080_00007	1666164..1666721	Cont23
citocromo B561	BJA5080_07855	3982055..3982612	Cont23
citocromo c	BJA5080_00050	1621185..1621916	Cont23
D-alanina-ligase	BJA5080_00152	1516033..1516725	Cont23
di-heme citocromo C	BJA5080_00044	1624723..1625544	Cont23
Mono-oxigenase dimetilnilina	BJA5080_06608	3453763..3455250	Cont23
DNA metilase domínio N-4/N-6	BJA5080_00925	619077..620432	Cont23
DNA polimerase	BJA5080_06558	3494863..3495198	Cont23
Putativo domínio EcoEI R	BJA5080_06280	3797996..3799954	Cont23
Epóxido hidratase	BJA5080_01438	35459..36331	Cont23
proteína flagelar	BJA5080_05008	1827620..1828885	Cont23
proteína flagelar 2	BJA5080_05012	1824141..1825829	Cont23
proteína flagelar Flis	BJA5080_05011	1825848..1826270	Cont23
Flagelina	BJA5080_05013	1823242..1824066	Cont23
putativo glutamato-1-semialdeído 2,1-aminomutase	BJA5080_00010	1660731..1661957	Cont23
helicase	BJA5080_06287	3783357..3786590	Cont23
Putativa proteína de utilização de Hidantoina	BJA5080_05084	1745768..1747450	Cont23
Putativa proteína de utilização de Hidantoina	BJA5080_05085	1743689..1745755	Cont23
Putativa hidrolase	BJA5080_01435	39677..40309	Cont23
Putativa hidrolase	BJA5080_01433	41366..42103	Cont23
putativo hydroxineurosporeno-O-metiltransferase	BJA5080_06615	3447150..3447773	Cont23
putativo L-alo-treonina desidrogenase	BJA5080_05080	1749987..1750790	Cont23
putativa proteína específica-efluxo macrolido	BJA5080_01456	13824..14996	Cont23
MFS transportador	BJA5080_07887	3954834..3955715	Cont23
metalofosfoesterase	BJA5080_00933	610985..611737	Cont23
Putativa metiltransferase	BJA5080_00061	1611133..1611780	Cont23
putativo molibdato de ligação proteína periplasmática	BJA5080_01420	53071..53835	Cont23
putativo N-6 DNA metilase	BJA5080_06281	3796530..3797999	Cont23
putativo nitrilotriacetato monooxigenase componente B	BJA5080_00159	1508327..1508932	Cont23
regulador de difosfato quinase	BJA5080_06613	3448938..3449399	Cont23
O-antígeno do sistema permease	BJA5080_05055	1775366..1776214	Cont23
O-metiltransferase, família 3	BJA5080_00922	623107..623769	Cont23
sensor de proteína PAS/PAC	BJA5080_07849	3985942..3986280	Cont23
histidina quinase	BJA5080_07858	3979218..3980636	Cont23
permease	BJA5080_01454	15815..16975	Cont23
proteína de montagem pilus, família RCPC/CPAB	BJA5080_00066	1605765..1606721	Cont23
proteína de montagem pilus, família TADB	BJA5080_00063	1609752..1610225	Cont23
proteína de montagem pilus, família TadC	BJA5080_00062	1610279..1611130	Cont23
proteína de montagem pilus, família Tadd	BJA5080_00055	1616729..1617145	Cont23
proteína de montagem pilus, família TADG	BJA5080_00059	1612314..1613642	Cont23
proteína de montagem pilus, família Tadz / CPAE	BJA5080_00065	1606721..1607923	Cont23
proteína biossíntese polissacarídeo	BJA5080_05057	1773435..1774577	Cont23
prepilin peptidase, família TadV/CpaA	BJA5080_00067	1605257..1605763	Cont23
Prolina iminopeptidase	BJA5080_04291	2586214..2587119	Cont23

Tabela 10 – Continuação

Anotação	ID	Coordenada	Contig
Protease degQ Precursor	BJA5080_06562	3492027..3493007	Cont23
proteína pseudopilin, família Tade	BJA5080_00058	1613639..1614055	Cont23
transportador	BJA5080_05075	1755044..1756138	Cont23
pirofosforilase	BJA5080_01419	54059..54988	Cont23
proteína de restrição	BJA5080_06282	3794779..3796533	Cont23
RNA polimerase fator sigma-E	BJA5080_05014	1822403..1823101	Cont23
S-adenosilmetionina: 2-metiltransferase demetilmenaquinona	BJA5080_05064	1764668..1765339	Cont23
secretina, família RcpA / CpaC	BJA5080_00057	1614456..1616120	Cont23
BAES sensor	BJA5080_00006	1666858..1667958	Cont23
desidrogenase	BJA5080_05066	1762640..1763230	Cont23
peptídeo sinal	BJA5080_00934	610125..610382	Cont23
histidina quinase	BJA5080_07866	3973060..3974232	Cont23
histidina quinase	BJA5080_08390	4069714..4070856	Cont23
precursor SoxD	BJA5080_00041	1628573..1629217	Cont23
Proteína soxs	BJA5080_00051	1620738..1621109	Cont23
proteína soxT	BJA5080_00053	1619198..1620277	Cont23
proteína SOXX	BJA5080_00047	1623436..1623879	Cont23
proteína Soxy	BJA5080_00046	1623898..1624350	Cont23
proteína soxZ	BJA5080_00045	1624368..1624694	Cont23
tiorredoxina Sox	BJA5080_00048	1622692..1623306	Cont23
regulador de transcrição	BJA5080_00924	621011..621949	Cont23
regulador de transcrição	BJA5080_01334	144973..145356	Cont23
proteína reguladora da transcrição	BJA5080_05065	1763600..1764571	Cont23
proteína reguladora da transcrição	BJA5080_07805	4026902..4027888	Cont23
SoxR proteína reguladora da transcrição	BJA5080_00052	1620350..1620700	Cont23
proteína de transporte transmembranar	BJA5080_01440	32643..34301	Cont23
sensor híbrido dois-componentes e regulador	BJA5080_01332	146818..149277	Cont23
sensor híbrido dois-componentes e regulador	BJA5080_06606	3456806..3457096	Cont23
sensor híbrido dois-componentes e regulador	BJA5080_00005	1667955..1668641	Cont23
sistema de secreção tipo I, ATP cassette de ligação	BJA5080_06337	3713894..3716149	Cont23
Sistema de secreção tipo I	BJA5080_06338	3712029..3713702	Cont23
Exoproteína -sistema de secreção I	BJA5080_08474	1671030..1679222	Cont23
Exoproteína -sistema de secreção I	BJA5080_08475	3718722..3728780	Cont23
Zn-dependente hidrolase	BJA5080_01439	34377..35393	Cont23
nuclease família RecB	BJA5080_06284	3789654..3793052	Cont23
partição plasmídeo RepB	BJA5080_06288	3782047..3782310	Cont23
Ribonuclease HII/HIII	BJA5080_07793	4042845..4043126	Cont23
Domínio RmlC	BJA5080_07840	3993827..3994684	Cont23
sinal de reconhecimento RNA	BJA5080_04274	2599249..2599611	Cont23
Schlafen, domínio AAA	BJA5080_00952	570312..571277	Cont23
desidrogenase/reductase SDR	BJA5080_07816	4012727..4012966	Cont23
desidrogenase/reductase SDR	BJA5080_07815	4012972..4013271	Cont23
ABC permease, transportador de açúcar	BJA5080_01348	130463..131521	Cont23
Tetratrico-peptídeo	BJA5080_07780	4057490..4058650	Cont23
fator de transcrição – Homeodomínio-like	BJA5080_00911	639116..639502	Cont23
regulador transcricional	BJA5080_01436	38732..39658	Cont23
regulador transcricional	BJA5080_01344	134970..135722	Cont23
regulador transcricional	BJA5080_00162	1504595..1505473	Cont23
transglutaminase	BJA5080_06335	3717606..3718661	Cont23
xilose isomerase	BJA5080_05051	1779652..1780482	Cont23
ATPase	BJA5080_07560	1645208..1647316	Cont24
proteína SMC, segregação de cromossomo	BJA5080_07565	1638900..1641353	Cont24
DTDP-4-dehidroramnose 3,5-epimerase	BJA5080_08374	388952..389167	Cont24
Heparinase II/III	BJA5080_03394	739936..741840	Cont24
domínio HTH	BJA5080_07539	1676886..1677710	Cont24
hidrogenase-4 componente E	BJA5080_03445	796500..797168	Cont24
Ácido ceto-reductoisomerase	BJA5080_03439	790683..791303	Cont24
Biossíntese de lipopolissacarídeo	BJA5080_03373	705154..706239	Cont24
protease	BJA5080_03111	427399..428922	Cont24
NADP	BJA5080_03060	371406..371621	Cont24
Nuclease SbcCD subunidade D	BJA5080_08362	1649578..1649925	Cont24

Tabela 10 – Continuação

Anotação	ID	Coordenada	Contig
Peptidase C39, processamento de bacteriocina	BJA5080_03440	791332..792588	Cont24
Polissacarídeo liase	BJA5080_03396	744087..746027	Cont24
2-oxoisovalerato desidrogenase	BJA5080_03435	785498..786730	Cont24
2-oxoisovalerato desidrogenase	BJA5080_03436	786734..787750	Cont24
3-hidroxi-isobutirato desidrogenase	BJA5080_03426	775744..776655	Cont24
3-isopropilmalato desidrogenase	BJA5080_03424	773004..774083	Cont24
ABC transportador permease	BJA5080_03431	782525..783523	Cont24
ABC transportador permease	BJA5080_03430	781473..782486	Cont24
acetoacetil-coenzima A sintetase	BJA5080_03393	734899..738213	Cont24
aciltransferase	BJA5080_08324	386207..387328	Cont24
aciltransferase	BJA5080_03058	369542..370597	Cont24
Adenilato-ciclase	BJA5080_07537	1678235..1679371	Cont24
N(3')-acetiltransferase IV	BJA5080_03386	724937..725848	Cont24
asparagina sintetase	BJA5080_03375	708328..710346	Cont24
asparagina sintetase	BJA5080_03388	728721..730661	Cont24
proteína de ligação	BJA5080_03376	710418..711359	Cont24
CDP-glicose 4,6-desidratase	BJA5080_03066	375368..376405	Cont24
cefalosporina hidroxilase	BJA5080_03077	387341..388099	Cont24
di-hidrolipoamida acetiltransferase	BJA5080_03437	787750..789033	Cont24
di-hidrolipoamida desidrogenase	BJA5080_03438	789047..790444	Cont24
exonuclease SbcD	BJA5080_08346	1647406..1647846	Cont24
produção de exopolissacarídeo	BJA5080_03402	754882..755940	Cont24
Formato hidrogenio liase, subunidade 4	BJA5080_03444	795544..796500	Cont24
Formato hidrogenio liase, subunidade 5	BJA5080_03448	799004..800146	Cont24
Formato hidrogenio liase, subunidade 7	BJA5080_03449	800160..800693	Cont24
GDP-manose 1-fosfato guanililtransferase	BJA5080_03059	370621..371022	Cont24
glucosil transferase	BJA5080_08322	727386..728552	Cont24
glucosil transferase 11	BJA5080_03064	373518..374393	Cont24
glucosil transferase	BJA5080_03087	398805..400010	Cont24
glucosil transferase	BJA5080_03062	372052..372963	Cont24
glucosil transferase	BJA5080_03070	379417..380979	Cont24
glucosil transferase	BJA5080_03071	381082..382086	Cont24
glucosil transferase	BJA5080_03074	384135..385223	Cont24
glucosil transferase 1	BJA5080_03387	726236..727384	Cont24
hidrogenase componente B	BJA5080_03443	793538..795553	Cont24
putativo Hidrogenase-4 componente F	BJA5080_03446	797165..798616	Cont24
proteína reguladora Leucina-responsiva	BJA5080_03433	784555..785055	Cont24
metalofosfoesterase	BJA5080_03109	425464..425907	Cont24
metalofosfoesterase	BJA5080_03114	429586..430470	Cont24
metiltransferase	BJA5080_03072	382079..383287	Cont24
metiltransferase	BJA5080_03399	750586..751371	Cont24
NAD-dependente epimerase/desidratase	BJA5080_03068	377862..378722	Cont24
NDP-açúcar oxidorreductase	BJA5080_03065	374440..375381	Cont24
Nuclease sbcCD subunidade C	BJA5080_07558	1649928..1653713	Cont24
nucleotidil transferase	BJA5080_03069	378719..379420	Cont24
O-antígeno ABC transportador, ligação de ATP	BJA5080_03080	389522..390853	Cont24
O-antígeno ABC transportador, permease	BJA5080_03081	390859..391713	Cont24
proteína imunogenica da membrana externa	BJA5080_03090	401828..402652	Cont24
oxidorreductase	BJA5080_03390	731632..733047	Cont24
biossíntese de polissacarídeo	BJA5080_03374	706448..708115	Cont24
biossíntese de polissacarídeo, CapD	BJA5080_03067	376612..377865	Cont24
oxidorreductase	BJA5080_03078	388096..388833	Cont24
biossíntese de polissacarídeo	BJA5080_03400	752730..754226	Cont24
proteína NusG	BJA5080_03082	391907..392422	Cont24
regulador de transcrição	BJA5080_07538	1677764..1678081	Cont24
proteína reguladora da transcrição	BJA5080_03107	422835..423197	Cont24
proteína reguladora da transcrição	BJA5080_03423	772097..772993	Cont24
UDP-glucose 4-epimerase	BJA5080_03404	756394..757302	Cont24
UDP-glucose 4-epimerase	BJA5080_03406	757698..758288	Cont24
UDP-glucose / proteína da família desidrogenase	BJA5080_03381	716711..717781	Cont24
UDP-N-acetilglucosamina-2-epimerase	BJA5080_03384	721126..723366	Cont24

Tabela 10 – Continuação

Anotação	ID	Coordenada	Contig
proteína CopG	BJA5080_03116	431829..432122	Cont24
S-adenosil-L-metionina-dependente metiltransferase	BJA5080_03379	714893..715531	Cont24
Terpenoide ciclase	BJA5080_03385	723710..724615	Cont24
regulador transcricional	BJA5080_07563	1642288..1642908	Cont24
domínio rico em triptofano	BJA5080_03084	393495..396551	Cont24
proteína de ligação	BJA5080_03083	392419..393063	Cont24
ABC transportador	BJA5080_02696	1813508..1815196	Cont26
Endopeptidase, domínio NLPC /P60	BJA5080_02671	1765761..1766204	Cont26
P-loop nucleosídeo trifosfato hidrolase	BJA5080_02195	1237350..1239482	Cont26
P-loop nucleosídeo trifosfato hidrolase	BJA5080_02204	1253020..1254165	Cont26
P-loop nucleosídeo trifosfato hidrolase	BJA5080_02207	1257181..1259523	Cont26
ABC transportador, ligação de ATP	BJA5080_02699	1817153..1818757	Cont26
ABC transportador, permease	BJA5080_02697	1815193..1816155	Cont26
ABC transportador, permease	BJA5080_02698	1816224..1817033	Cont26
Diguanilato ciclase	BJA5080_02660	1749404..1751026	Cont26
Reparo de DNA	BJA5080_02692	1805800..1808232	Cont26
domínio hélice-volta-hélice	BJA5080_02203	1252364..1252720	Cont26
repressor de manitol	BJA5080_02689	1803089..1803688	Cont26
proteína de utilização da metilamina	BJA5080_02661	1751415..1752437	Cont26
nitrilotriacetato monooxigenase componente A	BJA5080_02695	1812108..1813475	Cont26
fosfatase	BJA5080_02680	1792958..1793305	Cont26
regulador de transcrição	BJA5080_02351	1416390..1416659	Cont26
regulador de transcrição	BJA5080_02345	1408713..1409441	Cont26
proteína de secreção tipo I	BJA5080_02198	1243651..1247283	Cont26
transportador de aminoácidos	BJA5080_05483	365232..366665	Cont27
transportador de aminoácidos	BJA5080_05496	377979..379418	Cont27
arginina deaminase	BJA5080_05482	363943..365205	Cont27
proteína de secreção da hemolisina D	BJA5080_05176	43929..45005	Cont27
histidina quinase	BJA5080_05178	46126..46983	Cont27
proteína de membrana	BJA5080_05187	51541..52599	Cont27
domínio PepSY	BJA5080_05186	50416..51273	Cont27
família AcrB/ACRD/ACRF	BJA5080_05175	40786..43932	Cont27
ATPase (AAA + superfamília)	BJA5080_06101	1045761..1048916	Cont27
proteína de secreção HlyD	BJA5080_05177	45002..46129	Cont27
oxigenase	BJA5080_05574	471785..473338	Cont27
regulador de transcrição, a família TetR	BJA5080_05179	47196..47858	Cont27
putativa proteína associada a virulência E	BJA5080_06093	1034649..1035110	Cont27
regulador transcricional	BJA5080_05180	47877..48545	Cont27
sensor híbrido de dois componentes e regulador	BJA5080_05172	35824..38055	Cont27
regulador transcricional, família XRE	BJA5080_06098	1038136..1042194	Cont27
<i>Bradyrhizobium japonicum</i> CPAC 15			
proteína de nodulação NoeE	BJS_07601	328172..328786	-
dihidrorizobiotoxina sintase	BJS_07600	330515..332944	-
regulador transcricional, LuxR	BJS_00914	5699330..5699623	-
regulador transcricional dois-componente LuxR	BJS_07013	996857..997258	-
regulador transcricional, família LysR	BJS_00852	5625379..5626269	-
regulador transcricional, família LysR	BJS_00873	5654358..5655293	-
regulador transcricional, família LysR	BJS_06029	4414481..4415080	-
regulador transcricional, LysR	BJS_08763	981060..982115	-
regulador de resposta de dois-componentes	BJS_00008	4676809..4677030	-
lactonase	BJS_00019	4684779..4685723	-
carbamoiltransferase	BJS_00403	5101714..5103294	-
oxidorreductase	BJS_00404	5103709..5105733	-
UDP-glucose	BJS_00405	5105832..5107097	-
UDP-N-acetilglucosamina- 2-epimerase	BJS_00406	5107094..5108215	-
Heparinase II	BJS_00408	5109677..5111356	-
asparagina sintetase	BJS_00412	5115591..5117621	-
glucosiltransferase 1	BJS_00413	5117665..5119398	-
S-adenosil-L-metionina-dependente metiltransferase	BJS_00422	5128015..5129004	-
aciltransferase 3	BJS_00423	5129375..5131348	-
Sistema de secreção de protease tipo I	BJS_00424	5131458..5134082	-

Tabela 10 – Continuação

Anotação	ID	Coordenada	Contig
canal de cloreto	BJS_00425	5134639..5136561	-
UDP-glucose 4-epimerase	BJS_00428	5140581..5141513	-
Ubiquinona	BJS_00429	5141516..5142328	-
acetiltransferase	BJS_00430	5142357..5142962	-
asparagina sintetase	BJS_00432	5144100..5145938	-
endopeptidase	BJS_00435	5149078..5149461	-
Sistema de secreção tipo V	BJS_00439	5150880..5152511	-
proteína de membrana	BJS_00443	5155740..5157170	-
proteína de conjugação	BJS_00463	5177696..5178916	-
proteína de conjugação, trbG	BJS_00464	5178913..5179323	-
proteína de conjugação	BJS_00465	5179545..5180498	-
proteína de conjugação, trbL	BJS_00466	5180881..5181543	-
proteína de conjugação	BJS_00467	5181750..5182625	-
regulador transcricional, família CopG	BJS_00478	5199794..5200198	-
transglucosilase lítica	BJS_00480	5203215..5203880	-
replicação A	BJS_00483	5205377..5206165	-
regulador de transcrição	BJS_00488	5207924..5208154	-
HNS-dependente expressão A/B	BJS_00714	5446318..5446584	-
poliketide sintase	BJS_00720	5451917..5458816	-
fibronectina III	BJS_00721	5458861..5460462	-
NAD(P)	BJS_00724	5466568..5467845	-
Transportador de ion cromato	BJS_00726	5469729..5470703	-
NADP-dependente de oxidoreductase	BJS_00741	5489661..5490548	-
MFS transportador	BJS_00742	5490688..5492181	-
proteína de resistência a múltiplas drogas	BJS_00743	5492178..5493320	-
regulador transcricional, família TetR	BJS_00744	5493460..5494101	-
transportador de múltiplas drogas	BJS_00834	5602787..5603848	-
ABC transportDOR	BJS_00835	5603845..5606175	-
Isoprenilcisteine carboxil metiltransferase	BJS_00841	5612429..5612668	-
esterase	BJS_00843	5614656..5615156	-
aldeído desidrogenase	BJS_00846	5618001..5620214	-
adenilato ciclase	BJS_00849	5621518..5622675	-
oxidoreductase	BJS_00850	5622764..5624377	-
Proteína de Nodulação W	BJS_00853	5626633..5627268	-
histidina quinase	BJS_00854	5627878..5633763	-
alquil hidropéroxido reductase	BJS_00858	5639658..5640251	-
NADPH: quinona reductase	BJS_00860	5641454..5642335	-
nitro-reductase	BJS_00861	5642512..5643147	-
Alquil hidropoxidase	BJS_00862	5643176..5643727	-
metiltransferase	BJS_00864	5644488..5645432	-
cupin	BJS_00865	5645536..5645949	-
NADPH: quinona reductase	BJS_00867	5647565..5648494	-
arabinose permease – ABC transportador	BJS_00868	5649045..5650400	-
transportador	BJS_00869	5650567..5651907	-
citocromo P460	BJS_00871	5652789..5653364	-
alquil hidropéroxido reductase	BJS_00872	5653624..5654361	-
biossíntese de antibióticos, mono-oxigenase	BJS_00874	5655709..5656395	-
histidina quinase	BJS_00875	5656464..5657492	-
cupin	BJS_00877	5658055..5658420	-
proteína de ligação ao hospedeiro	BJS_00878	5658934..5659338	-
ABC transportador, permease	BJS_00881	5661316..5662176	-
ABC transportador, permease	BJS_00882	5662173..5663060	-
ABC transportador, substrato de ligação	BJS_00883	5663126..5664571	-
regulador transcricional, família LacI	BJS_00884	5664583..5665620	-
N-acetilglucosamina-6-fosfato-desacetilase	BJS_00886	5666529..5667677	-
desidrogenase	BJS_00887	5667688..5668734	-
desidrogenase	BJS_00888	5668742..5669731	-
ABC transportador, ATPase	BJS_00894	5675544..5677109	-
glutaminase	BJS_00895	5677312..5678310	-
MFS transportador	BJS_00896	5678384..5680033	-
transportador de citrato	BJS_00898	5681572..5683401	-

Tabela 10 – Continuação

Anotação	ID	Coordenada	Contig
transportador de citrato	BJS_00899	5683454..5685283	-
Na + / H + antiportador subunidade E	BJS_00902	5689075..5690343	-
citocromo c	BJS_01429	6252290..6252838	-
adenilato ciclase	BJS_01433	6254111..6256015	-
Sistema de secreção tipo V	BJS_01435	6256753..6259728	-
proteína de membrana	BJS_01439	6263383..6263913	-
transportador	BJS_01441	6264944..6266881	-
fosfotransferase I	BJS_01444	6268563..6270161	-
transporte de açúcar	BJS_01445	6270167..6270442	-
quinase	BJS_01446	6270474..6270875	-
PTS-dependente di-hidroxiacetona quinase	BJS_01447	6270872..6271453	-
ABC transportador, ATP de ligação	BJS_01448	6271483..6272490	-
histidina quinase, PAS/PAC	BJS_01454	6277577..6280045	-
MFS transportador	BJS_01455	6281661..6282557	-
Adenilato-ciclase	BJS_01457	6282951..6283661	-
proteínas osmoticamente induzidaC	BJS_01458	6284095..6284667	-
NADH: flavina oxidoredutase	BJS_01459	6284773..6285879	-
regulador transcricional, família HxIR	BJS_01460	6286033..6286410	-
Polifosfato quinase-2	BJS_01465	6291233..6291478	-
alfa-trealose-fosfato sintase	BJS_01472	6296200..6297654	-
proteína CsbD	BJS_01476	6300874..6301098	-
DNA ligase	BJS_01479	6303064..6303489	-
regulador de resposta	BJS_01483	6307710..6308108	-
ferro permease	BJS_01493	6314223..6314483	-
histidina quinase	BJS_01495	6315389..6316510	-
transportador	BJS_01496	6316574..6317716	-
histidina quinase	BJS_01505	6324772..6326169	-
sulfonatos alifáticos	BJS_01529	6352262..6352984	-
hidrolase	BJS_01532	6355970..6357274	-
sensor do sistema dois-componentes	BJS_01535	6359789..6361387	-
transglicosilase	BJS_01539	6368778..6369839	-
Tetratrico-peptídeo	BJS_01542	6371257..6372381	-
oxidase	BJS_01544	6373896..6375095	-
C4-dicarboxilato	BJS_01730	6565675..6567021	-
metalofofosfoesterase	BJS_01731	6567240..6568505	-
Tetratrico-peptídeo	BJS_01736	6570899..6572485	-
regulador transcricional, família LacI	BJS_02204	7050502..7051407	-
lipase	BJS_02206	7051538..7053004	-
ABC transportador, permease	BJS_02208	7053998..7055005	-
aldolase fosfato	BJS_02210	7056424..7057137	-
tioesterase	BJS_02212	7057416..7057877	-
desidrogenase	BJS_02255	7101986..7103071	-
proteína transmembranar TqsA	BJS_02263	1143088..1143990	-
Histidina quinase sensor de dois-componentes	BJS_02268	1146009..1147172	-
Policetídeo ciclase SnaoL	BJS_02269	1147460..1147792	-
histidina quinase	BJS_02272	1148765..1149655	-
biossíntese de glucano periplasmático	BJS_02279	1153613..1155073	-
glucosiltransferase H	BJS_02280	1155061..1157205	-
histidina quinase	BJS_02283	1158978..1160099	-
hidrolase	BJS_02284	1160099..1160830	-
álcool desidrogenase, GroES	BJS_02288	1165723..1166901	-
NADP	BJS_02289	1167040..1167783	-
proteína de ligação de pirofosfato de tiamina	BJS_02308	1185211..1186986	-
oxidoredutase	BJS_02309	1186989..1188014	-
UvrABC	BJS_02310	1188204..1190732	-
superfamília CheY	BJS_02311	1190999..1191226	-
proteína de ligação de leucina	BJS_02313	1193855..1194751	-
fosfodiesterase	BJS_02325	1205346..1205834	-
Histidina quinase sensor de dois-componentes	BJS_02339	1219049..1222195	-
putativa NADH desidrogenase	BJS_02341	1223228..1224091	-
OprB porina seletiva de carboidratos	BJS_02342	1224720..1224950	-

Tabela 10 – Continuação

Anotação	ID	Coordenada	Contig
permease, Fet4	BJS_02343	1226306..1226791	-
OprB porina seletiva de carboidratos	BJS_02344	1225064..1226104	-
OprB porina seletiva de carboidratos	BJS_02347	1229737..1230279	-
domínio relacionado a Ferritina	BJS_02349	1231409..1231864	-
piridoxamina 5'-fosfato oxidase	BJS_02361	1243538..1243969	-
regulador transcricional	BJS_02388	1270924..1271757	-
ABC transportador	BJS_02389	1271942..1273186	-
ABC transportador, permease	BJS_02390	1273287..1274207	-
ABC transportador, permease	BJS_02391	1274211..1275254	-
regulador de transcrição, família TetR	BJS_02784	1658206..1658739	-
proteína de membrana	BJS_02787	1661659..1662924	-
oxidorreductase	BJS_02788	1663258..1664106	-
GMC oxidorreductase	BJS_02793	1667120..1668898	-
citocromo-C oxidase	BJS_02794	1668981..1670141	-
Domínio RmlC	BJS_03352	2254782..2255081	-
ABC transportador	BJS_03353	2255295..2256005	-
DNA-citosina metiltransferase	BJS_03361	2271309..2272586	-
endonuclease de restrição, tipo II	BJS_03363	2273034..2274263	-
peptidase C13	BJS_03368	2280015..2280545	-
metiltransferase	BJS_03372	2284955..2285755	-
ABC transportador	BJS_03375	2287729..2288859	-
peptidase	BJS_03384	2299038..2299292	-
Família RecA	BJS_03395	2310259..2310579	-
proteína de ligação	BJS_03401	2315691..2317127	-
proteína de membrana	BJS_03496	2425083..2426081	-
proteína MoxR	BJS_03498	2426558..2427544	-
ATPase	BJS_03499	2427552..2428508	-
regulador transcricional AraC	BJS_03502	2431842..2432834	-
domínio PepSY	BJS_03507	2438568..2438804	-
transferase	BJS_03509	2444596..2446119	-
Glucosiltransferase	BJS_03515	2452569..2453243	-
Glucosiltransferase	BJS_03517	2454739..2455254	-
regulador transcricional	BJS_03531	2470402..2470626	-
DNA ligase	BJS_03537	2478732..2479334	-
Molibdênio sulfurase cofator	BJS_03550	2487163..2487522	-
proteína de membrana	BJS_03556	2495452..2496111	-
ABC transportador	BJS_04367	3327916..3329859	-
proteína de ligação de peptidoglicano	BJS_04368	3329886..3331622	-
proteína de resistência acriflavina	BJS_04373	3335971..3339117	-
proteína de resistência acriflavina	BJS_04375	3339640..3342717	-
regulador transcricional, família LacI	BJS_04611	7349957..7350991	-
6-carboxihexanoate - CoA-ligase	BJS_04612	7351014..7353101	-
enoil-CoA hidratase	BJS_04613	7353105..7353896	-
regulador transcricional, família IclR	BJS_04614	7354119..7354835	-
alfa/beta hidrolase	BJS_04620	7362933..7363715	-
nítrico óxido redutase	BJS_05549	8384901..8387222	-
hidroxilase salicilato	BJS_05550	8387333..8387773	-
hidroxilase salicilato	BJS_05551	8387982..8388500	-
histidina quinase	BJS_05552	8388585..8391323	-
proteína de fusão de membrana	BJS_05554	8391720..8392790	-
receptor regulador de resposta (proteína CheY)	BJS_05556	8395222..8395587	-
Glicosil hidrolase	BJS_05558	8396969..8398807	-
Mono-oxigenase	BJS_05559	8398872..8400482	-
proteína de stress universal A	BJS_05822	4653963..4654214	-
regulador de transcrição, AraC	BJS_05834	4639961..4640890	-
Metiltransferase tipo 11	BJS_05835	4639011..4639724	-
pilus tipo IV, Pilz	BJS_05836	4638040..4638282	-
ABC transportador	BJS_05845	4628780..4629268	-
2-dehidropantoato 2-redutase	BJS_05846	4627661..4628584	-
domínio Cupredoxin	BJS_05848	4625728..4626825	-
regulador de transcrição	BJS_05850	4622617..4623603	-

Tabela 10 – Continuação

Anotação	ID	Coordenada	Contig
arilamina N-acetiltransferase	BJS_05852	4619645..4620499	-
Domínio NTF2	BJS_05856	4615429..4615794	-
receptor de TonB	BJS_05866	4599350..4601740	-
metiltransferase	BJS_05982	4487823..4489466	-
P-loop nucleosídeo hidrolase	BJS_05983	4483761..4487813	-
particionamento de cromossomo parB	BJS_05987	4477149..4478033	-
particionamento de cromossomo parB	BJS_05988	4476241..4477146	-
domínio lipase	BJS_05989	4475125..4475838	-
MFS transportador	BJS_06001	4460454..4461032	-
histidina quinase	BJS_06004	4456686..4458524	-
regulador transcricional, família AraC	BJS_06023	4424177..4425232	-
Alquil hidroperóxido redutase	BJS_06025	4421447..4422121	-
Colina desidrogenase	BJS_06027	4417726..4419396	-
oxalato descarboxilase OxdD	BJS_06028	4416369..4417595	-
proteína secreção RTX	BJS_06765	3548765..3550303	-
metiltransferase	BJS_06766	3547989..3548675	-
AMP-dependente sintetase	BJS_06770	3540361..3543093	-
Asparagina sintase	BJS_06772	3536343..3538331	-
transferase	BJS_06774	3533863..3535086	-
regulador de resposta do sistema de dois-componentes	BJS_06775	3533102..3533623	-
epimerase NAD-dependente	BJS_06777	3530351..3531334	-
transferase	BJS_06778	3529538..3530281	-
Piridoxal fosfato-dependente transferase	BJS_06779	3528402..3529541	-
manose-1-fosfato guaniltransferase	BJS_06780	3527715..3528356	-
Acil-carreador	BJS_06781	3527020..3527298	-
acil-CoA sintetase	BJS_06782	3525407..3526969	-
asparagina sintetase	BJS_06783	3523383..3525410	-
NH (3) sintetase	BJS_06784	3522466..3523377	-
transferases	BJS_06787	3518449..3519129	-
NADP-dependente de domínio oxidoreductase	BJS_06793	3511675..3512394	-
transporte de manganês	BJS_06928	1020070..1021380	-
porina seletiva de carboidratos	BJS_06931	1023116..1024486	-
Ureohidrolase	BJS_06938	1031178..1031585	-
NADP	BJS_06939	1031601..1032386	-
álcool desidrogenase	BJS_06940	1032738..1033145	-
álcool desidrogenase	BJS_06941	1033142..1033681	-
proteína induzida por estresse	BJS_06943	1034883..1035110	-
piruvato desidrogenase	BJS_06951	1040067..1041506	-
DNA ligase	BJS_06952	1041534..1041899	-
histidina quinase	BJS_06958	1048283..1049452	-
DNA ligase	BJS_06960	1051765..1052406	-
proteína histona	BJS_06966	1059569..1059943	-
Epóxido hidrolase	BJS_06973	1067063..1068022	-
Domínio YidB	BJS_06983	1076534..1076821	-
oxidoreductase	BJS_06991	1083766..1084557	-
Alfa/Beta hidrolase	BJS_06994	1086485..1087834	-
Piridoxal fosfato-dependente transferase	BJS_07006	1009941..1010159	-
beta-lactamase	BJS_07011	1002374..1003360	-
Citocromo c oxidase subunidade II	BJS_07012	1001399..1001710	-
citocromo C peroxidase	BJS_07016	990039..990317	-
Domínio RmlC	BJS_07017	986542..986871	-
MFS transportador	BJS_07021	977444..978091	-
domínio hidrolase	BJS_07023	976654..976956	-
oxalato descarboxilase OxdD	BJS_07026	972904..974166	-
NAD(P)	BJS_07031	963834..964370	-
FAD-dependente oxigenase	BJS_07037	959559..960971	-
álcool desidrogenase	BJS_07038	958471..959343	-
fosfo-dependente de metal	BJS_07039	957436..957723	-
MFS transportador	BJS_07041	955798..956334	-
regulador de transcrição, a família GntR	BJS_07043	953584..955014	-
álcool desidrogenase	BJS_07044	952834..953556	-

Tabela 10 – Continuação

Anotação	ID	Coordenada	Contig
Proteína sensível rica em triptofano	BJS_07053	947001..947546	-
Nex18 simbioticamente induzida	BJS_07054	946411..946914	-
regulador transcricional	BJS_07055	945601..946200	-
proteína CsbD	BJS_07056	944755..944979	-
domínio hidrolase	BJS_07059	942179..942559	-
ATPase	BJS_07060	939639..941669	-
proteína KAIC	BJS_07061	938076..939635	-
Sacarose-6-fosfato	BJS_07064	933630..935534	-
riboquinase	BJS_07069	928385..929470	-
DNA helicase dependente de ATP	BJS_07073	920610..922067	-
regulador transcricional, HTH	BJS_07077	918266..918496	-
Fitanoil CoA-dioxigenase	BJS_07552	395852..396595	-
MFS transportador	BJS_07553	394531..395835	-
oxigenase	BJS_07554	392912..394534	-
glutamina sintetase	BJS_07595	337162..338499	-
proteína de membrana	BJS_07597	335181..335879	-
amidotransferase	BJS_07598	334031..334846	-
desaturase de ácidos graxos	BJS_07599	332976..334034	-
hidroxineurosporeno-O-metiltransferase	BJS_07701	8829875..8830873	-
Fe-S- oxidorreductase	BJS_07705	8832433..8833806	-
sulfatase	BJS_07706	8834238..8835713	-
adenilato-ciclase	BJS_07709	8839235..8840911	-
proteína precursora de translocação twin-arginina	BJS_07716	8848039..8849298	-
regulador transcricional, família XRE	BJS_08151	9342550..9342912	-
peptidase Zn	BJS_08152	9342912..9343796	-
pirofosfatase	BJS_08155	9345754..9346650	-
desoxicitidina trifosfato deaminase	BJS_08156	9347124..9347864	-
NUDIX hidrolase	BJS_08161	9350970..9351440	-
fosfodiesterase	BJS_08162	9351437..9351901	-
Taurina dioxigenase	BJS_08196	180855..181742	-
peptidase associada a resposta SOS	BJS_08197	179277..179591	-
proteína FixW	BJS_08204	170357..171238	-
regulador transcricional tipo HTH	BJS_08209	164754..165035	-
dioxigenase	BJS_08211	162920..163807	-
fosfohidrolase	BJS_08213	155642..158959	-
peptídeo de sinal	BJS_08214	155111..155404	-
regulador transcricional	BJS_08219	150685..151329	-
Domínio Kila	BJS_08224	145754..146512	-
Fosfoenolpiruvato	BJS_08233	135083..135352	-
xilose isomerase	BJS_08235	133502..133846	-
proteína de ligação	BJS_08236	133122..133352	-
DNA polimerase III	BJS_08241	124274..124591	-
proteína Hip A	BJS_08243	122844..123719	-
proteína Hip A	BJS_08244	122426..122842	-
fosfolipase D	BJS_08245	118911..121712	-
glutamil endopeptidase S2	BJS_08254	96080..96979	-
aspartato aminotransferase	BJS_08258	91505..92803	-
amidofosforibosiltransferase, PurF	BJS_08259	88648..90228	-
haloácido desidrogenase	BJS_08260	87965..88279	-
succinato semialdeído desidrogenase	BJS_08261	85326..86360	-
glucosil hidrolase	BJS_08263	83265..84179	-
regulador de transcrição, família LacI	BJS_08265	80225..81316	-
MFS transportador	BJS_08267	76029..77225	-
Enoil-CoA hidratase	BJS_08269	74404..75219	-
proteína de conjugação, trbI	BJS_08451	8705477..8706679	-
proteína de conjugação, trbG	BJS_08452	8704407..8705486	-
proteína de conjugação, trbF	BJS_08453	8703829..8704410	-
proteína de conjugação, trbL	BJS_08454	8702356..8703717	-
proteína de conjugação, trbJ	BJS_08456	8701270..8702028	-
proteína de conjugação, trbE	BJS_08457	8698835..8701273	-
proteína de conjugação	BJS_08458	8698552..8698827	-

Tabela 10 – Continuação

Anotação	ID	Coordenada	Contig
proteína de conjugação, trbC	BJS_08459	8698211..8698552	-
proteína de conjugação	BJS_08460	8697264..8698214	-
ABC transportador	BJS_08464	8693160..8694326	-
RNA polimerase	BJS_08474	8681604..8682056	-
Acil-CoA-N-aciltransferase	BJS_08476	8679780..8680283	-
oxidase	BJS_08478	8677158..8678489	-
proteína de conjugação TrbI	BJS_08494	9369367..9370569	-
proteína de conjugação	BJS_08495	9368447..9369370	-
proteína de conjugação, trbF	BJS_08496	9367656..9368339	-
proteína de conjugação, trbL	BJS_08497	9366651..9367562	-
celulose sintase	BJS_08517	9431369..9431776	-
DNase dependente-Mg	BJS_08522	9426467..9427231	-
ATPase	BJS_08525	9422457..9424337	-
proteína de membrana	BJS_08541	1122755..1123249	-
Exonuclease, RNase T	BJS_08547	1116978..1117478	-
DNA polimerase III	BJS_08548	1116685..1116909	-
Repetição SPW	BJS_08553	1113534..1113896	-
NADH desidrogenase	BJS_08556	1111131..1111553	-
domínio barril dimérica alfa-beta	BJS_08599	8785928..8786227	-
glioxalase	BJS_08600	8786500..8787123	-
hidrolase	BJS_08601	8787136..8787984	-
desidrogenase	BJS_08602	8787981..8788859	-
regulador transcricional HTH	BJS_08604	8789771..8790379	-
proteína de membrana	BJS_08605	8790644..8791228	-
quinase	BJS_08694	990458..993355	-
peptidase M20	BJS_08734	1672084..1673331	-
Histidina quinase sensor de dois-componentes	BJS_08736	2301967..2303196	-
transportador de cromato	BJS_08738	4623983..4625197	-
Metalo-beta-lactamase	BJS_08747	2321140..2322240	-
Domínio OmpA-like	BJS_08762	2488626..2489687	-
oxidorreductase quinona	BJS_08764	984329..985372	-
3 aciltransferase	BJS_08765	2460075..2461115	-
Dolicol-fosfato manosiltransferase	BJS_08779	5126957..5127931	-
domínio Cupredoxin	BJS_08780	5471596..5472570	-
domínio Cupredoxin	BJS_08781	5471231..5471605	-
domínio Cupredoxin	BJS_08815	4682656..4683405	-
Timidilato sintase	BJS_08839	9344023..9344634	-
proteína CopG	BJS_08930	9411255..9411584	-
OsmC / família Ohr	BJS_08934	5640831..5641148	-
regulador de resposta dos dois-componentes	BJS_08942	1092425..1092733	-
proteína de ligação	BJS_09005	1251062..1251304	-
MFS transportador	BJS_09024	6282564..6282794	-
Proteína CheY	BJS_09065	988595..989296	-

Artigos científicos a partir dos resultados obtidos nesta tese

- Barros-Carvalho, G.A. et al. An efficient approach to explore and discriminate anomalous regions in bacterial genomes based on maximum entropy. **Submetido para BMC Bioinformatics.**
- Barros-Carvalho, G.A. et al. A detailed survey of IS elements in the genomes of Brazilian *Bradyrhizobium* reveals putative impacts on biological nitrogen fixation. **Em preparação.**

Referências Bibliográficas

- AHUJA, N., KUMAR, P. & BHATNAGAR, R. (2004). The adenylate cyclase toxins. *Critical Reviews in Microbiology* **30**, 187–196. 53, 54
- AL SAFADI, R., AMOR, S., HERY-ARNAUD, G., SPELLERBERG, B. & LANOTTE, P. (2010). Enhanced expression of lmb gene encoding laminin-binding protein in *Streptococcus agalactiae* strains harboring IS1548 in scpB-lmb intergenic region. *PLoS ONE* **5**, e10794. 12, 42
- ALBRECHT, S., MAIER, R., HANUS, F., RUSSELL, S., EMERICH, D. & EVANS, H. (1979). Hydrogenase in *Rhizobium japonicum* increases nitrogen fixation by nodulated soybeans. *Science* **203**, 1255–1257. 46
- ALFANO, J. & COLLMER, A. (2004). Type III secretion system effector proteins: double agents in bacterial disease and plant defense. *Annual Review of Phytopathology* **42**, 385–414. 45
- ALTSCHUL, S., GISH, W., MILLER, W., MYERS, E. & LIPMAN, D. (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410. 14, 20, 22, 23, 30
- ALTSCHUL, S., MADDEN, T., SCHAFER, A., ZHANG, J., ZHANG, Z., MILLER, W. & LIPMAN, D. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389–3402. 14, 20, 22, 23
- AZAD, R. & LAWRENCE, J. (2011). Towards more robust methods of alien gene detection. *Nucleic Acids Research* **39**, e56. 65
- BALLEZA, E., LÓPEZ-BOJORQUEZ, L., MARTÍNEZ-ANTONIO, A., RESENDIS-ANTONIO, O., LOZADA-CHÁVEZ, I., BALDERAS-MARTÍNEZ, Y., ENCARNACIÓN, S. & COLLADO-VIDES, J. (2009). Regulation by transcription factors in bacteria: beyond description. *FEMS Microbiology Reviews* **33**, 133–151. 54

- BARRETT, A. (1994). Classification of peptidases. *Methods in Enzymology* **244**, 1–15. 62
- BATISTA, J., HUNGRIA, M., BARCELLOS, F., FERREIRA, M. & MENDES, I. (2007). 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* **53**, 270–284. 8
- BATISTA, J., TORRES, A. & HUNGRIA, M. (2010). Towards a two-dimensional proteomic reference map of *Bradyrhizobium japonicum* CPAC 15: spotlighting “hypothetical proteins”. *Proteomics* **10**, 3176–3189. 24, 49
- BENTLEY, S. & PARKHILL, J. (2004). Comparative genomic structure of prokaryotes. *Annual Review of Genetics* **38**, 771–792. 15, 69
- BERNABÉU-RODA, L., CALATRAVA-MORALES, N., CUÉLLAR, V. & M.J., S. (2015). Characterization of surface motility in *Sinorhizobium meliloti*: regulation and role in symbiosis. *Symbiosis* **67**, 79–90. 58
- BERNSTEIN, H. (2000). The biogenesis and assembly of bacterial membrane proteins. *Current Opinion in Microbiology* **3**, 203–209. 56
- BICKHART, D., GOGARTEN, J. & LAPIERRE, P. (2009). Insertion sequence content reflects genome plasticity in strains of the root nodule actinobacterium *Frankia*. *BMC Genomics* **10**, 468. 47
- BLANCAFLOR, E., WANG, Y. & MOTES, C. (2006). Organization and function of the actin cytoskeleton in developing root cells. *International Review of Cytology* **252**, 219–264. 4
- BLATCH, G. & LÄSSLE, M. (1999). The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *BioEssays* **21**, 932–939. 54
- BODDEY, L. & HUNGRIA, M. (1997). Phenotypic grouping of Brazilian *Bradyrhizobium* strains which nodulate soybean. *Biology and Fertility of Soils* **25**, 407–415. 8, 74
- BOHLIN, J. & SKJERVE, E. (2009). Examination of genome homogeneity in prokaryotes using genomic signatures. *PLoS ONE* **4**, e8113. 15, 24
- BOHLIN, J., SNIPEN, L., HARDY, S., KRISTOFFERSEN, A., LAGESEN, K., DØNSVIK, T., SKJERVE, E. & USSERY, D. (2010). Analysis of intra-genomic GC content homogeneity within prokaryotes. *BMC Genomics* **11**, 464. 24

- BOTHE, H., SCHMITZ, O., YATES, M. & NEWTON, W. (2010). Nitrogen fixation and hydrogen metabolism in cyanobacteria. *Microbiology and Molecular Biology Reviews* **74**, 529–551. 46
- BRETON, C., SNAJDROVÁ, L., JEANNEAU, C., KOCA, J. & IMBERTY, A. (2006). Structures and mechanisms of glycosyltransferases. *Glycobiology* **16**, 29R–37R. 52, 53
- BRU, C., COURCELLE, E., CARRÈRE, S., BEAUSSE, Y., DALMAR, S. & KAHN, D. (2005). The ProDom database of protein domain families: more emphasis on 3D. *Nucleic Acids Research* **33**, D212–D215. 14
- BURKOVSKI, A. & KRÄMER, R. (2002). Bacterial amino acid transport proteins: occurrence, functions, and significance for biotechnological applications. *Applied Microbiology and Biotechnology* **58**, 265–274. 56
- CALHOUN, D., BONNER, C., GU, W., XIE, G. & JENSEN, R. (2001). The emerging periplasm-localized subclass of AroQ chorismate mutases, exemplified by those from *Salmonella typhimurium* and *Pseudomonas aeruginosa*. *Genome Biology* **2**, RESEARCH0030. 52
- CASACUBERTA, E. & GONZÁLEZ, J. (2013). The impact of transposable elements in environmental adaptation. *Molecular Ecology* **22**, 1503–1517. 10, 11
- CASADEVALL, A. & PIROFSKI, L. (2001). Host-pathogen interactions: the attributes of virulence. *The Journal of Infectious Diseases* **184**, 337–344. 59
- CASJENS, S. (1998). The diverse and dynamic structure of bacterial genomes. *Annual Review of Genetics* **32**, 339–377. 15
- CERVEAU, N., LECLERCQ, S., BOUCHON, D. & CORDAUX, R. (2011). *Evolutionary Biology – Concepts, Biodiversity, Macroevolution and Genome Evolution*, chap. Evolutionary Dynamics and Genomic Impact of Prokaryote Transposable Elements. Springer Berlin Heidelberg, pp. 291–312. 11, 12, 35, 42
- CERVENY, L., STRASKOVA, A., DANKOVA, V., HARTLOVA, A., CECKOVA, M., STAUD, F. & STULIK, J. (2013). Tetra-tricopeptide repeat motifs in the world of bacterial pathogens: role in virulence mechanisms. *Infection and Immunity* **81**, 629–635. 54
- CONAB (2016). Companhia nacional de abastecimento. Soja, safra 2014/2015. URL <http://www.conab.gov.br/>. 3

- CONESA, A., GÖTZ, S., GARCÍA-GÓMEZ, J., TEROL, J., TALÓN, M. & ROBLES, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676. 23
- CONSORTIUM, T. U. (2008). The universal protein resource (uniprot). *Nucleic Acids Research* **36**, D190–D195. 21
- CONTRERAS-MOREIRA, B. & VINUESA, P. (2013). GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Applied and Environmental Microbiology* **79**, 7696–7701. 22
- COSTA, T., FELISBERTO-RODRIGUES, C., MEIR, A., PREVOST, M., REDZEJ, A., TROKTER, M. & WAKSMAN, G. (2015). Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nature Reviews Microbiology* **13**, 343–359. 59
- CYTRYN, E., SANGURDEKAR, D., STREETER, J., FRANCK, W., CHANG, W., STACEY, G., EMERICH, D., JOSHI, T., XU, D. & SADOWSKY, M. (2007). Transcriptional and physiological responses of *Bradyrhizobium japonicum* to desiccation-induced stress. *Journal of Bacteriology* **189**, 6751–6762. 75
- DA SILVA, A., FERRO, J., REINACH, F., FARAH, C., FURLAN, L., QUAGGIO, R., MONTEIRO-VITORELLO, C., VAN SLUYS, M., ALMEIDA, N., ALVES, L., DO AMARAL, A., BERTOLINI, M., CAMARGO, L., CAMAROTTE, G., CANNANAN, F., CARDOZO, J., CHAMBERGO, F., CIAPINA, L., CICARELLI, R., COUTINHO, L., CURSINO-SANTOS, J., EL-DORRY, H., FARIA, J., FERREIRA, A., FERREIRA, R., FERRO, M., FORMIGHIERI, E., FRANCO, M., GREGGIO, C., GRUBER, A., KATSUYAMA, A., KISHI, L., LEITE, R., LEMOS, E., LEMOS, M., LOCALI, E., MACHADO, M., MADEIRA, A., MARTINEZ-ROSSI, N., MARTINS, E., MEIDANIS, J., MENCK, C., MIYAKI, C., MOON, D., MOREIRA, L., NOVO, M., OKURA, V., OLIVEIRA, M., OLIVEIRA, V., PEREIRA, H., ROSSI, A., SENA, J., SILVA, C., DE SOUZA, R., SPINOLA, L., TAKITA, M., TAMURA, R., TEIXEIRA, E., TEZZA, R., TRINDADE DOS SANTOS, M., TRUFFI, D., TSAI, S., WHITE, F., SETUBAL, J. & KITAJIMA, J. (2002). Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature* **417**, 459–463. 67
- DALBEY, R., WANG, P. & KUHN, A. (2011). Assembly of bacterial inner membrane proteins. *Annual Review of Biochemistry* **80**, 161–187. 56
- DALE, C., PLAGUE, G., WANG, B., OCHMAN, H. & MORAN, N. (2002). Type III secretion systems and the evolution

- of mutualistic endosymbiosis. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 12397–12402. 45
- DARMON, E. & LEACH, D. (2014). Bacterial genome instability. *Microbiology and Molecular Biology Reviews* **78**, 1–39. 10, 15, 64, 67, 69, 72
- DAVIES, B. & WALKER, G. (2007). Identification of novel *Sinorhizobium meliloti* mutants compromised for oxidative stress protection and symbiosis. *Journal of Bacteriology* **189**, 2110–2113. 62
- DAVIS, J., RUBIN, A. & SAUER, R. (2011). Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Research* **39**, 1131–1141. 21
- DEAKIN, W. & BROUGHTON, W. (2009). Symbiotic use of pathogenic strategies: rhizobial protein secretion systems. *Nature Reviews Microbiology* **7**, 312–320. 59
- DEAKIN, W., MARIE, C., SAAD, M., KRISHNAN, H. & BROUGHTON, W. (2005). NopA is associated with cell surface appendages produced by the type III secretion system of *Rhizobium* sp. strain NGR234. *Molecular Plant-Microbe Interactions* **18**, 499–507. 56
- DEGRASSI, G., DEVESCOVI, G., BIGIRIMANA, J. & VENTURI, V. (2010). *Xanthomonas oryzae* pv. *oryzae* XKK.12 contains an AroQ gamma chorismate mutase that is involved in rice virulence. *Phytopathology* **100**, 262–270. 52
- DELAMUTA, J., RIBEIRO, R. & ORMEÑO-ORRILLO, E. (2013). Polyphasic evidence supporting the reclassification of *Bradyrhizobium japonicum* group Ia strains as *Bradyrhizobium diazoefficiens* sp. nov. *International Journal of Systematic and Evolutionary Microbiology* **63**, 3342–3351. 3, 7, 12, 74
- DELMOTTE, N., AHRENS, C., KNIEF, C., QELI, E., KOCH, M., FISCHER, H., VORHOLT, J., HENNECKE, H. & PESSI, G. (2010). An integrated proteomics and transcriptomics reference data set provides new insights into the *Bradyrhizobium japonicum* bacteroid metabolism in soybean root nodules. *Proteomics* **10**, 1391–1400. 7, 22, 24, 38, 41, 43, 49, 52, 62
- DIXON, R. & KAHN, D. (2004). Genetic regulation of biological nitrogen fixation. *Nature Reviews Microbiology* **2**, 621–631. 6

- DJORDJEVIC, M., CHEN, H., NATERA, S., VAN NOORDEN, G., MENZEL, C., TAYLOR, S., RENARD, C., GEIGER, O., WEILLER, G. & CONSORTIUM, S. D. S. (2003). A global analysis of protein expression profiles in *Sinorhizobium meliloti*: discovery of new genes for nodule occupancy and stress adaptation. *Molecular Plant-Microbe Interactions* **16**, 508–524. 75
- DOBRINDT, U., HOCHHUT, B., HENTSCHEL, U. & HACKER, J. (2004). Genomic islands in pathogenic and environmental microorganisms. *Nature Reviews Microbiology* **2**, 414–424. 69, 71, 72
- DOERKS, T., VAN NOORT, V., MINGUEZ, P. & BORK, P. (2012). Annotation of the *M. tuberculosis* hypothetical orfeome: adding functional information to more than half of the uncharacterized proteins. *PLoS One* **7**, e34302. 14
- DOLATABADIAN, A., SANAVY, S., GHANATI, F. & GRESSHOFF, P. (2012). Morphological and physiological response of soybean treated with the microsymbiont *Bradyrhizobium japonicum* pre-incubated with genistein. *South African Journal of Botany* **79**, 9–18. 4
- DUFRAIGNE, C., FERTIL, B., LESPINATS, S., GIRON, A. & DESCHAVANNE, P. (2005). Detection and characterization of horizontal transfers in prokaryotes using genomic signature. *Nucleic Acids Research* **33**, e6. 15, 16, 68
- EDQVIST, P., BRÖMS, J., BETTS, H., FORSBERG, A., PALLÉN, M. & FRANCIS, M. (2006). Tetratricopeptide repeats in the type III secretion chaperone, LcrH: their role in substrate binding and secretion. *Molecular Microbiology* **59**, 31–44. 54
- ESTREM, S., ROSS, W., GAAL, T., CHEN, Z., NIU, W., EBRIGHT, R. & GOURSE, R. (1999). Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase alpha subunit. *Genes & Development* **13**, 2134–2147. 21
- FAURE, D. (2002). The family-3 glycoside hydrolases: from housekeeping functions to host-microbe interactions. *Applied and Environmental Microbiology* **68**, 1485–1490. 52
- FERGUSON, B., INDRASUMUNAR, A., HAYASHI, S., LIN, M., LIN, Y., REID, D. & GRESSHOFF, P. (2010). Molecular analysis of legume nodule development and autoregulation. *Journal of Integrative Plant Biology* **52**, 61–76. 3, 4

- FERGUSON, B., LIN, M. & GRESSHOFF, P. (2013). Regulation of legume nodulation by acidic growth conditions. *Plant Signaling & Behavior* **8**, e23426. 8
- FILLOUX, A. (2010). Secretion signal and protein targeting in bacteria: a biological puzzle. *Journal of Bacteriology* **192**, 3847–3849. 63
- FINAN, T. (2002). Evolving insights: symbiosis islands and horizontal gene transfer. *Journal of Bacteriology* **184**, 2855–2856. 32, 69
- FINN, R., MISTRY, J., TATE, J., COGGILL, P., HEGER, A., POLLINGTON, J., GAVIN, O., GUNASEKARAN, P., CERIC, G., FORSLUND, K., HOLM, L., SONNHAMMER, E., EDDY, S. & BATEMAN, A. (2010). The Pfam protein families database. *Nucleic Acids Research* **38**, D211–D222. 14
- FRANCK, S., FRANCK, W., BIRKE, S., CHANG, W., SANGURDEKAR, D., CYTRYN, E., JOSHI, T., SADOWSKY, M., STACEY, G., XU, D. & EMERICH, D. (2014). Comparative transcriptomic analysis of symbiotic *Bradyrhizobium japonicum*. *Symbiosis* **63**, 123–135. 7, 22, 24, 49, 52
- FREES, D., BRØNDSTED, L. & INGMER, H. (2013). Bacterial proteases and virulence. *Subcellular Biochemistry* **66**, 161–192. 62
- FREIBERG, C., FELLAY, R., BAIROCH, A., BROUGHTON, W., ROSENTHAL, A. & PERRET, X. (1997). Molecular basis of symbiosis between *Rhizobium* and legumes. *Nature* **387**, 394–401. 33
- GAL-MOR, O. & FINLAY, B. (2006). Pathogenicity islands: a molecular toolbox for bacterial virulence. *Cellular Microbiology* **8**, 1707–1719. 71
- GALPERIN, M. (2001). Conserved 'hypothetical' proteins: new hints and new puzzles. *Comparative and Functional Genomics* **2**, 14–18. 13
- GALPERIN, M. & KOONIN, E. (2004). 'Conserved hypothetical' proteins: prioritization of targets for experimental study. *Nucleic Acids Research* **32**, 5452–5463. 13, 14
- GARCIA-VALLVE, S., GUZMAN, E., MONTERO, M. & ROMEU, A. (2003). HGT-DB: a database of putative horizontally transferred genes in prokaryotic complete genomes. *Nucleic Acids Research* **31**, 187–189. 16, 30

- GARDY, J., SPENCER, C., WANG, K., ESTER, M., TUSNÁDY, G., SIMON, I., HUA, S., deFAYS, K., LAMBERT, C., NAKAI, K. & BRINKMAN, F. (2003). PSORT-B: Improving protein subcellular localization prediction for Gram-negative bacteria. *Nucleic Acids Research* **31**, 3613–3617. 23
- GLOSTER, T. (2014). Advances in understanding glycosyltransferases from a structural perspective. *Current Opinion in Structural Biology* **28**, 131–141. 52
- GOMES, D., BATISTA, J., ROLLA, A., DA SILVA, L., BLOCH, C., GALLI-TERASAWA, L. & HUNGRIA, M. (2014). Proteomic analysis of free-living *Bradyrhizobium diazoefficiens*: highlighting potential determinants of a successful symbiosis. *BMC Genomics* **15**, 643. 24, 49
- GOODLASS, G. & SMITH, K. (1979). Effects of ethylene on root extension and nodulation of pea (*Pisum sativum* L.) and white clover (*Trifolium repens* L.). *Plant Soil* **51**, 387–395. 74
- GORDON, J., TOWSEY, M., HOGAN, J., MATHEWS, S. & TIMMS, P. (2006). Improved prediction of bacterial transcription start sites. *Bioinformatics* **22**(142–148). 21
- GRENNAN, A. (2008). Ethylene response factors in jasmonate signaling and defense response. *Plant Physiology* **146**, 1457–1458. 74
- GUIASU, S. & SHENITZER, A. (1985). The principle of maximum entropy. *The Mathematical Intelligencer* **7**, 42–48. 17
- GÖTTFERT, M., RÖTHLISBERGER, S. & KÜNDIG, C. (2001). Potential symbiosis-specific genes uncovered by sequencing a 410-kilobase DNA region of the *Bradyrhizobium japonicum* chromosome. *Journal of Bacteriology* **183**, 1405–1412. 7, 32, 73
- GÜRLEBECK, D., THIEME, F. & BONAS, U. (2006). Type III effector proteins from the plant pathogen *Xanthomonas* and their role in the interaction with the host plant. *Journal of Plant Physiology* **163**, 233–255. 59
- HAN, Y., LIU, X., BENNY, U., KISTLER, H. & VANETTEN, H. (2001). Genes determining pathogenicity to pea are clustered on a supernumerary chromosome in the fungal plant pathogen *Nectria haematococca*. *The Plant Journal* **25**, 305–314. 62

- HANIN, M., JABBOURI, S., QUESADA-VINCENS, D., FREIBERG, C., PERRET, X., PROMÉ, J., BROUGHTON, W. & FELLAY, R. (1997). Sulphation of *Rhizobium* sp. NGR234 Nod factors is dependent on *noeE*, a new host-specificity gene. *Molecular Microbiology* **24**, 1119–1129. 73
- HANSEL, A., AXELSSON, R., LINDBERG, P., TROSHINA, O., WUNSCHERS, R. & LINDBLAD, P. (2001). Cloning and characterisation of a *hyp* gene cluster in the filamentous *Cyanobacterium nostoc* sp. strain PCC 73102. *FEMS Microbiology Letters* **201**, 59–64. 46
- HEATON, B., HERROU, J., BLACKWELL, A., WYSOCKI, V. & CROSSON, S. (2012). Molecular structure and function of the novel BrnT/BrnA toxin-antitoxin system of *Brucella abortus*. *Journal of Biological Chemistry* **287**, 12098–12110. 59
- HEGDE, R. & BERNSTEIN, H. (2006). The surprising complexity of signal sequences. *Trends in Biochemical Sciences* **31**, 563–571. 63
- HENNECKE, H. (1990). Nitrogen fixation genes involved in the *Bradyrhizobium japonicum*-soybean symbiosis. *FEBS Letters* **268**, 422–426. 7
- HENTSCHEL, U., STEINERT, M. & HACKER, J. (2000). Common molecular mechanisms of symbiosis and pathogenesis. *Trends in Microbiology* **8**, 226–231. 58, 59
- HERSHEY, D., LU, X., ZI, J. & PETERS, R. (2014). Functional conservation of the capacity for ent-kaurene biosynthesis and an associated operon in certain rhizobia. *Journal of Bacteriology* **196**, 100–106. 50
- HONTELEZ, J., LANKHORST, R., KATINAKIS, P., VAN DEN BOS, R. & VAN KAMMEN, A. (1989). Characterization and nucleotide sequence of a novel gene *fixW* upstream of the *fixABC* operon in *Rhizobium leguminosarum*. *Molecular and General Genetics* **218**, 536–544. 73
- HOTSON, A. & MUDGETT, M. (2004). Cysteine proteases in phytopathogenic bacteria: identification of plant targets and activation of innate immunity. *Current Opinion in Plant Biology* **7**, 384–390. 43
- HSIAO, W., WAN, I., JONES, S. & BRINKMAN, F. (2003). IslandPath: aiding detection of genomic islands in prokaryotes. *Bioinformatics* **19**, 418–420. 16, 30
- HU, Y. & RIBBE, M. (2013). Nitrogenase assembly. *Biochimica et Biophysica Acta* **1827**(1112–1122). 38

- HUNGRIA, M., CAMPO, R., MENDES, I. & GRAHAM, P. (2006a). *Nitrogen nutrition and sustainable plant productivity*, chap. 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. Houston, Texas: Studium Press, LLC, pp. 43–93. 3, 8
- HUNGRIA, M., FRANCHINI, J. & CAMPO, R. (2006b). Nitrogen nutrition of soybean in Brazil: Contributions of biological N₂ fixation and n fertilizer to grain yield. *Canadian Journal of Plant Science* **86**, 927–939. 3
- HUNGRIA, M. & FRANCO, A. (1993). Effects of high temperature on nodulation and nitrogen fixation by *Phaseolus vulgaris* L. *Plant and Soil* **149**, 95–102. 8
- HUNGRIA, M., MENDES, I. & DE BRUIJN, F. (2014). *Biological Nitrogen Fixation*, chap. Nitrogen fixation with soybean: the perfect symbiosis? New Jersey: Wiley Publisher, Hoboken. 8
- HUNGRIA, M., NISHI, C. & STACEY, G. (1996). Comparasion between parental and variant soybean *Bradyrhizobium* strains with regard to the prodution of lipo-chitin nodulation signals, early stages of root infection, nodule occupancy, and N₂ fixation rates. *Plant and Soil* **186**, 331–341. 8, 71
- HUNGRIA, M. & VARGAS, M. (2000). Environmental factors affecting N₂ fixation in grain legumes in the tropics, with an emphasis on Brazil. *Field Crops Research* **65**, 151–164. 8
- HUNTER, S., APWEILER, R., ATTWOOD, T., BAIROCH, A., BATEMAN, A., BINNS, D., BORK, P., DAS, U., DAUGHERTY, L., DUQUENNE, L., FINN, R., GOUGH, J., HAFT, D., HULO, N., KAHN, D., KELLY, E., LAUGRAUD, A., LETUNIC, I., LONSDALE, D., LOPEZ, R., MADERA, M., MASLEN, J., McANULLA, C., McDOWALL, J., MISTRY, J., MITCHELL, A., MULDER, N., NATALE, D., ORENGO, C., QUINN, A., SELENGUT, J., SIGRIST, C., THIMMA, M., THOMAS, P., VALENTIN, F., WILSON, D., WU, C. & YEATS, C. (2009). InterPro: the integrative protein signature database. *Nucleic Acids Research* **37**, D211–D215. 14, 21, 23
- IIDA, T., ITAKURA, M., ANDA, M., SUGAWARA, M., ISAWA, T., OKUBO, T., SATO, S., CHIBA-KAKIZAKI, K. & MINAMISAWA, K. (2015). Symbiosis island shuffling with abundant insertion sequences in the genomes of extra-slow-growing strains of soybean bradyrhizobia. *Applied and Environmental Microbiology* **81**, 4143–4154. 13, 32
- IVANKOV, D., PAYNE, S., GALPERIN, M., BONISSONE, S., PEVZNER, P. & FRISHMAN, D. (2013). How many signal peptides are there in bacteria? *Environmental Microbiology* **15**, 983–990. 63

- IZQUIERDO, J., SIZOVA, M. & LYND, L. (2010). Diversity of bacteria and glycosyl hydrolase family 48 genes in cellulolytic consortia enriched from thermophilic biocompost. *Applied and Environmental Microbiology* **76**, 3545–3553. 52
- JANCZAREK, M. & SKORUPSKA, A. (2007). The *Rhizobium leguminosarum* bv. *trifolii* RosR: transcriptional regulator involved in exopolysaccharide production. *Molecular Plant-Microbe Interactions* **20**, 867–881. 54
- JANK, T., BELYI, Y. & AKTORIES, K. (2015). Bacterial glycosyltransferase toxins. *Cellular Microbiology* **17**, 1752–1765. 53
- JAYNES, E. (1957a). Information theory and statistical mechanics. *Physical Review* **106**, 620–630. 17, 26
- JAYNES, E. (1957b). Information theory and statistical mechanics ii. *Physical Review* **108**, 171–190. 17, 26
- JOHN, M., RÖHRIG, H., SCHMIDT, J., WIENEKE, U. & SCHELL, J. (1993). *Rhizobium* NodB protein involved in nodulation signal synthesis is a chitooligosaccharide deacetylase. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 625–629. 56
- JONES, K., KOBAYASHI, H., DAVIES, B., TAGA, M. & WALKER, G. (2007). How rhizobial symbionts invade plants: the *Sinorhizobium*–*Medicago* model. *Nature Reviews Microbiology* **5**, 619–633. 57, 58
- JORDAN, D. (1982). Transfer of *Rhizobium japonicum* Buchanan 1980 to *Bradyrhizobium* gen. nov., a genus of slow growing root-nodule bacteria from leguminous plants. *International Journal of Systematic and Evolutionary Microbiology* **32**, 136–139. 8
- JUHAS, M., VAN DER MEER, J., GAILLARD, M., HARDING, R., HOOD, D. & CROOK, D. (2009). Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiology Reviews* **33**, 376–393. 69
- KANEHISA, M. & GOTO, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* **28**, 27–30. 23
- KANEKO, T., MAITA, H., HIRAKAWA, H., UCHIIKE, N., MINAMISAWA, K., WATANABE, A. & SATO, S. (2011). Complete genome sequence of the soybean symbiont *Bradyrhizobium japonicum* strain USDA6^T. *Genes* **2**, 763–787. 7, 12, 32

- KANEKO, T., NAKAMURA, Y., SATO, S., ASAMIZU, E., KATO, T., SASAMOTO, S., WATANABE, A., IDESAWA, K., ISHIKAWA, A., KAWASHIMA, K., KIMURA, T., KISHIDA, Y., KİYOKAWA, C., KOHARA, M., MATSUMOTO, M., MATSUNO, A., MOCHIZUKI, Y., NAKAYAMA, S., NAKAZAKI, N., SHIMPO, S., SUGIMOTO, M., TAKEUCHI, C., YAMADA, M. & TABATA, S. (2000). Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. *DNA Research* **7**, 331–338. 32
- KANEKO, T., NAKAMURA, Y., SATO, S., MINAMISAWA, K., UCHIUMI, T., SASAMOTO, S., WATANABE, A., IDESAWA, K., IRIGUCHI, M., KAWASHIMA, K., KOHARA, M., MATSUMOTO, M., SHIMPO, S., TSURUOKA, H., WADA, T., YAMADA, M. & TABATA, S. (2002). Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA Research* **9**, 189–197. 7, 12, 19, 32
- KANHERE, A. & VINGRON, M. (2009). Horizontal gene transfers in prokaryotes show differential preferences for metabolic and translational genes. *BMC Evolutionary Biology* **9**, 9. 72
- KARLIN, S. & BURGE, C. (1995). Dinucleotide relative abundance extremes: a genomic signature. *Trends in Genetics* **11**, 283–290. 15, 16, 24, 69
- KARPENAHALLI, M., LUPAS, A. & SÖDING, J. (2007). TPRpred: a tool for prediction of TPR-, PPR- and SEL1-like repeats from protein sequences. *BMC Bioinformatics* **8**, 2. 23
- KAY, S. & BONAS, U. (2009). How *Xanthomonas* type III effectors manipulate the host plant. *Current Opinion in Microbiology* **12**, 37–43. 59
- KIMBREL, J., THOMAS, W., JIANG, Y., CREASON, A., THIREAULT, C., SACHS, J. & CHANG, J. (2013). Mutualistic co-evolution of type III effector genes in *Sinorhizobium fredii* and *Bradyrhizobium japonicum*. *PLOS Pathogens* **9**, e1003204. 43, 45
- KISTNER, C. & PARNISKE, M. (2002). Evolution of signal transduction in intracellular symbiosis. *Trends in Plant Science* **7**, 511–518. 54
- KOONIN, E., MAKAROVA, K. & ARAVIND, L. (2001). Horizontal gene transfer in prokaryotes: quantification and classification. *Annual Review of Microbiology* **55**, 709–742. 15, 64
- KUMAR, K., PRAKASH, A., TASLEEM, M., ISLAM, A., AHMAD, F. & HASSAN, M. (2014). Functional annotation of putative hypothetical proteins from *Candida dubliniensis*. *Gene* **543**, 93–100. 14

- KUNIN, V. & OUZOUNIS, C. (2003). The balance of driving forces during genome evolution in prokaryotes. *Genome Research* **13**, 1589–1594. 10
- KUYKENDALL, L. D., SAXENA, B., DEVINE, T. & UDELL, S. (1992). Genetic diversity in *Bradyrhizobium japonicum* Jordan 1982 and a proposal for *Bradyrhizobium elkanii* sp. nov. *Canadian Journal of Microbiology* **38**, 501–505. 74
- KÄLL, L., KROGH, A. & SONNHAMMER, E. (2004). A combined transmembrane topology and signal peptide prediction method. *Journal of Molecular Biology* **338**, 1027–1036. 23
- LAWRENCE, J. & HENDRICKSON, H. (2005). Genome evolution in bacteria: order beneath chaos. *Current Opinion in Microbiology* **8**, 572–578. 15
- LEE, A., STEWART, J., CLARDY, J. & GANEM, B. (1995). New insight into the catalytic mechanism of chorismate mutases from structural studies. *Chemistry & Biology* **2**, 195–203. 51
- LIBAULT, M., FARMER, A., BRECHENMACHER, L., DRNEVICH, J., LANGLEY, R., BILGIN, D., RADWAN, O., NEECE, D., CLOUGH, S., MAY, G. & STACEY, G. (2010). Complete transcriptome of the soybean root hair cell, a single-cell model, and its alteration in response to *Bradyrhizobium japonicum* infection. *Plant Physiology* **152**, 541–552. 7
- LIMA, W., PAQUOLA, A., VARANI, A., VAN SLUYS, M. & MENCK, C. (2008). Laterally transferred genomic islands in *Xanthomonadales* related to pathogenicity and primary metabolism. *FEMS Microbiology Letters* **281**, 87–97. 29, 66, 67
- LIMA, W., VAN SLUYS, M. & MENCK, C. (2005). Non-gamma-proteobacteria gene islands contribute to the *Xanthomonas* genome. *OMICS* **9**, 160–172. 29, 66, 67
- LING, A. & CORDAUX, R. (2010). Insertion sequence inversions mediated by ectopic recombination between terminal inverted repeats. *PLoS ONE* **5**, e15654. 12
- LIU, W., FENG, X., ZHENG, Y., HUANG, C., NAKANO, C., HOSHINO, T., BOGUE, S., KO, T., CHEN, C., CUI, Y., LI, J., WANG, I., HSU, S., OLDFIELD, E. & GUO, R. (2014). Structure, function and inhibition of ent-kaurene synthase from *Bradyrhizobium japonicum*. *Scientific Reports* **4**, 6214. 50

- LIU, X., INLOW, M. & VANETTEN, H. (2003). Expression profiles of pea pathogenicity (PEP) genes in vivo and in vitro, characterization of the flanking regions of the PEP cluster and evidence that the PEP cluster region resulted from horizontal gene transfer in the fungal pathogen *Nectria haematococca*. *Current Genetics* **44**, 95–103. 62
- LOH, J. & STACEY, G. (2003). Nodulation gene regulation in *Bradyrhizobium japonicum*: a unique integration of global regulatory circuits. *Applied and Environmental Microbiology* **69**, 10–17. 4, 7, 9
- LORY, S., WOLFGANG, M., LEE, V. & SMITH, R. (2004). The multi-talented bacterial adenylate cyclases. *International Journal of Medical Microbiology* **293**, 479–482. 53, 54
- LU, X., HERSHEY, D., WANG, L., BOGDANOVE, A. & PETERS, R. (2015). An ent-kaurene-derived diterpenoid virulence factor from *Xanthomonas oryzae* pv. *oryzicola*. *New Phytologist* **206**, 295–302. 51
- LUBEC, G., AFJEHI-SADAT, L., YANG, J. & JOHN, J. (2005). Searching for hypothetical proteins: theory and practice based upon original data and literature. *Progress in Neurobiology* **77**, 90–127. 13, 14
- LUO, L. & LU, D. (2014). Immunosuppression during *Rhizobium*-legume symbiosis. *Plant Signaling & Behavior* **9**, e28197. 58
- LUO, L., YAO, S.-Y., BECKER, A., RÜBERG, S., YU, G.-Q., ZHU, J.-B. & CHENG, H.-P. (2005). Two new *Sinorhizobium meliloti* LysR-type transcriptional regulators required for nodulation. *Journal of Bacteriology* **187**, 4562–4572. 73
- MACLEAN, A., ANSTEY, M. & FINAN, T. (2008). Binding site determinants for the LysR-type transcriptional regulator PcaQ in the legume endosymbiont *Sinorhizobium meliloti*. *Journal of Bacteriology* **190**, 1237–1246. 73
- MADDOCKS, S. & OYSTON, P. (2008). Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology* **154**, 3609–3623. 73
- MAHILLON, J. & CHANDLER, M. (1998). Insertion sequences. *Microbiology And Molecular Biology Reviews* **62**, 725–774. 10, 11, 34, 45
- MANTRI, Y. & WILLIAMS, K. (2004). Islander: a database of integrative islands in prokaryotic genomes, the associated integrases and their DNA site specificities. *Nucleic Acids Research* **32**, D55–D58. 16, 30

- MARTOGLIO, B. & DOBBERSTEIN, B. (1998). Signal sequences: more than just greasy peptides. *Trends in Cell Biology* **8**, 410–415. 63
- MARTÍNEZ, M., PALACIOS, J., IMPERIAL, J. & RUIZ-ARGÜESO, T. (2004). Symbiotic autoregulation of *nifA* expression in *Rhizobium leguminosarum* bv. *viciae*. *Journal of Bacteriology* **186**, 6586–6594. 73
- MATTHYSSE, A. & McMAHAN, S. (2001). The effect of the *Agrobacterium tumefaciens* attR mutation on attachment and root colonization differs between legumes and other dicots. *Applied and Environmental Microbiology* **67**, 1070–1075. 75
- MATTHYSSE, A., YARNALL, H., BOLES, S. & McMAHAN, S. (2000). A region of the *Agrobacterium tumefaciens* chromosome containing genes required for virulence and attachment to host cells. *Biochimica et Biophysica Acta* **1490**, 208–212. 75
- MAZANDU, G. & MULDER, N. (2012). Function prediction and analysis of *Mycobacterium tuberculosis* hypothetical proteins. *International Journal of Molecular Sciences* **13**, 7283–7302. 13, 14
- MCDONOUGH, K. & RODRIGUEZ, A. (2011). The myriad roles of cyclic AMP in microbial pathogens: from signal to sword. *Nature Reviews Microbiology* **10**, 27–38. 54
- MCGINNIS, S. & MADDEN, T. (2004). BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Research* **32**, W20–W25. 48
- MENDES, I., HUNGRIA, M. & VARGAS, M. (2004). Establishment of *Bradyrhizobium japonicum* and *B. elkanii* strains in a Brazilian cerrado oxisol. *Biology and Fertility of Soils* **40**, 28–35. 8, 71
- MITHÖFER, A. (2002). Suppression of plant defence in rhizobia–legume symbiosis. *Trends in Plant Science* **7**, 440–444. 4
- MITRA, A., KESARWANI, A., PAL, D. & NAGARAJA, V. (2010). WebGesTer DB- a transcription terminator database. *Nucleic Acids Research* **39**, D129–D135. 21, 45
- MONGIARDINI, E., AUSMEES, N., PÉREZ-GIMÉNEZ, J., JULIA ALTHABEGOITI, M., IGNACIO QUELAS, J., LÓPEZ-GARCÍA, S. & LODEIRO, A. (2008). The rhizobial adhesion protein RapA1 is involved in adsorption of rhizobia to plant roots but not in nodulation. *FEMS Microbiology Ecology* **65**, 279–288. 59

- MOODY, R. & WILLIAMSON, M. (2013). Structure and function of a bacterial Fasciclin I domain protein elucidates function of related cell adhesion proteins such as TGF β 1 and periostin. *FEBS Open Bio* **3**, 71–77. 74
- MYLONA, P., PAWLOWSKI, K. & BISSELING, T. (1995). Symbiotic nitrogen fixation. *The Plant Cell* **7**, 869–885. 6
- NAGY, Z. & CHANDLER, M. (2004). Regulation of transposition in bacteria. *Research in Microbiology* **155**, 387–398. 42
- NAN, J., BROSTROMER, E., LIU, X., KRISTENSEN, O. & SU, X. (2009). Bioinformatics and structural characterization of a hypothetical protein from *Streptococcus mutans*: Implication of antibiotic resistance. *PLoS ONE* **4**, e7245. 14
- NASSER, W. & REVERCHON, S. (2007). New insights into the regulatory mechanisms of the LuxR family of quorum sensing regulators. *Analytical and Bioanalytical Chemistry* **387**, 381–390. 72
- NELSON, M. & SADOWSKY, M. (2015). Secretion systems and signal exchange between nitrogen-fixing rhizobia and legumes. *Frontiers in Plant Science* **6**, 491. 59
- NEWELL, P., YOSHIOKA, S., HVORECNY, K., MONDS, R. & O'TOOLE, G. (2011). Systematic analysis of diguanylate cyclases that promote biofilm formation by *Pseudomonas fluorescens* Pf0-1. *Journal of Bacteriology* **193**, 4685–4698. 54
- NIEMANN, H., SCHUBERT, W. & HEINZ, D. (2004). Adhesins and invasins of pathogenic bacteria: a structural view. *Microbes and Infection* **6**, 101–112. 59
- NIGMATULLINA, L., LAVINA, A., VERSHININA, Z. & BAIMIEV, A. (2015). Role of bacterial adhesin RAPA1 in formation of efficient symbiosis of *Rhizobium leguminosarum* with bean plants. *Microbiology* **84**, 705–711. 59
- NISHI, C., BODDEY, L., VARGAS, M. & HUNGRIA, M. (1996). Morphological, physiological and genetic characterization of two new *Bradyrhizobium* strains recently recommended as Brazilian commercial inoculants for soybean. *Symbiosis* **20**, 147–162. 8
- NiÑO, M., KIM, J., LEE, H., ABDULA, S., NOU, I. & CHO, Y. (2014). Key roles of cysteine protease in different plant pathosystem. *Plant Breeding and Biotechnology* **2**, 97–109. 43

- OCHMAN, H., LAWRENCE, J. & GROISMAN, E. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304. 72
- OCHMAN, H. & MORAN, N. (2001). Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. *Science* **292**, 1096–1098. 69
- OKAZAKI, S., OKABE, S., HIGASHI, M., SHIMODA, Y., SATO, S., TABATA, S., HASHIGUCHI, M., AKASHI, R., GÖTTFERT, M. & SAEKI, K. (2010). Identification and functional analysis of type III effector proteins in *Mesorhizobium loti*. *Molecular Plant-Microbe Interactions* **23**, 223–234. 59
- OKE, V. & LONG, S. (1999a). Bacterial genes induced within the nodule during the *Rhizobium*-legume symbiosis. *Molecular Microbiology* **32**, 837–849. 74
- OKE, V. & LONG, S. (1999b). Bacteroid formation in the *Rhizobium*-legume symbiosis. *Current Opinion in Microbiology* **2**, 641–646. 74
- OKUBO, T., PIROMYOU, P., TITTABUTR, P., TEAUMROONG, N. & MINAMISAWA, K. (2016). Origin and evolution of nitrogen fixation genes on symbiosis islands and plasmid in *Bradyrhizobium*. *Microbes and Environments* **31**, 260–267. 32
- OLDROYD, G. & DOWNIE, J. (2008). Coordinating nodule morphogenesis with rhizobial infection in legumes. *Annual Review of Plant Biology* **59**, 519–546. 3
- OLDROYD, G., MURRAY, J., POOLE, P. & DOWNIE, J. (2011). The rules of engagement in the legume-rhizobial symbiosis. *Annual Review of Genetics* **45**, 119–144. 4
- OVERBEEK, R., FONSTEIN, M., D'SOUZA, M., PUSCH, G. & MALTSEV, N. (1999). The use of gene clusters to infer functional coupling. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 2896–2901. 22
- PEOPLES, M., BROCKWELL, J., HERRIDGE, D., ROCHESTER, I., ALVES, B., URQUIAGA, S., BODDEY, R., DAKORA, F., BHATTARAI, S., MASKEY, S., SAMPET, C., RERKASEM, B., KHAN, D., HAUGGAARD-NIELSEN, H. & JENSEN, E. (2009). The contributions of nitrogen-fixing crop legumes to the productivity of agricultural systems. *Symbiosis* **48**, 1–17. 3

- PERES, J., MENDES, L., SUHET, A. & VARGAS, M. (1993). Eficiência e competitividade de estirpes de rizóbio para soja em solos de cerrado. *Revista Brasileira de Ciência do Solo* **17**, 357–363. 8
- PESSI, G., AHRENS, C., REHRAUER, H., LINDEMANN, A., HAUSER, F., FISCHER, H. & HENNECKE, H. (2007). Genome-wide transcript analysis of *Bradyrhizobium japonicum* bacteroids in soybean root nodules. *Molecular Plant-Microbe Interactions* **20**, 1353–1363. 7, 22, 24, 38, 41, 45, 46, 47, 49, 52, 56, 58, 62
- PETERS, J., VANGELOFF, A. & LANDICK, R. (2011). Bacterial transcription terminators: the rna 3'-end chronicles. *Journal of Molecular Biology* **412**, 793–813. 21, 45
- PLAGUE, G. (2010). Intergenic transposable elements are not randomly distributed in bacteria. *Genome Biology and Evolution* **2**, 584–590. 37, 42
- POPP, C. & OTT, T. (2011). Regulation of signal transduction and bacterial infection during root nodule symbiosis. *Current Opinion in Plant Biology* **14**, 458–467. 54
- POWERS, D. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. *Journal of Machine Learning Technologies* **2**, 37–63. 29, 30
- PRELL, J., WHITE, J., BOURDES, A., BUNNEWELL, S., BONGAERTS, R. & POOLE, P. (2009). Legumes regulate *Rhizobium* bacteroid development and persistence by the supply of branched-chain amino acids. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 12477–12482. 3
- PRESTON, G., HAUBOLD, B. & RAINEY, P. (1998). Bacterial genomics and adaptation to life on plants: implications for the evolution of pathogenicity and symbiosis. *Current Opinion in Microbiology* **1**, 589–597. 71
- PROFT, T. & BAKER, E. (2009). Pili in Gram-negative and gram-positive bacteria - structure, assembly and their role in disease. *Cellular and Molecular Life Sciences* **66**, 613–635. 58
- PRUITT, K., TATUSOVA, T. & MAGLOTT, D. (2005). NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research* **33**, D501–D504. 20, 21
- QUESADA-VINCENS, D., HANIN, M., BROUGHTON, W. & JABBOURI, S. (1998). In vitro sulfotransferase activity of NoeE, a nodulation protein of *Rhizobium* sp. NGR234. *Molecular Plant-Microbe Interactions* **11**, 592–600.

- REES, D., AKIF TEZCAN, F., HAYNES, C., WALTON, M., ANDRADE, S., EINSLE, O. & HOWARD, J. (2005). Structural basis of biological nitrogen fixation. *Philosophical Transactions of the Royal Society A, Mathematical, Physical and Engineering Sciences* **363**, 971–984. 6
- RIBBE, M., HU, Y., HODGSON, K. & HEDMAN, B. (2014). Biosynthesis of nitrogenase metalloclusters. *Chemical Reviews* **114**, 4063–4080. 38
- RISSMAN, A., MAU, B., BIEHL, B., DARLING, A., GLASNER, J. & PERNA, N. (2009). Reordering contigs of draft genomes using the Mauve aligner. *Bioinformatics* **25**, 2071–2073. 19
- ROCHA, E. (2004). Order and disorder in bacterial genomes. *Current Opinion in Microbiology* **7**, 519–527. 15
- ROCHA, E. (2008). The organization of the bacterial genome. *Annual Review of Genetics* **42**, 211–233. 15
- RODRIGUEZ, H., SNOW, E., BHAT, U. & LOECHLER, E. (1992). An *Escherichia coli* plasmid-based, mutational system in which supF mutants are selectable: insertion elements dominate the spontaneous spectra. *Mutation Research* **270**, 219–231. 12
- ROSSBACH, S., MAI, D., CARTER, E., SAUVIAC, L., CAPELA, D., BRUAND, C. & DE BRUIJN, F. (2008). Response of *Sinorhizobium meliloti* to elevated concentrations of cadmium and zinc. *Applied and Environmental Microbiology* **74**, 4218–4221. 74
- RUAN, X. & PETERS, N. (1992). Isolation and characterization of rhizobitoxine mutants of *Bradyrhizobium japonicum*. *Journal of Bacteriology* **174**, 3467–3473. 74
- RUAN, X., ZHANG, C. & PETERS, N. (1993). *Bradyrhizobium japonicum* rhizobitoxine genes and putative enzyme functions: expression requires a translational frameshift. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 2641–2645. 74
- SALZBERG, S., DELCHER, A., KASIF, S. & WHITE, O. (1998). Microbial gene identification using interpolated markov models. *Nucleic Acids Research* **26**, 544–548. 20
- SANCHEZ, C. (2011). Bacterial pathogenesis: The importance of first impressions. *Nature Reviews Microbiology* **9**, 630–631. 59

- SANTOS, R., HÉROUART, D., PUPPO, A. & TOUATI, D. (2000). Critical protective role of bacterial superoxide dismutase in *rhizobium*-legume symbiosis. *Molecular Microbiology* **38**, 750–759. 62
- SASSO, S., RAMAKRISHNAN, C., GAMPER, M., HILVERT, D. & KAST, P. (2005). Characterization of the secreted chorismate mutase from the pathogen *Mycobacterium tuberculosis*. *The FEBS Journal* **272**, 375–389. 52
- SAUER, F., MULVEY, M., SCHILLING, J., MARTINEZ, J. & HULTGREN, S. (2000). Bacterial pili: molecular mechanisms of pathogenesis. *Current Opinion in Microbiology* **3**, 65 – 72. 58
- SCHNEIDER, C., PARISH, T., BASSO, L. & SANTOS, D. (2008). The two chorismate mutases from both *Mycobacterium tuberculosis* and *Mycobacterium smegmatis*: Biochemical analysis and limited regulation of promoter activity by aromatic amino acids. *Journal of Bacteriology* **190**, 122–134. 51
- SCHÄPER, S., KROL, E., SKOTNICKA, D., KAEVER, V., HILKER, R., SØGAARD-ANDERSEN, L. & BECKER, A. (2015). Cyclic di-GMP regulates multiple cellular functions in the symbiotic alphaproteobacterium *Sinorhizobium meliloti*. *Journal of Bacteriology* **198**, 521–535. 54
- SHAHBAAZ, M., HASSAN, M. & AHMAD, F. (2013). Functional annotation of conserved hypothetical proteins from *Haemophilus influenzae* Rd KW20. *PLoS ONE* **8**, e84263. 13, 14
- SHANNON, C. (1948). A mathematical theory of communication. *The Bell System Technical Journal* **27**, 379–423. 17
- SIGRIST, C., DE CASTRO, E. & CERUTTI, L. (2012). New and continuing developments at PROSITE. *Nucleic Acids Research* , 1–4. 14
- SIGUIER, P., FILÉE, J. & CHANDLER, M. (2006a). Insertion sequences in prokaryotic genomes. *Current Opinion in Microbiology* **9**, 526–531. 10, 11, 33
- SIGUIER, P., GOURBEYRE, E. & CHANDLER, M. (2014). Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiology Reviews* **38**, 865–891. 10, 11, 20, 34, 35
- SIGUIER, P., GOURBEYRE, E., VARANI, A., TON-HOANG, B. & CHANDLER, M. (2015). Everyman's guide to bacterial insertion sequences. *Microbiology Spectrum* **3**, MDNA3–0030–2014. 34, 35, 45

- SIGUIER, P., PEROCHON, J., LESTRADE, L., MAHILLON, J. & CHANDLER, M. (2006b). ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Research* **34**, D32–D36. 20
- SIGUIER, P., VARANI, A., PEROCHON, J. & CHANDLER, M. (2012). Exploring bacterial insertion sequences with ISfinder: objectives, uses, and future developments. *Methods in Molecular Biology* **859**, 91–103. 20
- SILLITOE, I., CUFF, A., DESSAILLY, B., DAWSON, N., FURNHAM, N., LEE, D., LEES, J., LEWIS, T., STUDER, R., RENTZSCH, R., YEATS, C., THORNTON, J. & ORENGO, C. (2013). New functional families (FunFams) in CATH to improve the mapping of conserved functional sites to 3D structures. *Nucleic Acids Research* **41**, D490–D498. 14
- SIMSER, J., RAHMAN, M., DREHER-LESNICK, S. & AZAD, A. (2005). A novel and naturally occurring transposon, ISRpe1 in the *Rickettsia peacockii* genome disrupting the *rickA* gene involved in actin-based motility. *Molecular Microbiology* **58**, 71–79. 12
- SIQUEIRA, A., ORMEÑO-ORRILLO, E., SOUZA, R., RODRIGUES, E., ALMEIDA, L., BARCELLOS, F., BATISTA, J., NAKATANI, A., MARTÍNEZ-ROMERO, E., VASCONCELOS, A. & HUNGRIA, M. (2014). Comparative genomics of *Bradyrhizobium japonicum* CPAC 15 and *Bradyrhizobium diazoefficiens* CPAC 7: elite model strains for understanding symbiotic performance with soybean. *BMC Genomics* **15**, 420. 9, 13, 16, 19, 22, 32, 46, 48, 63, 71, 72, 73, 74, 75, 76, 77
- SKORSKI, P., PROUX, F., CHERAITI, C., DREYFUS, M. & HERMANN-LE DENMAT, S. (2007). The deleterious effect of an insertion sequence removing the last twenty percent of the essential *Escherichia coli* *rpsA* gene is due to mRNA destabilization, not protein truncation. *Journal of Bacteriology* **189**, 6205–6212. 41
- SMIT, G., LOGMAN, T., BOERRIGTER, M., KIJNE, J. & LUGTENBERG, B. (1989). Purification and partial characterization of the *Rhizobium leguminosarum* biovar *viciae* Ca²⁺-dependent adhesin, which mediates the first step in attachment of cells of the family *Rhizobiaceae* to plant root hair tips. *Journal of Bacteriology* **171**, 4054–4062. 59
- SOTO, M., DOMÍNGUEZ-FERRERAS, A., PÉREZ-MENDOZA, D., SANJUÁN, J. & OLIVARES, J. (2009). Mutualism versus pathogenesis: the give-and-take in plant-bacteria interactions. *Cellular Microbiology* **11**, 381–388. 7, 57, 58, 59

- SOTO, M., SANJUÁN, J. & OLIVARES, J. (2006). Rhizobia and plant-pathogenic bacteria: common infection weapons. *Microbiology* **152**, 167–174. 7, 58, 59, 72
- SPAINK, H. (2000). Root nodulation and infection factors produced by rhizobial bacteria. *Annual Review of Microbiology* **54**, 257–288. 58
- STOUGAARD, J. (2000). Regulators and regulation of legumes root nodule development. *Plant Physiology* **124**, 531–540. 3
- SUBRAMANIAN, S., STACEY, G. & YU, O. (2006). Endogenous isoflavones are essential for the establishment of symbiosis between soybean and *Bradyrhizobium japonicum*. *The Plant Journal* **48**, 261–273. 4
- SUBRAMANIAN, S., STACEY, G. & YU, O. (2007). Distinct, crucial roles of flavonoids during legume nodulation. *Trends in Plant Science* **12**, 282–285. 4, 58
- SUBRAMONI, S., GONZALEZ, J., JOHNSON, A., PÉCHY-TARR, M., ROCHAT, L., PAULSEN, I., LOPER, J., KEELAND, C. & VENTURI, V. (2011). Bacterial subfamily of LuxR regulators that respond to plant compounds. *Applied and Environmental Microbiology* **77**, 4579–4588. 72
- SUGIYAMA, A., SHITAN, N. & YAZAKI, K. (2008). Signaling from soybean roots to *Rhizobium*. *Plant Signaling & Behavior* **3**, 38–40. 4
- SULLIVAN, J. & RONSON, C. (1998). Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 5145–5149. 32
- TAKEMURA, H., HORINOCHI, S. & BEPPU, T. (1991). Novel insertion sequence IS1380 from *Acetobacter pasteurianus* is involved in loss of ethanol-oxidizing ability. *Journal of Bacteriology* **173**, 7070–7076. 45
- TATUSOVA, T., CIUFO, S., FEDOROV, B., O'NEILL, K. & TOLSTOY, I. (2014). RefSeq microbial genomes database: new representation and annotation strategy. *Nucleic Acids Research* **42**, D553–D559. 20, 21, 22
- TATUSOVA, T. & MADDEN, T. (1999). BLAST 2 sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiology Letters* **174**, 247–250. 20

- TERAKADO, J., OKAMURA, M., FUJUHARA, M., OHMORI, M. & YONEYAMA, T. (1997). Cyclic AMP in rhizobia and symbiotic nodules. *Ann Bot* **80**, 499–503. 54
- TIAN, C., GARNERONE, A., MATHIEU-DEMAZIÈRE, C., MASSON-BOIVIN, C. & BATUT, J. (2012). Plant-activated bacterial receptor adenylate cyclases modulate epidermal infection in the *Sinorhizobium meliloti*-Medicago symbiosis. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 6751–6756. 54
- TOUCHON, M. & ROCHA, E. (2007). Causes of insertion sequences abundance in prokaryotic genomes. *Molecular Biology and Evolution* **24**, 969–981. 11
- UCHIUMI, T., OHWADA, T., ITAKURA, M., MITSUI, H., NUKUI, N., DAWADI, P., KANEKO, T., TABATA, S., YOKOYAMA, T., TEJIMA, K., SAEKI, K., OMORI, H., HAYASHI, M., MAEKAWA, T., SRIPRANG, R., MUROOKA, Y., TAJIMA, S., SIMOMURA, K., NOMURA, M., SUZUKI, A., SHIMODA, Y., SIOYA, K., ABE, M. & MINAMISAWA, K. (2004). Expression islands clustered on the symbiosis island of the *Mesorhizobium loti* genome. *Journal of Bacteriology* **186**, 2439–2448. 33
- URBAN, M., PANT, R., RAGHUNATH, A., IRVINE, A., PEDRO, H. & HAMMOND-KOSACK, K. (2015). The Pathogen-Host interactions database: additions and future developments. *Nucleic Acids Research* **43**, D645–D655. 24, 49
- VAN SLUYS, M., MONTEIRO-VITORELLO, C., CAMARGO, L., MENCK, C., DA SILVA, A., FERRO, J., OLIVEIRA, M., SETUBAL, J., KITAJIMA, J. & SIMPSON, A. (2002). Comparative genomic analysis of plant-associated bacteria. *Annual Review of Phytopathology* **40**, 169–189. 59
- VARANI, A., SIGUIER, P. & GOURBEYRE, E. (2011). ISSaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. *Genome Biology* **12**, R30. 19, 20, 35
- VERNIKOS, G. & PARKHILL, J. (2006). Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the *Salmonella* pathogenicity islands. *Bioinformatics* **22**, 2196–2203. 16, 30
- WAACK, S., KELLER, O., ASPER, R., BRODAG, T., DAMM, C., FRICKE, W., SUROVCIK, K., MEINICKE, P. & MERKL, R. (2006). Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. *BMC Bioinformatics* **7**, 142. 16, 30

- WAGNER, S. (2011). Biological nitrogen fixation. *Nature Education* **3**, 15. 3
- WEI, M., YOKOYAMA, T., MINAMISAWA, K., MITSUI, H., ITAKURA, M., KANEKO, T., TABATA, S., SAEKI, K., OMORI, H., TAJIMA, S., UCHIUMI, T., ABE, M. & OHWADA, T. (2008). Soybean seed extracts preferentially express genomic loci of *Bradyrhizobium japonicum* in the initial interaction with soybean, *Glycine max* (L.) Merr. *DNA Research* **15**, 201–214. 33
- WHITELEY, C. & LEE, D. (2015). Bacterial diguanylate cyclases: structure, function and mechanism in exopolysaccharide biofilm development. *Biotechnology Advances* **33**, 124–141. 54
- WIIG, J., HU, Y., LEE, C. & RIBBE, M. (2012). Radical SAM-dependent carbon insertion into the nitrogenase M-cluster. *Science* **337**, 1672–1675. 38
- XIA, T., SONG, J., ZHAO, G., ALDRICH, H. & JENSEN, R. (1993). The *aroQ*-encoded monofunctional chorismate mutase (CM-F) protein is a periplasmic enzyme in *Erwinia herbicola*. *Journal of Bacteriology* **175**, 4729–4737. 52
- XIA, Y. (2004). Proteases in pathogenesis and plant defence. *Cellular Microbiology* **6**, 905–913. 62
- YASUTA, T., OKAZAKI, S., MITSUI, H., YUHASHI, K., EZURA, H. & MINAMISAWA, K. (2001). DNA sequence and mutational analysis of rhizobitoxine biosynthesis genes in *Bradyrhizobium elkanii*. *Applied and Environmental Microbiology* **67**, 4999–5009. 74
- YU, Y., ADAMS, D. & YANG, S. (1979). 1-aminocyclo-propane-carboxylate synthase, a key enzyme in ethylene biosynthesis. *Archives of Biochemistry and Biophysics* **198**, 280–286. 74
- YUHASHI, K., ICHIKAWA, N., EZURA, H., AKAO, S., MINAKAWA, Y., NUKUI, N., YASUTA, T. & MINAMISAWA, K. (2000). Rhizobitoxine production by *Bradyrhizobium elkanii* enhances nodulation and competitiveness on *Macroptilium atropurpureum*. *Applied and Environmental Microbiology* **66**, 2658–2663. 74
- ZAGHLOUL, L., TANG, C., CHIN, H., BEK, E., LAN, R. & TANAKA, M. (2007). The distribution of insertion sequences in the genome of *Shigella flexneri* strain 2457T. *FEMS Microbiology Letters* **277**, 197–204. 37
- ZDOBNOV, E. & APWEILER, R. (2001). InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847–848. 23

- ZERILLO, M., VAN SLUYS, M., CAMARGO, L. & MONTEIRO-VITORELLO, C. (2008). Characterization of new IS elements and studies of their dispersion in two subspecies of *Leifsonia xyli*. *BMC Microbiology* **8**, 127. 47
- ZEYTUNI, N. & ZARIVACH, R. (2012). Structural and functional discussion of the tetra-trico-peptide repeat, a protein interaction module. *Structure* **20**, 397–405. 54
- ZHANG, Z. & SAIER, M. J. (2009). A novel mechanism of transposon-mediated gene activation. *PLoS Genetics* **5**, e1000689. 12, 42
- ZHU, F., WU, R., ZHANG, H. & WU, H. (2013). Structural and biochemical analysis of a bacterial glycosyltransferase. *Methods in Molecular Biology* **1022**, 29–39. 52, 53
- ZINSER, E., SCHNEIDER, D., BLOT, M. & KOLTER, R. (2003). Bacterial evolution through the selective loss of beneficial genes. trade-offs in expression involving two loci. *Genetics* **164**, 1271–1277. 12