

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE BIOCÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

**Genômica comparativa do operon e regulon *nod* em
*Bradyrhizobium elkanii***

IVAN LUIS ZENZEN

Dissertação submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do título de Mestre em Genética e Biologia Molecular

Orientação: Prof^ª. Dr^ª. Luciane M. P. Passaglia
Porto Alegre, dezembro de 2015.

O presente estudo foi realizado no Núcleo de Microbiologia Agrícola do Departamento de Genética da Universidade Federal do Rio Grande do Sul, financiado pelo Conselho Nacional de Desenvolvimento Tecnológico (CNPq) e pelo Instituto Nacional de Ciência e Tecnologia (INCT) da Fixação Biológica de Nitrogênio.

AGRADECIMENTOS

Sem dúvidas, a realização deste trabalho conta com a contribuição de muitos que desempenharam um papel ativo durante minha formação, apoiando e oferecendo o suporte necessário, direta e indiretamente, para que o mesmo fosse concluído. O reconhecimento deste portanto, faço questão de compartilhar com todos formadores e colegas que fizeram e continuam a fazer parte desta trajetória. São muitos, entretanto, quase certo do risco de ser injusto por não mencionar alguém, não posso deixar de nominar aqueles que foram fundamentais para o êxito desta etapa.

Inicialmente, à minha orientadora e conselheira Profa. Dra. Luciane Passaglia, pelos ensinamentos, conselhos e discussões científicas que me fizeram crescer e amadurecer pessoal e profissionalmente; pela compreensão e amizade durante o período de convívio; sobretudo pelo apoio, pela confiança, motivação e enorme paciência, por acreditar no meu potencial e me proporcionar oportunidades acadêmicas altamente qualificadas.

Aos membros integrantes da banca examinadora, Prof^a. Dr^a. Adriana Giongo, Prof. Dr. Luciano do Amarante e Prof. Dr. Rogério Margis pela pronta disponibilidade em avaliar e contribuir com nosso trabalho.

Aos professores e funcionários do Programa de Pós-Graduação em Genética e Biologia Molecular da Universidade Federal do Rio Grande do Sul, pela excelência profissional e pela oportunidade de aprendizado.

Aos colegas da Pós-Graduação com quem tive a oportunidade e o prazer de dividir as risadas e as “angústias” do percurso, em especial à Edilena Sperb pelo auxílio infindável, pela parceria de trabalho, pelo companheirismo e sobretudo pela amizade sincera.

Finalmente e de forma especial à minha família, meus pais Edemir Zenzen e Ivone Zenzen, minha irmã Ana Carolina Zenzen, mais recentemente ao nosso pequeno-gigante sopro de vida, a Clarinha, por todo amor, pelo apoio incondicional e um suporte sem o qual seria impossível esta conquista.

SUMÁRIO

RESUMO	5
ABSTRACT	7
1. INTRODUÇÃO	8
1.1 Nitrogênio: um desequilíbrio necessário?	8
1.2 Fixação Biológica de Nitrogênio: o processo de simbiose e seus protagonistas	9
1.3 Circuito de genes <i>nod</i> regulatórios: o diálogo molecular na simbiose	13
1.4 Regulação da transcrição dos genes relacionados à simbiose	16
1.5 Soja x <i>Bradyrhizobium</i> : simbiose de sucesso	17
1.6 Genômica comparativa	20
2. OBJETIVOS	22
2.1 Objetivo geral	22
2.2 Objetivos específicos	22
3. MANUSCRITO	23
4. CONSIDERAÇÕES FINAIS E PERSPECTIVAS	63
5. REFERÊNCIAS BIBLIOGRÁFICAS	67

RESUMO

A Fixação Biológica de Nitrogênio (FBN), processo no qual o nitrogênio atmosférico é convertido à amônia, é muito bem estabelecida entre bactérias diazotróficas coletivamente chamadas de rizóbios e espécies leguminosas. No Brasil, esse tipo de associação (simbiótica) supre totalmente a necessidade de nitrogênio na cultura da soja. Para que a infecção seja efetiva e possa resultar na formação de um nódulo capaz de sustentar o processo de FBN conduzido pelo bacterioide, o rizóbio necessita, previamente, reconhecer e responder à presença das raízes da planta compatível. As associações simbióticas entre rizóbios e plantas leguminosas são altamente específicas, de forma que cada espécie, ou até mesmo estirpe de rizóbio, possui uma gama definida de plantas às quais está apto a se associar, e vice-versa. A principal função dos produtos dos genes de nodulação (*nod*) é garantir a troca de sinais entre os dois organismos envolvidos na relação simbiótica, onde os produtos dos genes *nod* regulatórios atuam no controle da expressão de genes *nod* estruturais. A expressão de genes *nod* estruturais, via de regra, não ocorre de forma autônoma nos micro-organismos simbiontes do gênero *Bradyrhizobium*, requerendo assim a presença de moléculas sinalizadoras secretadas pelas raízes das plantas (predominantemente flavonoides) e ativadores transcricionais do tipo-LysR – as proteínas NodD regulatórias. Neste contexto, motivos específicos das proteínas NodD ligam-se a sequências conservadas na região promotora do operon *nod*, conhecidas como *nod boxes*, mediando a transcrição dos genes *nod*. Aparentemente, este sistema regulatório envolvendo as proteínas NodD está presente na maioria das estirpes de *Rhizobium*, *Bradyrhizobium* e *Azorhizobium*, sugerindo um mecanismo geral de controle da nodulação. Em *B. diazoefficiens*, duas proteínas NodD foram identificadas, com funções e padrões de expressão distintos: NodD₁, ativador da transcrição dos genes *nod* responsivo aos flavonoides liberados pelas raízes das plantas, e NodD₂, com ação contrária, atuando como repressor da transcrição desses genes. Enquanto existe uma quantidade relativamente grande de conhecimento em relação à genética e aos mecanismos moleculares que regulam a expressão dos genes envolvidos na nodulação de *B. diazoefficiens*, incluindo a sequência completa de seu genoma, informações sobre a genética de *B. elkanii* ainda são relativamente escassas, mesmo que existam alguns dados genômicos disponíveis. Neste trabalho, sequências genômicas de seis linhagens de *B. elkanii* foram comparadas com o genoma da linhagem de referência *B. diazoefficiens* USDA 110 com o objetivo de elucidar

mecanismos envolvidos na expressão dos genes *nod*, especialmente aqueles relacionados ao operon e ao regulon *nod*. Os resultados obtidos permitiram acrescentar aspectos importantes no modelo de regulação apresentado para a linhagem de referência e que pode ser estendido para linhagens de *B. elkanii*.

ABSTRACT

The Biological Nitrogen Fixation (BNF), process in which atmospheric nitrogen is converted to ammonia, is well established among diazotrophs collectively called rhizobia and legume species. In Brazil, this type of symbiotic association fully meets the need for nitrogen in soybean crop. To infection be effective and result in the formation of a nodule able to sustain the BNF process lead by the bacterioid, the rhizobia need previously to recognize and respond to the presence of the root of a compatible plant. The symbiotic associations between rhizobia and leguminous plants are highly specific, so that each species or even strain of rhizobia has a defined range of plants to which it is able to associate, and *vice-versa*. The main function of the products of nodulation (*nod*) genes is to guarantee the exchange of signals between the two organisms involved in the symbiotic relationship, where the products of regulatory *nod* genes act to control the expression of structural *nod* genes. The expression of structural *nod* genes usually does not occur independently in the symbiotic microorganisms of the *Bradyrhizobium* genus, thus requiring the presence of signalling molecules secreted by plant roots (predominantly flavonoids) and transcriptional activators – the LysR-type regulatory NodD proteins. In this context, NodD proteins bind to specific motifs of conserved sequences in the promoter region of the *nod* operon, known as *nod* boxes, mediating transcription of the *nod* genes. Apparently, this regulatory system involving NodD proteins is present in most strains of *Rhizobium*, *Bradyrhizobium* and *Azorhizobium*, suggesting a general mechanism of nodulation control. In *B. diazoefficiens* two NodD proteins were identified with distinct functions and expression patterns: NodD₁, a flavonoid responsive transcriptional activator of *nod* genes, and NodD₂ with counteraction, acting as a transcriptional repressor of these genes. While there is a relatively large amount of knowledge about the genetics and molecular mechanisms that regulate the expression of genes involved in *B. diazoefficiens* nodulation, including the complete sequence of its genome, genetic information concerning *B. elkanii* is still relatively sparse, even if there are some available genomic data. In this study, genomic sequences of six strains of *B. elkanii* were compared with the genome of the reference strain *B. diazoefficiens* USDA 110 in order to elucidate mechanisms involved in the expression of *nod* genes, especially those related to the *nod* operon and the *nod* regulon. The results obtained allowed to add important aspects in the regulatory model presented to the reference strain and that could be extended to strains of *B. elkanii*.

1 INTRODUÇÃO

1.1 Nitrogênio: um desequilíbrio necessário?

O nitrogênio (N) é considerado nutriente mineral essencial aos seres vivos, sendo constituinte primário de biomoléculas indispensáveis à manutenção da vida como ácidos nucleicos, determinantes dos aspectos genéticos, e proteínas, responsáveis pelo metabolismo celular. Apesar da abundância na natureza, perfazendo aproximadamente 78% do volume total da atmosfera terrestre, este elemento existe como gás diatômico, também conhecido como dinitrogênio (N_2) ou nitrogênio molecular, uma substância relativamente inerte e biologicamente indisponível para a grande maioria dos organismos. Neste contexto, sua utilização fica previamente sujeita à quebra da ligação tripla entre os átomos de N e a subsequente ligação dos átomos resultantes com um ou mais elementos essenciais: (i) oxigênio e/ou hidrogênio, pelo processo de fixação, ou (ii) carbono, por meio da assimilação de N (Galloway, 1998, Galloway *et al.*, 2002).

A dinâmica e produtividade de grande parte dos ecossistemas marinhos e terrestres não-manejados, bem como ecossistemas agrícolas e florestais sob manejo, são limitadas pelo suprimento de nitrogênio em uma forma reativa ou biologicamente disponível (Vitousek *et al.*, 1997). O processo físico-químico de fixação de N, também conhecido como processo de Haber-Bosch (Haber, 2002), possibilitou a produção de fertilizantes nitrogenados em escala industrial, promovendo uma rápida ascensão do volume de N reativo no ambiente, computados em mais de 113 Tg para o ano de 2015 (Smil, 1999, FAO, 2011). Com a adoção de práticas de fertilização mineral em várias regiões do globo, a produtividade das principais culturas agrícolas registrou incrementos consideráveis, estimando-se aumento de 1,9 para 4,3 na capacidade de suporte de habitantes por hectare de terra arável entre 1908 e 2008 (Erisman *et al.*, 2008), e provendo assim, sustento a um adicional de 5,4 bilhões na população mundial ao longo do mesmo período (Smil, 1999).

Embora o enriquecimento de N mineral na agricultura demonstre consistência no suporte às atividades, acarretando no aumento de produtividade e acúmulo substancial de biomassa em curto e médio prazos, a ampliação na disponibilidade de N reativo nos ecossistemas por meio do processo de fixação industrial apresenta várias questões negativas. A absorção de carbono na biosfera e posterior implicância desta no sistema climático, por exemplo, é fortemente influenciada pela ciclagem de nutrientes como o N.

Em sistemas globais acoplados, alterações neste componente biogeoquímico terrestre podem alterar o ciclo global de carbono (C), afetando tanto o aumento na taxa de dióxido de carbono na atmosfera quanto a resposta dos ecossistemas a este aumento. Em última instância, alterações fundamentais no comportamento de mecanismos de *feedback* críticos que operam entre a biosfera terrestre e o sistema climático global podem ser observadas, eventualmente ocasionando redução da biodiversidade (Thornton *et al.*, 2007, Schimel *et al.*, 2015). Adicionalmente, a acidificação e eutrofização de ecossistemas terrestres e aquáticos, além da poluição atmosférica, decorrentes da produção e utilização excessiva de N mineral, impactam direta e/ou indiretamente a saúde humana (Cowling *et al.*, 2002, Swanston *et al.*, 2004, Trentman *et al.*, 2015); estes, por sua vez, aliados a fatores de ordem econômica, como o elevado custo energético para a síntese de amônia pelo processo Haber-Bosch, ressurgem a discussão quanto à sustentabilidade desta prática (Smil, 1999, Galloway *et al.*, 2004).

1.2 Fixação Biológica de Nitrogênio: o processo de simbiose e seus protagonistas

O acúmulo de evidências científicas, relacionando os atuais padrões de fertilização nitrogenada nos principais sistemas de cultivo a diversos efeitos adversos, tem enfatizado a importância do desenvolvimento e da adoção de estratégias produtivas mais sustentáveis. Entretanto, o grande desafio destas práticas, e em especial, a redução ou substituição da fertilização mineral, reside na conciliação de atividades ambientalmente coerentes capazes de produzir alimentos seguros para o consumo humano, de forma rentável e sem comprometer a produtividade dos cultivos (1994). Sob condições naturais, o N reativo é majoritariamente oriundo do processo de fixação biológica de nitrogênio (FBN), responsável por um aporte anual estimado entre 90-130 Tg de N em ecossistemas terrestres, e 40-200 Tg de N em ecossistemas marinhos, além da precipitação atmosférica em função de descargas elétricas, cuja contribuição, mais modesta, é na ordem de 3-10 Tg de N por ano (Galloway, 1998, Galloway *et al.*, 2008).

A conversão de N_2 à amônia ou outras formas reativas de N é considerado o segundo processo de maior relevância biológica após a fotossíntese (Sylvia, 2005), tendo a FBN como principal mecanismo e responsável por um montante total de N fixado equivalente à quantidade global de amônia produzida pela indústria química de fertilizantes (Denarie *et al.*, 1993). O processo de FBN é essencialmente catalisado pelo

complexo enzimático da nitrogenase e é realizado por um grupo especializado de procariotos, chamados de diazotróficos. Esse grupo inclui bactérias de vida livre, cianobactérias, e, sobretudo, endossimbiontes da classe alfa-proteobactéria pertencentes à família Rhizobiaceae, mais especificamente aos gêneros *Rhizobium*, *Mesorhizobium*, *Ensifer*, *Bradyrhizobium* e *Azorhizobium*, coletivamente conhecidos como rizóbios (Gage, 2004, Carrareto Alves *et al.*, 2014).

O complexo enzimático da nitrogenase é composto de duas subunidades: dinitrogenase redutase (produto do gene *nifH*), também chamada de Fe-proteína, e dinitrogenase (produto dos genes *nifDK*), também chamada de MoFe-proteína (Rangaraj *et al.*, 2001, Dixon *et al.*, 2004). Durante a catálise, a dinitrogenase é reduzida pela dinitrogenase redutase, um elétron por vez, até a quantidade suficiente de elétrons (em geral oito) ser acumulada para subsequente redução do substrato (Dixon *et al.*, 2004). Esse processo demanda um aporte considerável de energia: pelo menos 16 moléculas de ATP são necessárias para cada molécula de N₂ reduzida. Consequentemente, existem diversos níveis de controle na expressão dos genes envolvidos, além de mecanismos de controle em nível posterior à transcrição (Fu *et al.*, 1989, Halbleib *et al.*, 2000, Rangaraj *et al.*, 2001, Dixon *et al.*, 2004).

Os genes envolvidos no processo de FBN, ou genes *nif* (de *nitrogen fixation*), foram primeiramente caracterizados na bactéria diazotrófica de vida livre *Klebsiella pneumoniae*, e em sua maioria estão envolvidos na montagem e estabilização do complexo enzimático da nitrogenase. Existe uma grande similaridade com relação à estrutura, organização e função dos genes *nif* nos diversos organismos estudados quando comparados à *K. pneumoniae*. Nesta bactéria, os genes estão localizados em uma região de 24.000 pares de base, organizados em oito operons (Arnold *et al.*, 1988). Os genes estruturais da nitrogenase estão localizados no mesmo operon (*nifHDK*). Muitos dos outros genes *nif* estão envolvidos na síntese do cofator da enzima, FeMo-co, entre eles *nifE*, *nifN*, *nifV*, *nifB* e *nifQ*. Existem ainda genes envolvidos no transporte de elétrons, como os genes *nifJ* e *nifE*, e genes envolvidos na regulação da expressão dos demais genes *nif*, como *nifA* e *nifL* (Dixon, 1984, Arnold *et al.*, 1988).

Os rizóbios possuem a habilidade particular de estabelecer associações simbióticas com determinadas espécies vegetais da família Leguminosae, no intuito de reduzir o N₂ à amônia (Sadowsky *et al.*, 2013). Posteriormente, o N fixado pelo micro-organismo é

disponibilizado à planta em troca de uma fonte carbono fixado pelo vegetal, de forma que ambos os organismos se beneficiem desta relação (Desbrosses *et al.*, 2011). O processo de simbiose entre rizóbios e plantas leguminosas culmina com a formação de órgãos específicos elicitados pelo procarioto sobretudo nas raízes do vegetal, os nódulos, onde ocorre a FBN. Como o complexo enzimático da nitrogenase é irreversivelmente inativado pelo oxigênio, há necessidade de manutenção deste em condições microaeróbicas, próximas à anoxia. Este ambiente é viabilizado pelo sequestro da maquinaria metabólica no interior do nódulo: um conjunto de células diferenciadas com características morfológicas e bioquímicas que limitam a exposição do complexo da nitrogenase ao oxigênio (Gage, 2004).

Apesar das particularidades conferidas pelo amplo aparato bioquímico e molecular utilizado pelos rizóbios para infectar legumes, a nodulação (i.e. a formação do nódulo) apresenta uma sequência de eventos comuns ao processo, resultando na sincronização de dois processos altamente coordenados que ocorrem paralelamente: (i) o re-início da divisão celular no córtex da raiz, levando à formação do primórdio do nódulo; e (ii) o processo de infecção bacteriana, que tem como alvo tal primórdio (Madsen *et al.*, 2010). Nos estádios iniciais da simbiose, os rizóbios necessitam deslocar-se da superfície até o interior do tecido radicular, onde colonizarão as células do primórdio do nódulo. Em nível celular, esta infecção inicia com o aprisionamento da bactéria entre as paredes celulares na extremidade de um pelo radicular deformado, geralmente apresentando uma curvatura acentuada ou “enrolamento” induzido pelo próprio micro-organismo (McCoy, 1932). Em seguida, a secreção de enzimas hidrolíticas pelo simbionte promove a degradação local da parede celular e a subsequente invaginação da membrana plasmática das células vegetais, dando origem a uma estrutura tubular que atravessa o pelo em direção ao córtex radicular, o chamado cordão de infecção (Callaham *et al.*, 1981, Madsen *et al.*, 2010). Ao atingir o primórdio do nódulo, o cordão de infecção ramifica-se e a crescente população de rizóbios contida ao longo desta estrutura é liberada nas células vegetais, passando por um processo de endocitose preliminar pelo qual as células procarióticas são subcompartimentalizadas por uma membrana peribacterioide, essencial ao estabelecimento funcional da endossimbiose (Verma, 1992, Verma *et al.*, 1993). A continuidade dos processos de divisão celular e concomitante infecção dão origem ao nódulo propriamente dito.

Entretanto, para que a infecção seja efetiva e possa resultar na formação de um nódulo capaz de sustentar o processo de FBN conduzido pelo bacteroide, previamente o rizóbio necessita reconhecer e responder à presença das raízes da planta compatível. As associações simbióticas entre rizóbios e plantas leguminosas são altamente específicas, de forma que cada espécie, ou até mesmo estirpe de rizóbio, possui uma gama definida de plantas às quais está apto a se associar, e vice-versa (Desbrosses *et al.*, 2011, Wang *et al.*, 2012). Além do programa genético governado pelo vegetal, um número considerável de genes procarióticos está envolvido no controle da especificidade da interação, infecção e nodulação.

Os genes *nod*, por exemplo, são classificados em duas categorias: genes *nod* estruturais e genes *nod* regulatórios (Tabela 1). A principal função dos produtos dos genes *nod* é garantir a troca de sinais entre os dois organismos envolvidos na relação simbiótica, onde os produtos dos genes *nod* regulatórios atuam no controle da expressão de genes *nod* estruturais. Em um primeiro momento, os produtos dos genes *nod* regulatórios estão envolvidos na detecção de moléculas sinalizadoras provenientes da planta, presentes nos exudatos radiculares, denominadas de flavonoides. Os produtos dos genes *nod* regulatórios ativam, então, a expressão dos genes *nod* estruturais, determinando a produção de lipoquitooligosacarídeos (LCOs) por parte da bactéria, conhecidos como fatores Nod. Por fim, estas moléculas são secretadas no meio e atuam em uma via de sinalização de *feedback* para com o vegetal (Denarie *et al.*, 1993).

Além disso, vários outros compostos liberados pelas plantas (betaínas, xantonas, ácidos, jasmonatos e compostos fenólicos simples) podem atuar como indutores dos genes de nodulação de rizóbios (genes *nod*); fito-hormônios (auxinas, citocininas e etileno), bem como peróxido de hidrogênio (H₂O₂) e ácido nítrico (NO) também exercem papel na simbiose (Janczarek *et al.*, 2015). De fato, a bactéria exerce um impacto essencial no estabelecimento de uma simbiose efetiva com a leguminosa compatível. Além do principal sinal do rizóbio – os fatores Nod – muitos componentes de superfície e proteínas estão envolvidos em várias etapas das interações simbióticas. Estes incluem polissacarídeos complexos localizados na membrana externa ou acumulados na superfície das células bacterianas, incluindo, por exemplo, lipopolissacarídeos, polissacarídeos antígeno-K, polissacarídeo extracelular, polissacarídeo neutro e polissacarídeo formador de gel (Frayse *et al.*, 2003). Outros compostos bacterianos como N-acil-homoserina-lactonas,

opanoídes, luminocromo, bradióxetina, ácido indol-acético, além de proteínas específicas, principalmente as secretadas pelos sistemas de secreção do tipo I e III, também participam na sinalização simbiótica (Janczarek *et al.*, 2015).

Tabela 1. Relação dos genes *nod* identificados em *Bradyrhizobium diazoefficiens* USDA 110. Genes *nod* regulatórios encontram-se em negrito; todos os demais representam genes *nod* estruturais.

Gene	Anotação		Ontologia genética	Tamanho	
	ID	Produto		Transcrito (nt)	Peptídeo (aa)
<i>nodW</i>	bll1714	regulador sistema de dois componentes	GO:0006355	684	227
<i>nodV</i>	bll1715	regulador sistema de dois componentes	GO:0018106	2670	889
<i>nodY</i>	bll2016	ND*	ND	729	242
<i>nodZ</i>	bsl2015	ND	ND	228	75
<i>nodA</i>	bll2019	regulador transcricional tipo-MerR	GO:0006355	573	190
<i>nodD₂</i>	bll2021	regulador transcricional tipo-LysR	GO:0006355	993	330
<i>nodD₁</i>	bll2023	regulador transcricional tipo-LysR	GO:0006355	945	314
<i>nodY</i>	blr2024	ND	ND	639	212
<i>nodA</i>	blr2025	acil-transferase	GO:0008415	633	210
<i>nodB</i>	blr2026	quito-oligossacarídeo desacetilase	GO:0005975	456	151
<i>nodC</i>	blr2027	quitina-sintase	GO:0016758	1458	485
<i>nodS</i>	blr2028	N-metil-transferase	GO:0009877	528	175
<i>nodU</i>	blr2029	6-O-carbamoyl transferase	GO:0009058	1710	569
<i>nodI</i>	blr2030	transportador fatores Nod	GO:0009877	921	306
<i>nodJ</i>	blr6578	transportador ABC / permease	GO:0006810	789	262
<i>nodM</i>	bsr2032	ND	ND	174	57
<i>nodN</i>	blr2033	ND	ND	396	131
<i>nodO</i>	blr2034	carbamoyl transferase	GO:0009058	1608	535
<i>nodZ</i>	Blr2035	hexosil-transferase	GO:0009877	1113	370

*ND = não determinado/confirmado

1.3 Circuito de genes *nod* regulatórios: o diálogo molecular na simbiose

A expressão de genes *nod* estruturais não ocorre de forma autônoma nos micro-organismos simbiontes, requerendo, assim, a presença de moléculas sinalizadoras secretadas pelas raízes das plantas (predominantemente os flavonoides) e ativadores

transcricionais do tipo-LysR próprios - as proteínas NodD regulatórias. Neste contexto, motivos específicos das proteínas NodD ligam-se às sequências conservadas na região promotora dos operons *nod*, conhecidas como *nod boxes* (Goethals *et al.*, 1992), mediando a transcrição dos genes de nodulação (*nod*). Aparentemente, este sistema regulatório envolvendo as proteínas NodD está presente na maioria das estirpes de *Rhizobium*, *Bradyrhizobium* e *Azorhizobium*, sugerindo um mecanismo geral de controle da nodulação (Denarie *et al.*, 1993).

A especificidade da interação simbiótica entre a planta leguminosa e a estirpe de rizóbio resulta em parte da regulação dos genes *nod* dependente de NodD em resposta aos coquetéis de flavonoides exudados pelas raízes de cada planta. Em outros termos, a expressão de genes *nod* é elicitada em função de diferentes combinações de flavonoides de forma altamente específica, mediada pela(s) proteína(s) NodD do micro-organismo envolvido (Denarie *et al.*, 1992, Peck *et al.*, 2006). Consistentemente, estirpes de rizóbios, cuja associação restringe-se a um menor número de plantas, apresentam ativação de NodD por um grupo limitado de moléculas sinalizadoras, enquanto estirpes de rizóbios com NodD ativada por um amplo espectro de moléculas demonstram maior gama de espécies vegetais com as quais são capazes de estabelecer a relação de simbiose (Györgypal *et al.*, 1991).

Apesar da falta de evidências bioquímicas diretas, aspectos genéticos suportam a hipótese de uma interação direta de NodD na percepção de flavonoides. Inicialmente, a presença de uma cópia funcional do gene *nodD* demonstra-se necessária e suficiente para induzir a expressão de genes *nod* estruturais na presença de flavonoides (Mulligan *et al.*, 1985), enquanto mutações pontuais em NodD são capazes de alterar esta expressão em resposta a diferentes moléculas sinalizadoras (Burn *et al.*, 1987). A construção de quimeras de NodD, pela transferência de determinados genes *nodD* para diferentes espécies/estirpes de rizóbios, também altera a sensibilidade a determinadas combinações de flavonoides quanto à transcrição de genes *nod* estruturais, indicando uma percepção potencial do espectro de moléculas indutoras pelas distintas proteínas NodD (Spaink *et al.*, 1987). Finalmente, a localização subcelular sugere que NodD encontra-se primariamente associada à porção interna da monocamada, na membrana plasmática bacteriana (Schlaman *et al.*, 1989), coincidindo com a região de maior acúmulo de flavonoides, onde possivelmente ocorra a interação de NodD com estas moléculas (Recourt *et al.*, 1989).

Subsequente à percepção das moléculas sinalizadoras e consequente “ativação”, NodD liga-se aos respectivos *nod boxes* com maior afinidade, sugerindo a existência de uma forma solúvel passível de translocação da membrana para o citosol sob condições de elevada concentração da proteína (Schlaman *et al.*, 1989). Sendo o produto gênico de *nodD* a única proteína celular solúvel capaz de se ligar especificamente às sequências dos *nod boxes*, o envolvimento de uma proteína intermediária mediando a transferência da informação parece pouco plausível (Fisher *et al.*, 1988). Adicionalmente, a região N-terminal de NodD apresenta o motivo de ligação ao DNA *helix-turn-helix* (H-T-H) homólogo àquele presente em ativadores transcricionais do tipo-LysR (Henikoff *et al.*, 1988), alocando, assim, a proteína na mesma família. Apesar de uma descrição inicial dos *nod boxes* como sequências altamente conservadas de aproximadamente 47-49 pares de bases, estudos posteriores envolvendo regiões promotoras dos genes *nod* em espécies dos gêneros *Bradyrhizobium* e *Azorhizobium* constataram menor grau de conservação dos *nod boxes*, revelando a existência de ao menos duas sequências palíndromes invertidas contendo o motivo ATC-N₉-GAT, às quais NodD₁ liga-se de forma tetramérica, induzindo a expressão dos respectivos genes *nod* estruturais (Wang *et al.*, 1991, Goethals *et al.*, 1992).

O primeiro passo na formação dos fatores Nod se dá a partir do produto do gene *nodC*, uma N-acetil-glucosamil-transferase. Essa enzima sintetiza um quitoooligossacarídeo, alongando a cadeia em terminações não reduzidas. A desacetilase NodB (produto de *nodB*) remove o N-acetil da extremidade dos oligossacarídeos da N-acetilglucosamina, e, finalmente, uma acetiltransferase, codificada por *nodA*, liga a cadeia acil ao acetil livre da extremidade do oligossacarídeo (Mergaert *et al.*, 1997, Perret *et al.*, 2000). As proteínas NodI e NodJ, apesar de não serem essenciais para a formação dos fatores Nod, possivelmente estejam envolvidas na exportação destes (Perret *et al.*, 2000). As proteínas NodABC são suficientes para a indução da nodulação na maioria das plantas leguminosas hospedeiras, uma vez que formam a cadeia principal dos fatores Nod. Os outros componentes Nod, no entanto, possuem papel mais sutil na nodulação, podendo ampliar a gama de plantas hospedeiras, ou mesmo agir na proteção da degradação dos fatores Nod (Perret *et al.*, 2000). Entre os genes cujos produtos estão envolvidos nesses processos estão *nodEF*, responsáveis por adicionar ácidos graxos à cadeia principal; *noeC*, responsável pela arabinosilação; *nodZ* e *nolK*, responsáveis pela fucosilação; *nodH* e *noeE*,

responsáveis pela sulfatação; *nodL*, *nodX* e *nolL*, responsáveis pela acetilação; *nodS*, *nodU* e *nolO*, responsáveis pela N metilação e carbomilação, e *noeI*, responsável pela 2-O-metilação (Carlson *et al.*, 1993, Mergaert *et al.*, 1997, Perret *et al.*, 2000, Geurts *et al.*, 2002).

1.4 Regulação da transcrição dos genes relacionados à simbiose

Os genes de nodulação de rizóbios, via de regra, estão organizados em operons, e estes podem estar localizados no cromossomo ou em grandes plasmídeos, dependendo da espécie (Chen *et al.*, 2005). Conforme mencionado, genes *nod* estruturais são regulados pelas proteínas NodD (produto dos genes *nodD*), que pertence à família de ativadores transcricionais do tipo LysR em bactérias. De acordo, a proteína NodD, uma vez em contato com o flavonóide específico exudado pela planta hospedeira, sofre uma mudança conformacional e liga-se aos *nod boxes* situados nas regiões promotoras dos genes *nod*, *nol* e *noe* (Stacey, 1995, Viprey *et al.*, 1998, Chen *et al.*, 2005, Bais *et al.*, 2006). A expressão dos genes *nod* por sua vez, leva à síntese dos fatores Nod (Perret *et al.*, 2000, Bais *et al.*, 2006).

Em *B. diazoefficiens*, duas proteínas NodD foram identificadas, com funções e padrões de expressão distintos: NodD₁, ativador da transcrição dos genes *nod* responsivo aos flavonoides liberados pelas raízes das plantas (à exemplo genisteína e daidzeína); NodD₂, com ação contrária, atuando como repressor da transcrição desses genes. Um mecanismo similar de ação destas proteínas também ocorre em *Rhizobium* sp.NGR234 (Fellay *et al.*, 1998, Loh *et al.*, 2001, Loh *et al.*, 2003). Curiosamente, além da indução por flavonoides como genisteína e daidzeína, NodD₁ em *B. diazoefficiens* apresenta auto-regulação, um mecanismo distinto daquele adotado em *Rhizobium* sp. cuja expressão é constitutiva (Loh *et al.*, 2003).

Apesar da relevância de NodD₁ para expressão dos genes *nod*, mutantes de *B. diazoefficiens* defectivos nesta proteína ainda são capazes de nodular suas plantas hospedeiras, mesmo que com menor eficiência quando comparados à bactéria de tipo selvagem (Sanjuan *et al.*, 1994, Loh *et al.*, 1997). Aparentemente contraditório, este paradoxo foi posteriormente resolvido com a descoberta de um sistema regulatório adicional, NodVW, também envolvido na ativação de genes *nod* via isoflavonas (Sanjuan *et al.*, 1994). O sistema NodVW é especialmente importante na formação de nódulos em

plantas hospedeiras alternativas, à exemplo *Vigna unguiculata*, *V. radiata* e *Macroptilium atropurpureum* (Göttfert *et al.*, 1992, Sanjuan *et al.*, 1994), entretanto não é requerido para a nodulação de soja (Loh *et al.*, 1997). A análise genômica de *nodVW* indica que ambos produtos pertencem à família de proteínas reguladoras do tipo dois-componentes (*two-component system*). Membros desta família são agrupados em duas subclasses: sensores e reguladores. Após a detecção do estímulo ambiental correto, o sensor transduz o sinal para o regulador, que por sua vez, controla a expressão dos genes alvo em um processo mediado por diversos passos de fosforilação. Ao receber o estímulo, o domínio quinase do sensor é autofosforilado em um resíduo de histidina conservado. Esta quinase ativada transfere então o grupamento fosforil para um resíduo de aspartato no domínio N-terminal da proteína reguladora, promovendo sua ativação. No caso das proteínas NodVW, sugere-se que NodV funcione como o sensor e NodW como o regulador, controlando a transcrição dos genes *nod* em *B. diazoefficiens* (Sanjuan *et al.*, 1994, Loh *et al.*, 1997).

Adicionalmente, verifica-se que NolA também é essencial para nodulação efetiva de uma planta hospedeira por *B. diazoefficiens*, atuando como um dos dois reguladores negativos da transcrição de genes de nodulação, juntamente com a proteína NodD₂ (Loh *et al.*, 2001, Krause *et al.*, 2002). Diferentemente dos genes de nodulação, *nolA* não é induzido por isoflavonas, com evidências indicando que sua transcrição possa ser induzida em resposta a quitooligossacarídeos presentes nas células (Loh *et al.*, 1999). Além disso, a presença de quitooligossacarídeos pode inibir a expressão dos genes de nodulação que governam a sua própria síntese. Esse mecanismo de *feedback* ocorre com a participação de NolA e NodD₂ (Loh & Stacey, 2001).

1.5 Soja x *Bradyrhizobium*: simbiose de sucesso

A soja [*Glycine max* (L.) Merr.] representa a espécie de maior interesse comercial dentre as leguminosas e culturas oleaginosas, sendo a mais importante fonte de proteína vegetal utilizada no mundo. Apresentando teor proteico em média de 40%, correspondente a 6,5% de N, é evidente a demanda deste elemento pela cultura (Hymowitz *et al.*, 1974, Bonato *et al.*, 2000). Em termos práticos, estima-se que sejam necessários aproximadamente 80 kg de N para a produção de 1.000 kg de grãos de soja, incluindo neste montante as quantidades alocadas em folhas, caules e raízes. Com os índices de produtividade atualmente acima de 3.000 kg de grãos por hectare (ha) e em contínua

ascensão, presume-se a necessidade de um aporte superior a 240 kg de N ha⁻¹ para a cultura, o qual pode ser suprido na totalidade pela FBN, consequência da associação simbiótica desta espécie vegetal com bactérias do gênero *Bradyrhizobium* (Hungria *et al.*, 2001).

A FBN desempenhou e continua a desempenhar papel estratégico crucial no sucesso da cultura da soja em termos globais, atenuando ou, em muitos casos, eliminando a prática de fertilização mineral de N. Como consequência, os custos e energia empregados neste sistema produtivo são consideravelmente inferiores quando comparados aos demais sistemas culturais que demandam fertilizantes nitrogenados comerciais (Hardy *et al.*, 1971, Pimentel, 2009). Com o objetivo de otimizar o processo de FBN, cultivares de soja e estirpes de *Bradyrhizobium* são constantemente avaliadas e selecionadas de forma a garantir máxima atividade biológica nas diferentes zonas de cultivo da leguminosa. Diversas condições ambientais podem atuar como fatores limitantes ao crescimento e atividade das espécies de rizóbios presentes no solo, afetando a diversidade e tamanho das populações das mesmas. A presença da planta hospedeira com a qual a bactéria é capaz de estabelecer a relação de simbiose, o histórico de cultivo da área, condições extremas de pH e práticas de fertilização mineral, dentre outros, são alguns aspectos que podem moldar o habitat destes micro-organismos, seja no solo em si ou na rizosfera (Brockwell *et al.*, 1991, Lima *et al.*, 1998, Parker, 1999, Andrade *et al.*, 2002, Sharma *et al.*, 2005). Tais fatores, normalmente atuando de forma combinada em condições de campo, possuem a capacidade de modificar a composição e atividade das comunidades microbianas (Paffetti *et al.*, 1996, Saleena *et al.*, 2001, Noisangiam *et al.*, 2012, Shiro *et al.*, 2013).

Os solos brasileiros apresentam originalmente populações baixas ou inexistentes de simbioses capazes de nodular de modo eficaz plantas de soja (Hungria *et al.*, 2001). Entretanto, inoculações massivas com as poucas estirpes de *Bradyrhizobium* recomendadas para a cultura resultaram em populações relativamente bem estabelecidas, estimadas entre 10³ e 10⁵ células por g de solo na maioria das áreas agrícolas em que há o cultivo da leguminosa (Giongo *et al.*, 2008). Mesmo nestas áreas com populações consideráveis de *Bradyrhizobium* estabelecidas por inoculações em cultivos prévios é possível verificar incrementos médios no rendimento de até 8% pela adoção da prática de reinoculação anual, ressaltando uma possível competição por sítios nodulares entre estirpes utilizadas nos inoculantes e rizóbios já presentes no solo (Lima *et al.*, 1998, Hungria *et al.*, 2007).

No Brasil, a Lei nº 6.894, de 16 de dezembro de 1980 que “*Dispõe sobre a inspeção e fiscalização da produção e do comércio de fertilizantes, corretivos, inoculantes, estimulantes ou biofertilizantes, destinados à agricultura*”, juntamente com os respectivos decretos, instruções normativas, portarias, leis ordinárias e resoluções à esta vinculados, regulamenta a atividade industrial de produção de inoculantes no país com base nos avanços técnico-científicos da área. A legislação prevê somente a comercialização de inoculantes produzidos com as espécies/estirpes recomendadas pelo Ministério da Agricultura Pecuária e Abastecimento (MAPA), garantindo, assim, produtos de elevada qualidade e eficiência técnica, com pureza e alta concentração de células viáveis de *Bradyrhizobium*. Atualmente, para a cultura da soja, são recomendadas as estirpes SEMIA 587 e SEMIA 5019 de *B. elkanii*, além de SEMIA 5079 e SEMIA 5080 de *B. japonicum*, por apresentarem maior eficiência em relação às estirpes previamente empregadas (Campos *et al.*, 2001, Brasil, 2011).

De acordo com os critérios legais, a composição de inoculantes industriais leva em consideração a seleção de estirpes de rizóbios altamente eficientes no processo de FBN, rigorosamente avaliadas sob condições ambientais particulares. Entretanto, o sucesso de tais formulações é frequentemente limitado pela presença de populações de rizóbios já estabelecidas no solo, eventualmente apresentando menor potencial para fixação de N. A maior habilidade competitiva por parte deste último grupo, em função do tamanho das populações, sua ampla distribuição no perfil do solo, e características adaptativas superiores, geralmente lhes proporciona maior ocupação de nódulos em condições de campo (Batista *et al.*, 2015). Além da competição com os micro-organismos “residentes”, variações na habilidade competitiva entre espécies, entre estirpes (em inoculantes multi-estirpe), ou até mesmo entre variantes de uma mesma estirpe (em inoculantes de estirpe única), podem ser inconvenientes e afetar a eficiência do inoculante, demonstrando a necessidade de avaliações contínuas, tanto na qualidade dos inoculantes produzidos, quanto na eficiência de novas estirpes (Kober *et al.*, 2004).

Desta forma, com a necessidade de conciliação dos aspectos técnicos à legislação, e somando-se a estes a dinâmica ecológica dos organismos envolvidos sob as mais diversas condições ambientais, encontrar as combinações mais apropriadas de genótipos vegetais e estirpes de rizóbio torna-se um grande desafio, seja pelo emprego de estratégias de melhoramento vegetal que visem plantas capazes de discriminar e permitir a nodulação

apenas pelos genótipos bacterianos mais eficientes quanto à FBN, ou pelo emprego de inoculantes contendo estirpes de rizóbios “ultracompetentes”, capazes de superar em termos competitivos as espécies/estirpes menos eficientes presentes no solo (Lindström *et al.*, 2010, Olivares *et al.*, 2013).

1.6 Genômica comparativa

O rápido progresso tecnológico das plataformas de sequenciamento genético e consequente redução dos custos promoveram uma revolução na forma de abordar questões biológicas fundamentais para obtenção de *insights* críticos sobre a funcionalidade celular e até mesmo em nível de organismo. Como resultado, a exorbitante quantidade de dados de sequenciamento gerados e acumulados nos anos recentes ultrapassa a capacidade de condução de análises sensíveis dos mesmos por especialistas das mais diversas áreas, viabilizando apenas parte do conhecimento do ponto de vista do potencial de informação que estas sequências carregam (Galperin *et al.*, 2014).

A genômica comparativa é fruto do crescente número de genomas recentemente sequenciados. Nesta nova área de pesquisa, comparam-se diferentes genomas com o intuito de obter informações sobre o *background* genético e inferir possíveis relações funcionais entre os organismos envolvidos, possibilitando sua distinção em nível molecular. Em sua essência, na genômica comparativa, assume-se que sequências que permanecem conservadas (similares) em múltiplas espécies ou em espécies distantes tendem a sê-lo em função de pressões evolutivas equiparáveis, implicando em manutenção da sua função biológica. O contrário, entretanto, não necessariamente seja verdadeiro: uma sequencia de DNA pode possuir determinada função biológica sem ser conservada em relação ao genoma de qualquer outra espécie. Desta mesma forma, conservação também não está necessariamente associada à identidade da sequênciã (Alföldi *et al.*, 2013).

Questões distintas podem ser abordadas comparando-se genomas com diferentes distâncias filogenéticas, sob diferentes graus de resolução. A genômica comparativa fornece ferramentas poderosas, tanto para o estudo de alterações/adaptações evolutivas entre organismos, auxiliando a identificação de genes conservados ou comuns entre espécies, quanto para os genes que fornecem a cada organismo características únicas. Neste sentido, amplos *insights* a respeito das categorias de genes podem ser obtidos pela comparação de genomas, mesmo que distantes do ponto de vista filogenético,

incorporando o estudo da evolução gênica e as relações entre os organismos e seu ambiente (Hardison, 2003). Uma análise integrada deste cenário, envolvendo a construção e utilização de banco de dados de ácidos nucleicos e proteínas, em associação a diferentes técnicas experimentais, como obtenção de organismos com genes *nocauteados* e hibridização fluorescente *in situ* (FISH), pode revelar diferenças e semelhanças genéticas entre organismos a ponto de proporcionar alguma compreensão sobre como estas, por sua vez, contribuem para as diferenças fenotípicas (Klug *et al.*, 2009).

Apesar de uma ampla variedade de ferramentas estar disponível para o estudo de diversos aspectos de sequências genômicas, recursos que permitem a comparação de múltiplos genomas simultaneamente, de forma a explorar os dados em sua totalidade, ainda são relativamente limitados, dificultando comparações diretas (Galperin *et al.*, 2014). Adicionalmente, muitas ferramentas de bioinformática estão implementadas em diferentes servidores e/ou *websites*, ou ainda sob diferentes plataformas operacionais, utilizando nomenclaturas gênicas e formatos de arquivos de dados distintos, tornando inconveniente a análise rebuscada de um único gene por vez e acrescentando mais um grau de complexidade ao processo (Alm *et al.*, 2005).

A comparação entre diferentes genomas procariotos, seja de diferentes espécies de bactérias pertencentes ao mesmo gênero ou diferentes linhagens de um mesmo gênero, permite, por exemplo, o estabelecimento de conjuntos de genes essenciais para aquela espécie, gênero ou grupo de linhagens (seu respectivo *core genome*), como também genes específicos em cada genoma (também conhecido como genoma acessório), que fornecem informações sobre a adaptação do organismo a determinadas situações ambientais (Ozer *et al.*, 2014). Tais comparações também possibilitam inferir sobre a história evolutiva de grupos gênicos (*operons*), se foram transmitidos verticalmente (herança parental) ou se foram adquiridos através de transferência gênica horizontal, fenômeno comum em bactérias. Neste sentido, os procariotos certamente representam um modelo de estudos interessante para compreender mecanismos evolutivos, considerando ainda sua diversidade intraespecífica, dentre e entre populações, as quais, por sua vez, podem apresentar tamanhos variados, sobrevivendo e prosperando em uma ampla gama de ambientes, dos mais regulados aos mais inóspitos (Abby *et al.*, 2007).

2. OBJETIVOS

2.1. Objetivo geral

Esse trabalho teve como objetivo contribuir, através da análise comparativa de dados genômicos de diferentes linhagens de *Bradyrhizobium elkanii*, para um melhor entendimento em relação à regulação dos genes *nod* nessa espécie.

2.2. Objetivos específicos

- Anotação dos genes envolvidos no processo de nodulação (genes *nod* e *nol*) nos genomas de diferentes linhagens de *B. elkanii*, disponíveis nos bancos de dados, a fim de comparar a organização destes em diferentes linhagens dessa espécie;
- Analisar as regiões reguladoras dos genes *nodD₁*, *nodD₂*, *nolA* e *nodVW*, buscando por motivos de ligação de proteínas ativadoras ou repressoras;
- Comparar as proteínas NodD₁ e NodD₂ em relação as suas homologias, aos seus domínios funcionais e as suas estruturas terciárias.

3. MANUSCRITO

Intenção de submissão ao periódico *Molecular Plant-Microbe Interactions (MPMI)*

Fator de impacto: 3.944

Bradyrhizobium elkanii nod regulon: insights through genomic analysis

Ivan Luis Zenzen and Luciane M. P. Passaglia*

Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul. Av. Bento Gonçalves, 9500, Caixa Postal 15.053, 91501-970, Porto Alegre, RS, Brazil.

*Corresponding author at Departamento de Genética, Instituto de Biociências, UFRGS. Av. Bento Gonçalves, 9500, Caixa Postal 15.053, Prédio 43312, sala 207b, Porto Alegre, RS, CEP 91501-970, Brazil. Tel.: +55 51 3308 9813; fax +55 51 3308 7311. E-mail address: luciane.passaglia@ufrgs.br (L. M. P. Passaglia)

Abstract

Many species of legume plants and a group of soil bacteria, collectively called rhizobia, are able to engage in a particular type of symbiotic relationship. In this symbiosis, plant supplies bacteria carbohydrate demands enabling them to fix nitrogen inside a new formed organ – the nodule, which further provides this nutrient to the plant. Establishment of a successful mutualistic relationship requires the products of bacterial structural *nod* genes expression, which encodes for the synthesis of lipochitooligosaccharide nodulation signal molecules, known as Nod Factors (NFs). *B. diazoefficiens* USDA 110 (former *B. japonicum* USDA 110) possesses a wide nodulation gene repertory to undertake NFs assembly and modification, and the transcription of *nodYABCSUIJnolMNO* operon in this strain depends upon transcriptional activators, i.e. products of regulatory *nod* genes responsive to signaling molecules exuded by roots of host plants, usually flavonoid compounds. Central to this regulatory circuit of *nod* genes expression are NodD proteins, members of the LysR-type regulator family. As the major source for soybean [*Glycine max* (L.) Merr.] inoculant industry, strains of *B. japonicum* and *B. elkanii* species dramatically

differ in their physiology and have been shown differential competitive fitness. In this study, we compare public available *B. elkanii* sequenced genomes with the close relative *B. diazoefficiens* USDA 110 reference genome, aiming to get some insights on *B. elkanii* mechanisms of *nod* gene expression, especially concerning to *nod* operon and *nod* regulon. The results obtained allowed to add important aspects in the regulatory model presented to the reference strain that could be extended to strains of *B. elkanii*.

Keywords: *Bradyrhizobium*, NodD₁ protein, *nod* genes, *nod* box.

Introduction

Soybean symbiotic partners mainly belong to the genus *Bradyrhizobium*, initially proposed as a group of slow-growing, alkaline producing root nodule, nitrogen-fixing bacteria (Hollis *et al.*, 1981, Somasegaran *et al.*, 1994). Additional genetic and physiological information, including biochemical profile, DNA homology, and phylogenomics were and are still being employed to clarify differences between isolates, supporting taxonomic classification to over 25 species into the *Bradyrhizobium* genus to date (Jordan, 1982, Garrity *et al.*, 2004, Rivas *et al.*, 2009). Among those species, considerable research efforts have been focused on *B. japonicum* and *B. elkanii* due to their commercial use as source for inoculant formulations, with many effective long-term programs for elite strains identification and selection being carried out in different countries (Campos *et al.*, 2001, Kober *et al.*, 2004, Hungria *et al.*, 2006, Melchiorre *et al.*, 2011).

Characterization of symbiosis itself is not an easy task since it shows a high degree of specialization in a way that a restrict group of rhizobial species/strains interacts with only a select range of plant species/varieties and *vice-versa*. Moreover, specificity may occur at distinct stages of the interaction, from the very beginning bacterial contact with the plant root to late nodule development and nitrogen fixation, resulting in a sort of biological nitrogen fixation (BNF) outcome patterns according to host-microsymbiont combinations (Schumpp *et al.*, 2010, Wang *et al.*, 2012). In view of that, symbiosis can be conceived as a complex framework promoted by strong evolutionary forces that involves a stringent initial molecular dialogue and signal exchange between the symbiotic partners. Establishment of a successful mutualistic relationship requires the products of the bacterial

structural *nod* genes expression, which encodes for the synthesis of lipochitooligosaccharide (LCOs) nodulation signal molecules, also known as Nod factors (NFs). Next, recognition of symbiotic NFs by the plant triggers a signaling cascade, which ultimately allows bacterial infection and induces *de novo* organogenesis of the nodule to accommodate the symbiont and further support the nitrogen fixation process (Loh *et al.*, 2003, Tóth *et al.*, 2015). *B. diazoefficiens* USDA 110 (former *B. japonicum* USDA 110) possesses a wide nodulation gene repertoire to undertake NFs assembly and modification.

In several species of rhizobia *nod* genes are frequently organized in an operon (*nod* operon), suggesting that the regulation of their expression integrates common mechanisms (Denarie *et al.*, 1993). Indeed, transcription of *nodYABCSTUIJnolMNO* operon in *B. diazoefficiens* USDA 110 depends upon transcriptional activators, i.e. products of regulatory *nod* genes responsive to signaling molecules exuded by roots of host plants, usually flavonoid compounds (Luka *et al.*, 1993, Loh *et al.*, 2003). Central to this regulatory circuit of *nod* genes expression are NodD proteins, members of the LysR-type regulator family. Upon activation by a particular flavonoid ligand NodD proteins are able to bind to specific DNA motifs upstream of the *nod* operon, the so called *nod boxes*, and selectively control the expression of the structural *nod* genes in the early stages of plant-bacteria interaction (Hong *et al.*, 1987, Henikoff *et al.*, 1988, Denarie *et al.*, 1992, Goethals *et al.*, 1992). Although multiple isoforms of NodD proteins have been identified in distinct rhizobial species, even suggesting a role in broadening plant host spectrum for those symbionts, only two were found in *B. diazoefficiens* USDA 110 genome, NodD₁ and NodD₂, products of *nodD*₁ and *nodD*₂ genes, respectively (Gottfert *et al.*, 1989). These two proteins show distinct expression patterns and play different functional roles in the regulation of structural *nod* gene expression. Active NodD₁, i.e. in the presence of the ligand flavonoid molecule (e.g. genistein in soybean root exudates), operates as a positive transcriptional regulator of *nod* operon in *B. diazoefficiens*. Unique for this organism when compared to other known rhizobial species, NodD₁ is not constitutively expressed; instead it is induced by flavonoid compounds and shows autoregulation (Banfalvi *et al.*, 1988). Although nodulation requires *nod* gene induction by flavonoids in most diazotrophs, efficiency of this process depends on quantitatively, spatially and temporally suitable expression of those genes. Hence it is not surprising that, in addition to positive

transcriptional regulators, *nod* genes are also under control of repressor elements, as is the case for NodD₂ protein, that acts as a negative regulator of *nod* operon (Loh *et al.*, 2003).

In *B. diazoefficiens* USDA 110 the core regulatory mechanism involving NodD₁ and NodD₂ is extended with additional regulators that act synergistically with NodD proteins to modulate the expression of *nod* genes. Of these, it is noteworthy to mention the role of Nola, a MerR-type regulator encoded by *nolA*, and NodVW, product of *nodVW*, which comprises a two-component regulatory system. Initially identified as a soybean genotype-specific nodulation factor, Nola was later proved to be an activator of *nodD₂*, therefore taking part in negative regulation with NodD₂, both playing together the feedback and quorum regulation of *nod* genes (Sadowsky *et al.*, 1991, Garcia *et al.*, 1996, Loh *et al.*, 2003). On the other hand, the NodVW two-component regulatory system provide an alternative flavonoid responsive pathway for *nod* gene activation, which explains the residual nodulation of soybean plants in NodD₁ mutants (Göttfert *et al.*, 1990, Loh *et al.*, 1997).

As the major source for soybean inoculant industry, strains of *B. japonicum* and *B. elkanii* species dramatically differ in their physiology and have been shown differential competitive fitness (Minamisawa, 1989, Teaney *et al.*, 1993, Vasilas *et al.*, 1993, Dobert *et al.*, 1994). Such differences prove the need for identify and characterize the genetic nature of specificity in the symbiotic relationship as a crucial step to develop guided strategies to enhance soybean inoculants effectiveness, given that BNF ultimately provides a mean resource for more sustainable agricultural systems. While there is a relatively great deal of knowledge regarding the genetics and molecular mechanisms of the soybean symbiont *B. diazoefficiens*, including the complete genome sequence, information about *B. elkanii* genetics is still restrict mainly to research focused on specific features. Despite a few draft genome releases and genomic data availability for this organism, comparisons between species are poorly explored. In this study, we compare public available *B. elkanii* sequenced genomes with the close relative *B. diazoefficiens* USDA 110 reference genome (Kaneko *et al.*, 2002), aiming to get some insights on *B. elkanii* mechanisms of *nod* gene expression, especially concerning to *nod* operon and *nod* regulon, thus providing a more comprehensive understanding in molecular dynamics and complexity of mechanisms involved in fine-tuning signal communication between this symbiont and its host plants.

Material and Methods

Strains and genomic data

Genomic data consisting of complete genome from *Bradyrhizobium diazoefficiens* USDA 110 (reference genome) and draft genomes from *B. elkanii* strains SEMIA 587, CCBAU 05737, CCBAU 43297, USDA 94, USDA 3254 and USDA 3259 were obtained from the public available database of The National Center for Biotechnology Information (NCBI) at <http://www.ncbi.nlm.nih.gov>. Genome features and accession numbers are described in Table 1. Type-strain taxonomy of organism's genomes was confirmed by 16S rRNA sequence analysis using RDP SeqMatch k-nearest-neighbor (k-NN) classifier (Wang *et al.*, 2007), and checked with Basic Local Alignment Search Tool (BLAST) searches and pairwise global sequence alignments implemented in EzTaxon server database (Kim *et al.*, 2012).

Annotation of genomes

Functional annotation was performed with the Rapid Annotation using Subsystem Technology (RAST) server (Aziz *et al.*, 2008), with Glimmer set for gene calling (Salzberg *et al.*, 1998), allowing frameshift correction, backfilling of gaps, and automatic fixing errors. Assigned functional features were triple-checked with InterProScan (Zdobnov *et al.*, 2001) by signature-recognition method in InterPro database (Hunter *et al.*, 2009), ScanProsite (de Castro *et al.*, 2006) for protein signature matches in PROSITE database (Sigrist *et al.*, 2010), and BLASTp against UniProtKB database (Magrane *et al.*, 2011). An inventory of genes involved in nodulation (structural- and regulatory-*nod/nol* genes) for each genome in this study is listed in Supplementary Table 1. *B. diazoefficiens* USDA 110 genome (Kaneko *et al.*, 2002) was taken as reference, and missing *nod/nol* genes were searched for in other genomes with BLASTn using homologous nucleotide sequence of the closely related reference species. Possible frameshift annotation errors in assigned genes were corrected, and Open Reading Frames (ORFs) checked with the Expert Protein Analysis System (ExPASy) translate tool (Gasteiger *et al.*, 2003) in contrast to respective reference.

Table 1. Characteristics of analyzed genomes.

	<i>B. diazoefficiens</i> USDA 110	<i>B. elkanii</i> SEMIA 587	<i>B. elkanii</i> CCBAU 05737	<i>B. elkanii</i> CCBAU 43297	<i>B. elkanii</i> USDA 94	<i>B. elkanii</i> USDA 3254	<i>B. elkanii</i> USDA 3259
Geographic location	USA	Brazil	China	China	USA	USA	USA
Number of contigs	1*	2,431	751	654	174	87	102
Size (Mb)	9.11	8.68	9.77	9.35	9.56	8.98	8.72
Av. read coverage	-	32x	109x	120x	NA**	NA	NA
N50 (Kb)	-	6,929	25,010	25,032	138,573	249,311	245,986
G+C (%)	64.1	63.6	63.5	63.8	63.7	63.8	63.9
Number RNA calls	3 rRNA 49 tRNA	3 rRNA 40 tRNA	3 rRNA 47 tRNA	3 rRNA 48 tRNA	3 rRNA 47 tRNA	3 rRNA 47 tRNA	3 rRNA 47 tRNA
Number CDS calls	8648	8236	9396	8928	9122	8565	8310
NCBI BioSample	GCF000011365.1	SAMN02471378	SAMN02469485	SAMN02469464	SAMN02584917	SAMN02441448	SAMN02440775
NCBI BioProject	PRJNA17	PRJNA86995	PRJNA77219	PRJNA77219	PRJNA165317	PRJNA165319	PRJNA162999
NCBI assembly accession	GCA_000011365.1	GCA_000257685.1	GCA_000261505.1	GCA_000261525.1	GCA_000519225.1	GCA_000472765.1	GCA_000473005.1

*chromosome fully sequenced

**NA = not available

Prediction of subcellular localization of proteins in cells was performed with sequence-based tools in a coordinate fashion (Emanuelsson *et al.*, 2007). Initially, prospection for the presence and location of signal peptide cleavage sites characteristic of secretory proteins was carried out in amino acid sequences using SignalP 4.1 (Petersen *et al.*, 2011), followed by *ab initio* prediction of non-classical protein secretion, i.e. not triggered by signal peptide, with SecretomeP 2.0 (Bendtsen *et al.*, 2004). Lipoprotein signal peptides and N-terminal membrane helices prediction was performed using LipoP 1.0 (Rahman *et al.*, 2008), while presence and location of potential twin-arginine translocation signal peptide cleavage sites was verified with TatP 1.0 (Bendtsen *et al.*, 2005). Finally, prediction of transmembrane topology in proteins was accessed using a combined approach based on hidden Markov model (HMM) algorithm implemented in TMHMM 2.0 server (Krogh *et al.*, 2001) and Phobius (Käll *et al.*, 2004).

Phylogenetic and sequence analysis

Multiple sequence alignment of NodD proteins was carried out by distance estimation using *kmer* counting and progressive alignment with log-expectation score, with subsequent refinement using tree dependent restricted partitioning employed by MUSCLE (Edgar, 2004). Algorithm was implemented in the Molecular Evolutionary Genetics Analysis - MEGA 6.0 package (Tamura *et al.*, 2013), and parameters set for Neighbor-Joining clustering method in all interactions, with -2.9 and 0 for gap opening and gap extension penalties, respectively (center specified as 0). Subsequent phylogenetic analysis and tree reconstruction were performed using Neighbor-Joining method in the same package, with molecular distances of aligned sequences computed according to p-distance parameters and 1,000 bootstrap replicates and pairwise deletion treatment for gaps. Point accepted mutations (PAM) 250 calculations have been adopted as substitution matrix model for scoring sequence alignments in this study (Dayhoff *et al.*, 1978).

Protein structure prediction and alignment

Protein 3D structure prediction was performed by SWISS-MODEL web server (Biasini *et al.*, 2014) based on evolutionary related structures using amino acid sequence and protein structure homology. The technique employs HMM sensitive searches against the SWISS-MODEL template library (SMTL) to generate a structural model of a protein of interest, and QMEAN potential for model quality estimation with independent accuracy evaluation by the Continuous Automated Model EvaluatiOn project – CAMEO (Haas *et*

et al., 2013) based on target sequences pre-released weekly by Protein Data Bank (PDB) (Berman *et al.*, 2007). Alignment of predicted protein structures was accomplished by combinatorial extension (CE) algorithm (Shindyalov *et al.*, 1998), implemented in the Research Collaboratory for Structural Bioinformatics (RCSB) PDB Protein Comparison Tool at www.rcsb.org (Bernstein *et al.*, 1977, Berman *et al.*, 2007, Goodsell *et al.*, 2015), with parameters set as 30 for maximum gap size allowed during aligned fragment pairs (AFP) extension in fragment size $m=8$, with gap open and gap extension penalties of 5.0 and 0.5, respectively.

Results

nod operon and regulatory genes

B. elkanii genome annotation followed by manual curation revealed a conserved operon structure and organization of *nodKABCSUIJ* genes in strains CCBAU 05737, CCBAU 43297, USDA 94, USDA 3254, and USDA 3259, similar to the organization presented in *B. diazoefficiens* USDA 110 *nod* operon, with substitution of *nodY* for the corresponding *nodK*. Curiously, in *B. elkanii* SEMIA 587 a distinct pattern for this “canonical” gene organization was observed, with these *nod* genes scattered in the chromosome and organized as *nodKABCS*, a second sparse block containing *nodI*, *nodJ*, *nolO*, and *nodZ* separated by approximately 890 kb from the *nod* operon, and then *nodU*, also at approximately 890 kb of distance from *nodZ*. *nolO* and *nodZ* were also present in all analyzed genomes, located in the vicinity of each other and physically close to *nodJ* as a part of the *nod* operon, with the exception of SEMIA 587 where they are in the upstream region of *nod* operon, near to *nodJ*. Other structural *nol* genes such as *nolZ* and *nolMN*, could not be identified in the genomes of any *B. elkanii* strain, as well as *nolY* in the genomes of strains USDA 94, USDA 3254, and USDA 3259 (Fig. 1).

Annotation also highlighted the presence of *nodD*₁ and *nodD*₂ regulatory genes in all six *B. elkanii* genomes analyzed, as well as in *B. diazoefficiens* USDA 110 genome. Organization of these two genes for most *B. elkanii* genomes involved in this study follows the pattern observed for *B. diazoefficiens* USDA 110, i.e. they are positioned close to each other and upstream to the *nod* operon. The two exceptions of this organization are found in the genomes of CCBAU 05737 and USDA 3254 strains, in which these genes are located downstream to the *nod* operon. Similarly, the presence of *nodVW* that code for the two-

component regulatory elements was ubiquitous in all six *B. elkanii* genomes. *nodVW* relative location and distances related to *nod* operon vary according to each genome, from a relatively far upstream region in SEMIA 587, CCBAU 43297, and USDA 94 genomes, to a relatively far downstream region in CCBAU 05737, USDA 3254, and USDA 3259 genomes, compared to their relatively near location in the upstream region (~300 kb) of *B. diazoefficiens* USDA 110 *nod* operon. Conversely, *nolA* gene apparently shows a conserved location within *B. elkanii* genomes, close to *nodD*₂. Exception was for strain SEMIA 587, where it is located near to *nodVW*, very distant from *nodD*₂ (Fig. 1).

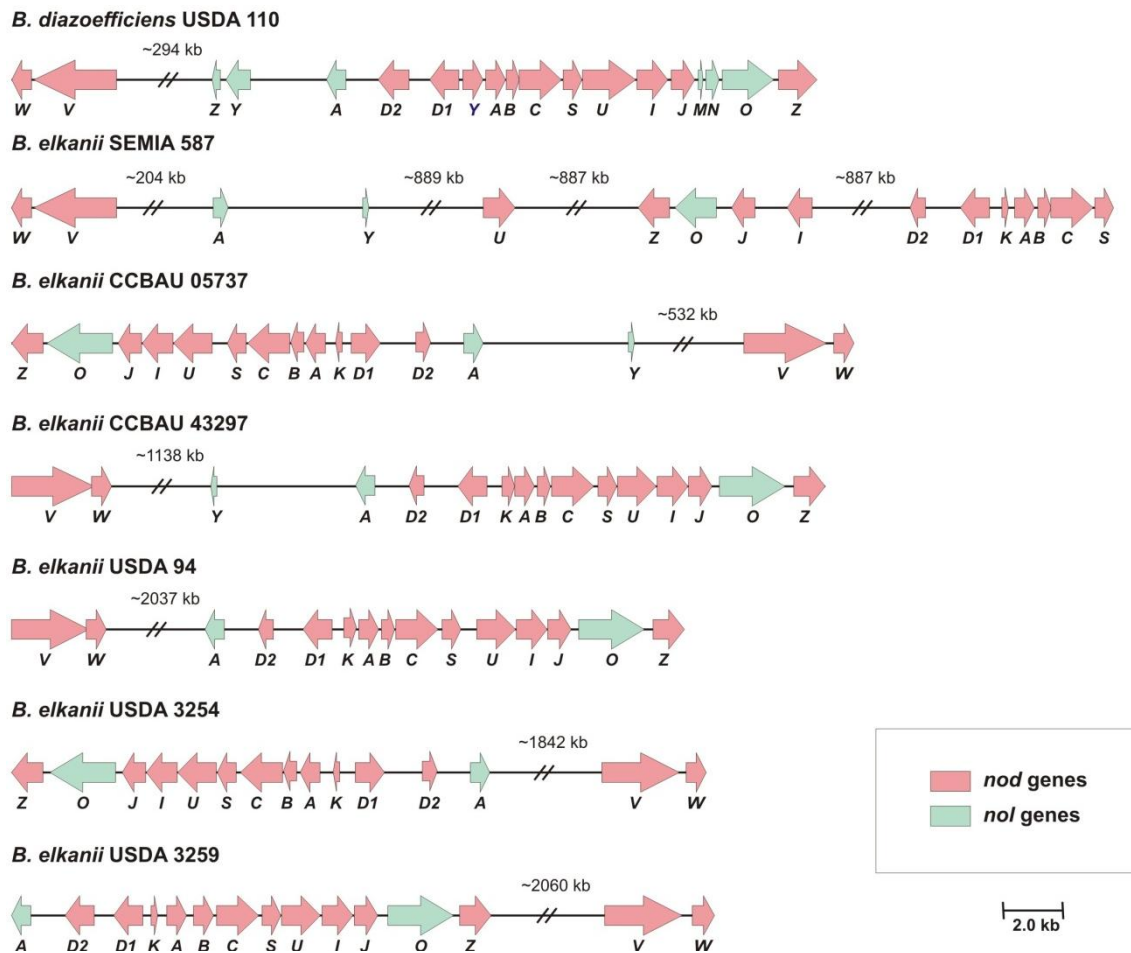


Fig. 1 Genetic map with organization of *nod/nol* genes in *B. diazoefficiens* USDA 110 (reference genome) and six *B. elkanii* genomes. Letters indicate the name of each gene present in the genome. Gene class is color code assigned, and transcription orientation is arrow oriented. Relative gene distances are indicated otherwise follows the scale presented.

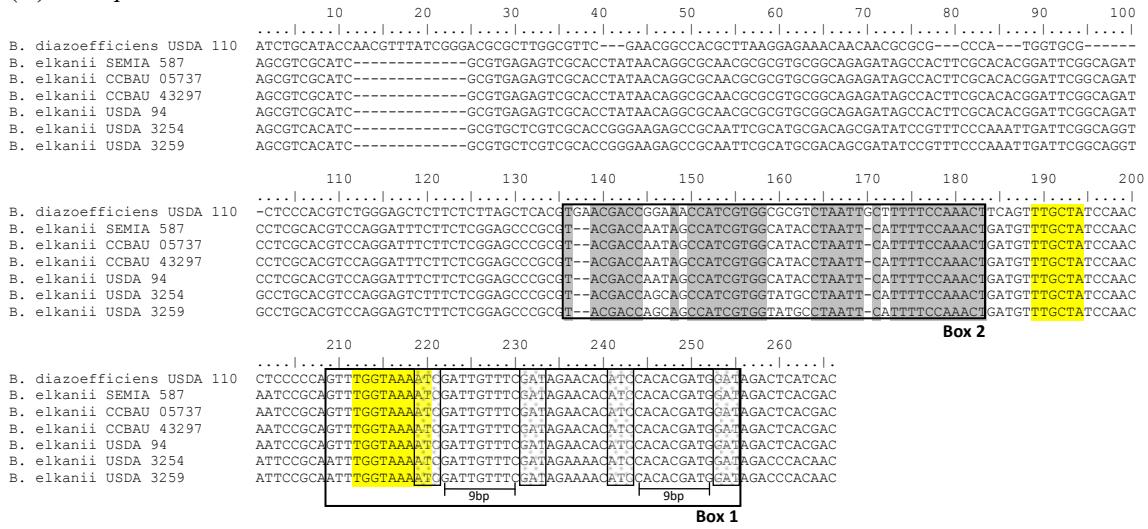
Analysis of *nodD* promoter regions

Analysis of a 250 bp region upstream of *nodD*₁ ORFs revealed that all six *B. elkanii* genomes, as well as *B. diazoefficiens* USDA 110 genome, contain one -10/-35 σ^{70} potential promoter TTGCTA-N₁₇-TGGTAAAAT located at 46 bp upstream from *nodD*₁ CDS start site. Additionally, all sequences show two 47 bp *nod* boxes, one of which corresponding to a consensus *nod* box sequence with the respective palindromic structure, located at 11 bp from *nodD*₁ CDS start site that controls the transcription of *nod* operon. The second one is a presumptive *nod* box-like sequence that is located in the upstream region (84 bp) of *nodD*₁ CDS start site (Fig. 2A). Interestingly, the consensus *nod* box sequence (Box1) overlaps the sequence of -10/-35 σ^{70} putative *nodD*₁ promoter, while the presumptive *nod* box-like sequence is located at just 6 bp upstream to it (Fig. 2A).

On the other hand, the analysis of the 250 bp region upstream of *nodD*₂ revealed that only *B. elkanii* SEMIA 587, CCBAU 05737, CCBAU 43297, and USDA 94 genomes exhibited a -10/-35 σ^{70} potential promoter with sequence TTGTCG-N₁₃-CTGTAAAAG, along with a partial *nod* box-like sequence structure located at 152 bp upstream from the putative σ^{70} promoter. In *B. diazoefficiens* USDA 110 as well as in *B. elkanii* USDA 3254 and USDA 3259 genomes any of those elements could be identified in the *nodD*₂ promoter region, i.e. apparently they do not display potential -10/-35 σ^{70} promoters neither a partial (or complete) *nod* box-like sequence (Fig. 2B).

nolA promoter regions (250 bp) from all genomes included in this study exhibited a -10/-35 σ^{70} potential promoter sequence TTGAAT-N₁₆-TTGTAGGCT, except for *B. elkanii* SEMIA 587. Additionally, apart from *B. elkanii* SEMIA 587, *nolA* promoter regions display high degree of sequence similarity among genomes (Supp. Fig. 1A). Potential -10/-35 σ^{70} promoter sequences were found in all genomes for *nodVW* regulatory gene sequences, although differences in structure and location of these promoters were observed. In contrast to *nolA*, promoter regions of *nodVW* genes were considerably dissimilar in sequence conservation among genomes (Supp. Fig. 1B). Search for *nod* box in *nolA* and *nodVW* promoter regions revealed the absence of this regulatory motif in their respective 250 bp upstream gene sequences for all genomes (Supp. Fig. 1A-B).

(A) *nodD*₁



(B) *nodD*₂

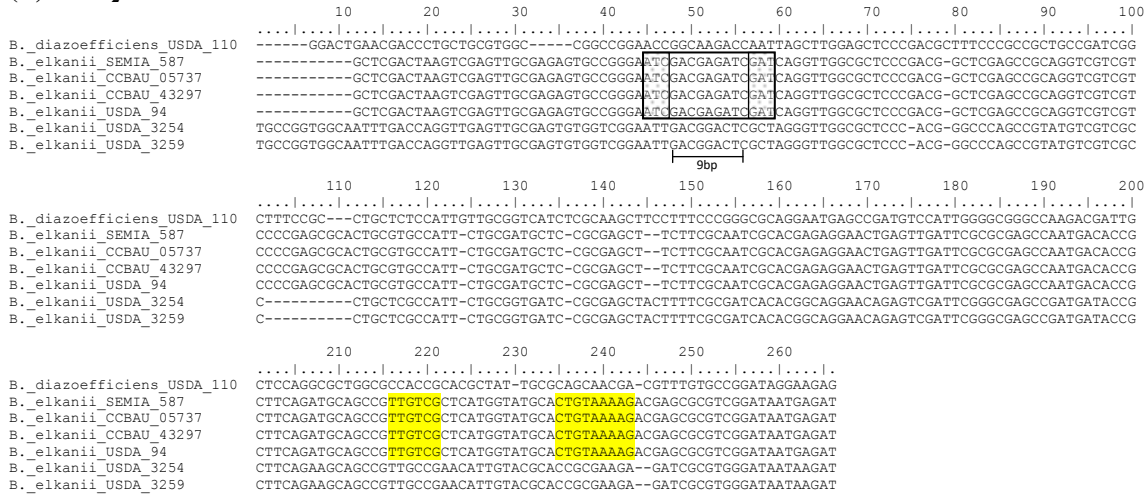


Figure 2. Alignment of the ~250 bp upstream region from *nodD* ORFs in *Bradyrhizobium* genomes. (A) The *nodD*₁ *nod* box sequences are boxed, showing a consensus *nod* box sequence near the *nodD*₁ ORF start codon (Box1), and a presumptive *nod* box-like sequence (Box2) upstream with conserved residues highlighted in grey; (B) *nodD*₂ partial *nod* box sequence structure found in some *B. elkanii* genomes is boxed. Inner boxes denote the putative *nod* box motifs ATC-N₉-GAT with their respective palindromic structure. Yellow dashed sequences show a -10/-35 σ^{70} potential promoter. Sequences read 5' to 3' from right to left, with last nucleotide in the sequence representing that immediately prior to ATG from *nodD*₁ (A) or *nodD*₂ (B) CDS start site.

Regulatory Nod proteins sequence analysis

All *B. elkanii* analyzed genomes display NodD₁ open reading frames (ORFs) of 314 amino acid residues, in agreement with NodD₁ protein of *B. diazoefficiens* USDA 110. Conversely, NodD₂ ORFs size show discreet variation among the organisms, with ORFs consisting of 330 amino acid residues in *B. diazoefficiens* USDA 110, 331 residues in *B. elkanii* strains SEMIA 587, CCBAU 05737, CCBAU 43297, and USDA 94, and 329 residues in *B. elkanii* strains USDA 3254 and USDA 3259 (Supp. Table 1).

NodD₁ and NodD₂ exhibit a relatively high degree of conservation, although some dissimilarity between organisms ultimately clusters each *B. elkanii* protein set into two groups (Fig. 3). Indeed, PAM250 matrix shows that *B. elkanii* strains SEMIA 587, CCBAU 05737, CCBAU 43297, and USDA 94 share 100% global sequence similarity among themselves for both NodD₁ and NodD₂; consequently, each protein set of these strains group together. The same was noticed between USDA 3254 and USDA 3259, although sequence similarity decreases 7% for NodD₁ and 12% for NodD₂ compared to their respective homologs from other strains within the same species. NodD regulator proteins from *B. elkanii* strains are still conserved when contrasted to *B. diazoefficiens* USDA 110 orthologs, especially NodD₁ which exhibited at least 93% sequence similarity, while for NodD₂ no less than 73% was observed (Supp. Fig. 2A-B). Additionally, conservation plots of NodD₁ and NodD₂ proteins from *B. elkanii* analyzed genomes revealed lower conservation at the carboxy-terminus region of proteins, especially for NodD₂, when compared to the reference genome of *B. diazoefficiens* USDA 110 (Supp. Fig. 3A-B).

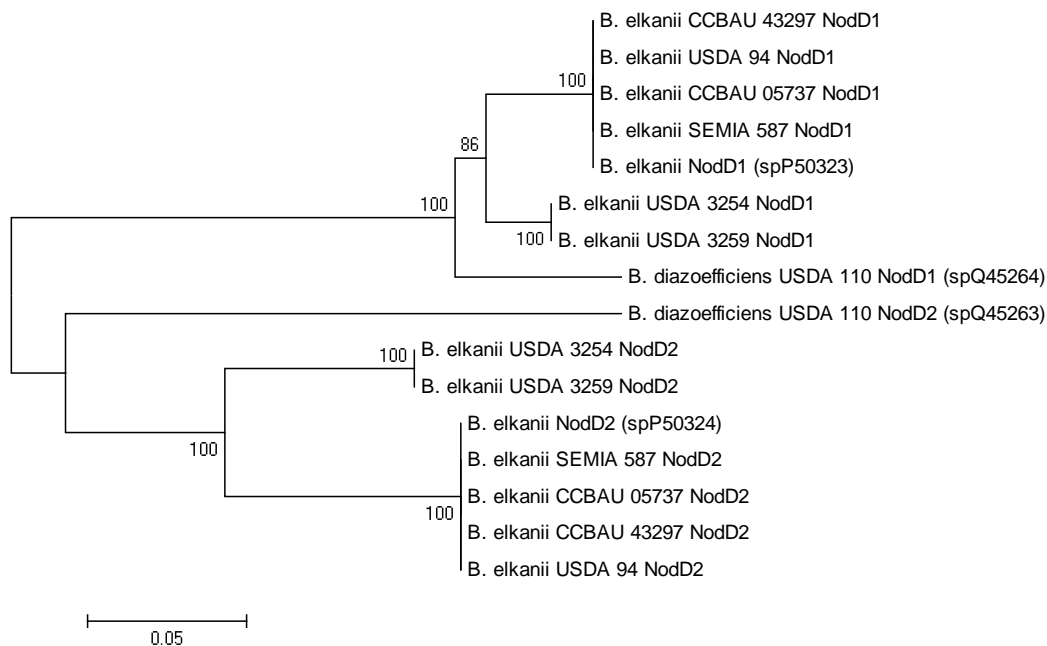


Figure 3. Evolutionary relationship of *nod* regulatory proteins NodD₁ and NodD₂ from *B. elkanii* strains analyzed. Included curated reference proteins NodD₁ (spP50323) and NodD₂ (spP50324) from *B. elkanii*. The evolutionary history was inferred using the Neighbor-Joining method. Optimal tree with the sum of branch length = 0.63145170 is shown, with the percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree, which were computed using the p-distance method and are represented in the units of the number of amino acid differences per site. All ambiguous positions were removed for each sequence pair and there were a total of 331 positions in the final dataset.

All analyzed sequences belong to the LysR family of transcriptional regulators, and exhibited a 57-amino acid helix-turn-helix (HTH) LysR-type domain located at position 6-63. These domains also contain a 20-amino acid HTH DNA-binding motif from residue 23-42 at the amino-terminus region of both NodD₁ and NodD₂ proteins. NodD₁ DNA-binding motif from *B. elkanii* strains SEMIA 587, CCBAU 05737, CCBAU 43297, and USDA 94 is highly conserved, and shares 100% identity for the sequence LTAAARQINLSQPAMSAAIA (Fig. 4A). Curiously, *B. elkanii* USDA 3254 and USDA 3259 share 100% identity with *B. diazoefficiens* USDA 110 NodD₁ DNA-binding motif LTAAARKINLSQPAMSAAIA, where a single amino acid substitution takes place at position 29, changing a glutamine (Q) for a lysine (K) amino acid residue (Fig. 4A).

(A) NodD₁



(B) NodD₂

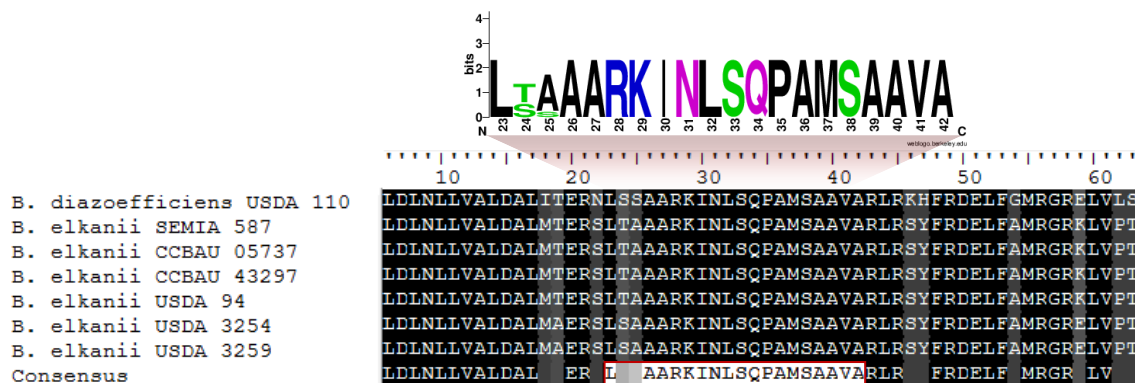


Figure 4. Multiple sequence alignment (MSA) of *Bradyrhizobium elkanii* NodD HTH LysR-type domain alignment and sequence logo showing HTH DNA-binding motif conservation. HTH DNA-binding sequence conservation among organisms is red boxed. Showing in (A) NodD₁ domain sequence and (B) NodD₂ domain sequence.

Likewise, *B. elkanii* NodD₂ DNA-binding motif also split into two groups, with strains SEMIA 587, CCBAU 05737, CCBAU 43297, and USDA 94 showing 100% identity for the sequence LTAAARKINLSQPAMSAAVA, while for strains USDA 3254 and USDA 3259 a serine (S) replaces the threonine (T) at position 24 of NodD₂, resulting in the DNA-binding motif LSAAARKINLSQPAMSAAVA (Fig. 4B). Comparing NodD₂ from *Bradyrhizobium* species, *B. diazoefficiens* USDA 110 shows higher similarity to *B. elkanii* strains USDA 3254 and USDA 3259, although an additional amino acid residue divergence at position 25 could be observed, with replacement of an alanine (A) in *B. elkanii* for a serine (S) in *B. diazoefficiens*, providing the sequence LSSAARKINLSQPAMSAAVA (Fig. 4B).

In silico analysis of subcellular localization of NodD₁ and NodD₂ proteins did not predict the presence of any significant ordinary signal peptide cleavage site in amino acid sequences, as well as any twin-arginine signal peptide cleavage site. Lipoprotein signal peptides and non-classical protein secretion patterns were also not identified. In accordance, additional protein topology and signal peptide examination did not point out for the presence of transmembrane helices. Ultimately, calculations using an integrative approach algorithm indicate NodD₁ and NodD₂ likely as cytoplasmic components and less probable as membrane-bound proteins.

Regulatory Nod proteins structure analysis

Global protein structure alignment of NodD₁ from *B. elkanii* strains exhibits high identity and similarity to NodD₁ protein from reference genome *B. diazoefficiens* USDA 110. Particularly, genomes of *B. elkanii* strains USDA 3254 and USDA 3259 which display values of 92.62% and 97.65% for NodD₁ structural identity and similarity compared to orthologous protein from *B. diazoefficiens* USDA 110, respectively. Despite slightly lower, NodD₁ from genomes of *B. elkanii* strains SEMIA 587, CCBAU 05737, CCBAU 43297, and USDA 94, still conserve considerable structural identity and similarity to NodD₁ from *B. diazoefficiens* USDA 110, in the order of 91.61% and 95.97%, respectively. Comparing NodD₁ structural alignments, it seems that higher identity and similarity observed for *B. elkanii* strains USDA 3254 and USDA 3259 to *B. diazoefficiens* USDA 110 resides mainly in the HTH motif of the DNA-binding domain (Supp. Fig. 4A-F). Conversely, NodD₂ global protein structure alignment from *B. elkanii* SEMIA 587,

CCBAU 05737, CCBAU 43297, and USDA 94 genomes demonstrate higher identity (75.17%) and similarity (84.56%) to ortholog NodD₂ from *B. diazoefficiens* USDA 110 than *B. elkanii* strains USDA 3254 and USDA 3259 genomes, for which identity and similarity no higher than 74.16% and 82.89% was observed, respectively (Supp. Fig. 5A-F). Additionally, even though substantial identity and similarity could be observed for structural alignment of NodD₂ between *Bradyrhizobium* species, values were lower compared to structural alignments showed for NodD₁ proteins.

Interestingly, when global structural alignment of proteins NodD₁ versus NodD₂ is performed within each genome, identity and similarity still can be considered relatively significant. In *B. elkanii* strains USDA 3254 and USDA 3259 genomes for example, NodD₁ versus NodD₂ structural alignment achieve 70.47% identity and 81.88% similarity, while for *B. elkanii* strains SEMIA 587, CCBAU 05737, CCBAU 43297, and USDA 94 genomes, identity and similarity attained values of 69.46% and 81.54%, respectively. The reference genome of *B. diazoefficiens* USDA 110 display the more dissimilar structural alignment of NodD₁ versus NodD₂, and consequently shows the lowest identity (63.76%) and similarity (80.54%) values (Supp. Fig. 6A-F). However, taking into account only the 57-residue protein segment corresponding to the DNA-binding domain containing the HTH motif, NodD₁ versus NodD₂ structural alignment values are considerably higher in all genomes, being identical in *B. elkanii* strains USDA 3254 and USDA 3259 genomes, i.e. reaching 100% of identity and similarity. *B. elkanii* strains SEMIA 587, CCBAU 05737, CCBAU 43297, and USDA 94 genomes also showed better structural alignment of amino-terminus region that accommodates the DNA-binding domain for NodD₁ versus NodD₂ comparison, with 91.38% identity and 98.28% similarity, as well as in *B. diazoefficiens* USDA 110, with no less than 81.03% and 93.10% for identity and similarity, respectively (Fig. 5).

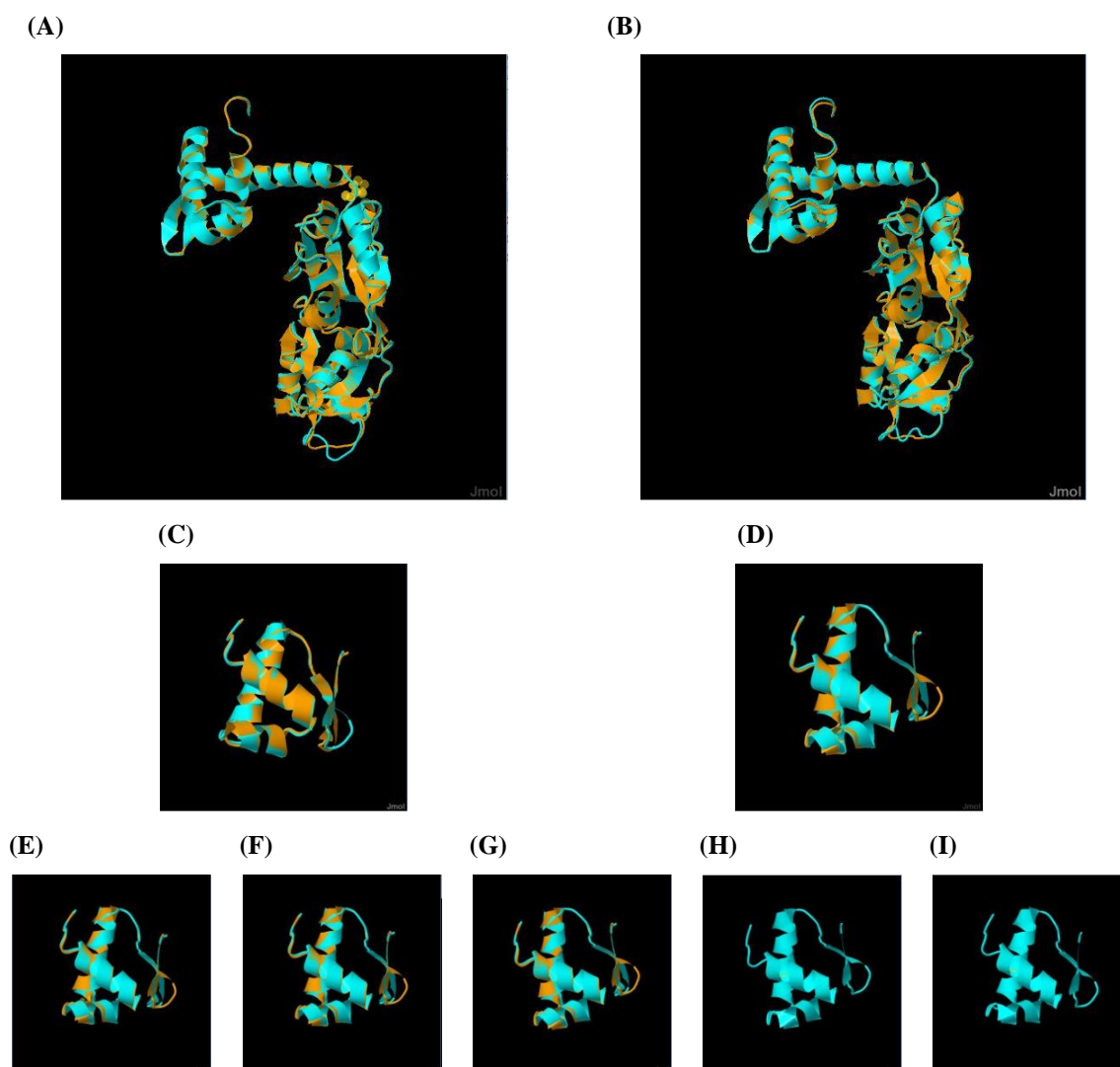


Figure 5. Structural alignment of NodD₁ and NodD₂ proteins within each *Bradyrhizobium* genome using Combinatorial Extension (CE) algorithm. Global NodD₁-NodD₂ protein structure alignment from *B. diazoefficiens* USDA 110 (A) and *B. elkanii* SEMIA 587 (B). Structural alignment of NodD₁-NodD₂ 57-residue HTH domain in *B. diazoefficiens* USDA 110 (C), *B. elkanii* SEMIA 587 (D), *B. elkanii* CCBAU 05737 (E), *B. elkanii* CCBAU 43297 (F), *B. elkanii* USDA 94 (G), *B. elkanii* USDA 3254 (H), and *B. elkanii* USDA 3259 (I). Orange colour residues represent NodD₁, superimposed by light-blue residues representing NodD₂.

Discussion

nod operon/regulon in *Bradyrhizobium* genomes

Lipochitooligosaccharides (LCOs), resulting of the action of the bacterial nodulation (*nod*, *nol*, *noe*) gene products, are key signal molecules in *Bradyrhizobium*-legume symbiosis establishment that triggers formation of a new organ (i.e. the nodule) within which biological nitrogen fixation occurs (Day *et al.*, 2000). Both, *B. diazoefficiens* and *B. elkanii* are able to nodulate soybean plants. Despite this common ability, refined

genetic and biochemical studies involving different *Bradyrhizobium* species have shown distinct physiological profiles that ultimately can affect their symbiotic behavior, including nodulation capacity, nitrogen fixation efficiency and competitive ability for nodule formation (Kober *et al.*, 2004). Most of the structural and regulatory *nod/nol* genes with known function seem to be conserved between *B. diazoefficiens* USDA 110 and *B. elkanii* genomes. Moreover, the presence of the structural *nodY/KABCSUIJnolOnodZ* genes, disposed in an operon organization for *B. diazoefficiens* and most *B. elkanii* genomes, may indicate that similar qualitative LCO outcome patterns can be expected for these organisms, depending primarily upon particular operon regulation in each genome. Organization of *nod* genes in clusters/operons is a common and often conserved feature among various rhizobia genomes, many of them, as in the case of *B. diazoefficiens*, located on the chromosome, thus facilitating coordination of the expression of these respective genes (Schlaman *et al.*, 1998). Although the operon organization of *nod* genes in *Bradyrhizobium* is frequent, it is not a fixed pattern, as can be observed for *B. elkanii* SEMIA 587 genome which shows the reduced *nodKABCS* operon with additional *nod/nol* genes scattered along the chromosome. Interestingly, the separation of the genes from this operon may result on differences in regulation of *nod* genes, ultimately providing an altered qualitative LCO pattern for this strain compared to other *B. elkanii* genomes, even though this is not set as mandatory for divergent *nod* gene arrangement (Vázquez *et al.*, 1991). As in most rhizobia, expression of *nod/nol* operon in *Bradyrhizobium* is dependent on distinct flavonoid compound cocktails produced and secreted by each leguminous host plant, with quantity and spectrum of these molecules varying according to the age and physiological state of the plant (Phillips *et al.*, 1995, Loh *et al.*, 2003, Hassan *et al.*, 2012).

Regulation of *nod/nol* operon in *B. diazoefficiens* and *B. elkanii* is essentially under control of the same elements, i.e. the *nod* regulon, consisting on *nodD*₁, *nodD*₂, *nolA* and *nodVW* genes, all of them present in the genomes of both species. The pattern of gene organization observed for *nodD*₁ and *nodD*₂, which are invariably located in the close vicinity of each other right upstream of and divergent from the *nod* operon in *Bradyrhizobium* genomes, reinforces their importance as key components in the core regulatory mechanism of operonic *nod* gene expression. Indeed, products of *nodD* genes are assumed as central to the coordination of structural *nod* genes in many symbionts, including *Bradyrhizobium* species (Mulligan *et al.*, 1989, Honma *et al.*, 1990, Loh *et al.*,

2003, del Cerro *et al.*, 2015). For instance, in *B. diazoefficiens* products of *nodD*₁ and *nodD*₂ regulatory genes orchestrate *nodYABCSTUIInolOnodZ* operon expression by its activation or repression, respectively (Loh *et al.*, 2003). Such regulatory system suggests a common mechanism that may apply for *B. elkanii* genomes with conserved *nod* gene organization. Besides their divergent transcription orientation, components of the LysR-type transcriptional regulator (LTTR) family, such as NodD₁ and NodD₂, also exhibit another characteristic feature, i.e. the property of autoregulation (Maddocks *et al.*, 2008), which ultimately indicates that expression of its associated operon largely depends on the regulation of these regulatory genes themselves. Since *B. diazoefficiens nodD* mutants maintain at least a marginal ability to nodulate soybean plants, alternative transcriptional activators were proposed to take part in the process. In fact, identification of a two-component regulatory system consisting of NodV and NodW, products of *nodVW*, respectively, shed some light on this paradox, providing a suitable explanation. In such a system, the sensor kinase component NodV can detect the environmental stimulus (e.g. a specific isoflavone) resulting in its autophosphorylation and subsequent signal transduction by transfer of the phosphoryl group to its cognate response regulator protein NodW, which in turn is then able to activate its target *nod* operon (Loh *et al.*, 1997). Further evidence that *Bradyrhizobium nodVW* products are essential for efficient nodulation of mung bean [*Vigna radiata* (L.) Wilezek], cowpea [*Vigna unguiculata* (L.) Walp.], and siratro [*Macroptilium atropurpureum* (Moc. & Sessé ex DC.) Urb.], but are not required for soybean, point towards a host-specificity role of these proteins, probably by recognition of specific flavonoid inducers produced by these host plant species by the correlate sensor component NodV (Göttfert *et al.*, 1990, Sanjuan *et al.*, 1994). Here we could identify the presence of the same system in *B. elkanii* genomes, suggesting that both species possess an alternative pathway for *nod* operon activation and may provide a strategy to broaden their respective host ranges.

Whereas successful soybean nodulation depends on proper structural *nod* gene expression timely, at the right place, and in suitable amounts, it is reasonable to assume that negative regulation also takes place in addition to their activation. Although this negative control in *B. diazoefficiens* is mediated by the product of *nodD*₂ that operates as a *nod* operon repressor as mentioned before, this process is assisted by *nolA* gene product, which set-up an additional level of regulation by inducing *nodD*₂ expression under

appropriate conditions. In other words, the product of *nolA* acts as an transcriptional inducer of the repressor *nodD*₂ (Garcia *et al.*, 1996). As already mentioned, Nola is assigned as a member of the MerR family of transcriptional regulators that activates transcription upon binding to specific DNA motifs and consequently induces a DNA bending leading to its appropriate alignment for RNA polymerase positioning and resulting in transcription activation (Ansari *et al.*, 1992, Philips *et al.*, 2015). Moreover, *nolA* has the peculiar feature to encode for the three distinct proteins Nola₁, Nola₂, and Nola₃, originated from three in-frame ATG start codons, with Nola₁ controlling the expression of the other two and being the one involved in activation of *nodD*₂ (Loh *et al.*, 1999). This mechanism is also shared with high homology by *B. elkanii* genomes, and for the most it shows a very similar gene organization and transcription orientation together with *nodD*₁ and *nodD*₂, as previously demonstrated (Dobert *et al.*, 1994), thus pointing to integrate *nod* gene regulation in such species as well.

***cis-regulatory nod* elements and *trans-acting* factors**

Considering the nature of genomic elements present in the *nod* operon and especially in the *nod* regulon from *B. diazoefficiens* and their markedly similarity with those in *B. elkanii* genomes, both in terms of structure and organization, we could hypothesize that major phenotypic differences in LCOs production and secretion patterns in these two species are given at some transcriptional or post-transcriptional regulatory level. Accordingly, in the presence of the correspondent flavonoid compounds, activated NodD proteins specifically bind to conserved *cis-regulatory* elements on bacterial DNA, the *nod* boxes. These regulatory structures basically consist of a 47 bp conserved region containing the *nod* box consensus sequence with the palindromic ATC-N₉-GAT motif, being located upstream to *nod* operon and controlling its expression (Rostas *et al.*, 1986, Nieuwkoop *et al.*, 1987, Goethals *et al.*, 1992). Conservation of such *cis-regulatory* element and its essentiality in many rhizobia species studied so far strongly suggests NodD₁ as central in *nod* operon gene expression for those species, including *Bradyrhizobium* (Spaink *et al.*, 1987). *nod* operon activation by NodD₁ in *B. elkanii* genomes are not an exception to this rule, and seem to be under a similar if not identical regulatory mechanism, since the same genetic features are observed in this context, highly conserved according to *B. diazoefficiens* USDA 110.

Besides activation of *nod* gene expression, an autoregulatory function unique to *Bradyrhizobium* is assigned for *nodD*₁ by its own product NodD₁ upon binding the same isoflavone inducing molecules that activate *nod* operon. Due to divergent transcription orientation, *nod* box sequences are in the upstream region from both, the *nod* operon containing the structural *nod* genes, as well as *nodD*₁ and *nodD*₂ regulatory genes. This location of *nod* boxes in between *nod* operon and *nodD*₁/*nodD*₂ within the genome makes it more difficult and complex to analyze their regulation. Although autoregulation performed by NodD₁ is also accomplished by a DNA-binding mechanism, it is claim to be achieved by binding to an alternative presumptive *nod* box-like sequence, located upstream to the consensus *nod* box in *B. diazoefficiens* USDA 110 (Banfalvi *et al.*, 1988, Wang *et al.*, 1991). Indeed, conservation of this regulatory sequence in *B. elkanii* genomes reinforce its proposed function, which is also supported by the close proximity of this additional *cis*-regulatory region to a -10/-35 σ^{70} potential promoter. Despite similarities such as high sequence conservation and location in the genome, a few differences still exist from presumptive *nod* box-like sequences in distinct *Bradyrhizobium* species (e.g. a slightly smaller size and some nucleotide divergence in *B. elkanii* genomes compared to *B. diazoefficiens*), which ultimately may affect the DNA-binding process itself.

Based on *nodD*₁ and *nodD*₂ sequence-homology and their resulting description as members of LysR-type regulators, it is wise at least to consider a potential autoregulatory role also for NodD₂, akin to what can be observed for NodD₁. Previous work on *Rhizobium japonicum* USDA 191 scanning its DNA sequences for the presence of a *nod* box, revealed that only *nodD*₁ promoter region show a high homology sequence to other *nod* boxes, while no extensive similar motifs were identified upstream to *nodD*₂, even though several alignments of up to four homologous base pairs could be observed (Appelbaum *et al.*, 1988). Although none complete *nod* box has been found in the promoter region of *nodD*₂ from *Bradyrhizobium* genomes, a remarkable conserved sequence containing only one copy of the palindrome ATC-N₉-GAT, characteristic of a *nod* box sequence structure, was identified in some *B. elkanii* genomes. Curiously, only the same genomes that contain this palindromic sequence displayed a potential -10/-35 σ^{70} binding site. Since it was not reported before, the concrete biological significance of such genomic signature is still unclear and may represent another mechanism in addition to NolA, by which *nodD*₂ transcription is regulated in this species, an issue that certainly deserve some further

investigation. Nowadays, the proposed model for *nodD*₂ expression in *B. diazoefficiens* essentially considers induction of transcription by Nola upon binding to a putative Nola binding-site present upstream to *nodD*₂ (Garcia *et al.*, 1996, Loh *et al.*, 2003). Sequence homology shows that Nola is a member of the MerR transcriptional regulator protein family, which is known to activate suboptimal σ^{70} -dependent promoters through protein-dependent DNA distortion that ultimately provides an appropriate alignment of -10/-35 σ^{70} and the correct positioning of RNA polymerase to contact the respective promoter (Brown *et al.*, 2003). Surprisingly, in the present work, a preliminary screening for potential Nola binding site upstream to *nodD*₂ was not able to identify any correspondent conserved *cis*-element in this region according to current literature, revealing a lack of further information to carry out exhaustive exploratory analysis in greater detail.

NodD protein sequences and structure conservation

In addition to *cis*-regulatory regions in the bacterial genome, many aspects regarding to the biochemical nature of *trans*-acting factors can affect transcriptional regulation. The high degree of protein conservation observed for NodD₁ and NodD₂ among *Bradyrhizobium* genomes provides concise evidence about the evolutionary importance of the mechanism whereupon these regulators operate. In fact, NodD amino acid sequence can display a wide range of conservation, with values higher than 90% of sequence similarity when comparing NodD proteins from organisms of the same genus, to lower than 50% for comparison of NodD sequences from distant related organisms that possess these proteins (Göttfert *et al.*, 1992). Although NodD global sequence similarity can vary considerably among different organisms, a general feature is the high-to-low conservation in amino acid sequence within the protein from the amino to the carboxy-terminus of the polypeptide, respectively. This structural pattern can be observed in many rhizobia species and is consistent with the function addressed to each domain of the protein (Burn *et al.*, 1987, Horvath *et al.*, 1987, Burn *et al.*, 1989). The N-terminus region, the most conserved part of the protein, anchors the HTH DNA-binding motif responsible for the recognition of the *nod* box *cis*-elements in bacterial DNA. Despite this higher conservation in sequence similarity, subtle amino acid substitutions observed among *B. elkanii* genomes apparently occurs, which ultimately may modify the affinity of such motif for the DNA, enhancing or even reducing the pattern of binding and consequent activity of these regulators. On the other hand, NodD C-terminus region is believed to be primarily involved in protein

oligomerization, an important aspect since it is strongly accepted that NodD₁ is active in its tetrameric form that arises from their cognate homodimers (Peck *et al.*, 2013). In this context, it is suggested that multiple weak interactions of non-specific amino acids residues in this region of NodD proteins take place, contributing to oligomerization itself (Ezezika *et al.*, 2007, Peck *et al.*, 2013). As observed in other rhizobia, NodD C-terminus region is also clearly less conserved in *B. elkanii* genomes. Even though this lower sequence similarity did not affect in great extent the protein structure, it may play a role in the oligomerization process, depending on the variability in qualitative and quantitative terms, and thus conferring distinct patterns of oligomerization.

Taking into account all the available information of the reference genome from *B. diazoefficiens* USDA 110, the similarities between *B. diazoefficiens* and *B. elkanii*, and the results obtained from this comparative genomic analysis, an extension in the model of *nod* gene regulation proposed for *B. diazoefficiens* can be applied to *B. elkanii* with some new added potential features (Fig. 6). Besides the activity of all known elements in the core mechanism of *nod* operon regulation by the already discussed *nod* regulon, the presence of a partial *nod* box sequence upstream to *nodD*₂ may suggest a putative NodD₁ binding site in this region with a possible regulation effect on *nodD*₂ as a transcriptional activation. We can speculate that such mechanism may act under higher concentrations of NodD₁ as a consequence of autoregulation of *nodD*₁ expression by its respective product. Under those conditions, a surplus of NodD₁ may be “trapped” by this partial *nod* box sequence, which probably show lower affinity to the regulator protein relative to the conserved consensus *nod* box sequence, and ultimately activate transcription of the repressor protein NodD₂ to balance the expression of *nod* operon. Additionally, prediction of highly conserved NodD protein structure patterns between NodD₁ and NodD₂ and their structural alignment led to another hypothesis regarding to the potential NodD₁-NodD₂ heterodimers formation, since their protein structures are very similar and no specific amino acid residues responsible for oligomerization could be identified in experiments carried out to date. If that is the case and NodD₁-NodD₂ heterodimers may be formed in *Bradyrhizobium*, they can affect the regulation driven by NodD₁ in this symbiont, also interfering in *nod* operon gene expression. Although data may enable inferences on these events, these are not conclusive evidences and experimental confirmation of such mechanisms should be performed to prove the validity of this model.

expression. Our work has enabled to draw a general framework of several similarities between *B. elkanii* genomes and the close relative *B. diazoefficiens*, from which it was possible to infer correlate functional mechanisms and key elements that play an essential role in the regulation of *nod* gene expression. Besides, it revealed new genomic features that were not clearly explored before, some of them unique for some *B. elkanii* genomes, raising new interesting questions for further research.

Acknowledgements

The authors gratefully acknowledge the valuable suggestions given by Prof. Dr. Maria Cátira Bortolini and MSc. Lucas Viscardi (UFRGS). This work was financed by a Grant and fellowships from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq/Brazil) and the Instituto Nacional de Ciência e Tecnologia (INCT) da Fixação Biológica de Nitrogênio (Brazil).

References

- Ansari, A.Z., Chael, M.L. and O'Halloran, T.V. (1992). Allosteric underwinding of DNA is a critical step in positive control of transcription by Hg-MerR. *Nature* **355**, 87-89.
- Appelbaum, E.R., Thompson, D.V., Idler, K. and Chartrain, N. (1988). Rhizobium japonicum USDA 191 has two nodD genes that differ in primary structure and function. *J Bacteriol* **170**, 12-20.
- Aziz, R., Bartels, D., Best, A., DeJongh, M., Disz, T., Edwards, R., *et al.* (2008). The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics* **9**, 75.
- Banfalvi, Z., Nieuwkoop, A., Schell, M., Besl, L. and Stacey, G. (1988). Regulation of nod gene expression in Bradyrhizobium japonicum. *Molecular & general genetics : MGG* **214**, 420-424.
- Bendtsen, J., Nielsen, H., Widdick, D., Palmer, T. and Brunak, S. (2005). Prediction of twin-arginine signal peptides. *BMC bioinformatics* **6**, 167.
- Bendtsen, J.D., Jensen, L.J., Blom, N., von Heijne, G. and Brunak, S. (2004). Feature-based prediction of non-classical and leaderless protein secretion. *Protein Engineering Design and Selection* **17**, 349-356.
- Berman, H., Henrick, K., Nakamura, H. and Markley, J.L. (2007). The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Research* **35**, D301-D303.
- Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer, E.F., Jr., Brice, M.D., Rodgers, J.R., *et al.* (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J Mol Biol* **112**, 535-542.
- Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., *et al.* (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research* **42**, W252-W258.
- Brown, N.L., Stoyanov, J.V., Kidd, S.P. and Hobman, J.L. (2003). The MerR family of transcriptional regulators. *FEMS microbiology reviews* **27**, 145-163.
- Burn, J., Rossen, L. and Johnston, A.W.B. (1987). Four classes of mutations in the nodD gene of Rhizobium leguminosarum biovar. viciae that affect its ability to autoregulate and/or activate other nod genes in the presence of flavonoid inducers. *Genes & Development* **1**, 456-464.
- Burn, J.E., Hamilton, W.D., Wootton, J.C. and Johnston, A.W. (1989). Single and multiple mutations affecting properties of the regulatory gene nodD of Rhizobium. *Mol Microbiol* **3**, 1567-1577.
- Campos, B.C., Hungria, M. and Tedesco, V. (2001). Eficiência da fixação biológica de N₂ por estirpes de Bradyrhizobium na soja em plantio direto. *Revista Brasileira de Ciência do Solo* **25**, 583-592.
- Day, R.B., Loh, J., Cohn, J.R. and Stacey, G. (2000) Signal exchange involved in the establishment of the Bradyrhizobium-legume symbiosis. In *Prokaryotic Nitrogen Fixation, a Model System for the Analysis of a Biological Process.*, E. Triplett (ed.). Norfolk, England, Horizon Scientific Press, pp. 385-414.
- Dayhoff, M.O. and Schwartz, R.M. (1978) A model of evolutionary change in proteins. In *In Atlas of protein sequence and structure*. Citeseer.
- de Castro, E., Sigrist, C.J.A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P.S., Gasteiger, E., *et al.* (2006). ScanProsite: detection of PROSITE signature matches

- and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Research* **34**, W362-W365.
- del Cerro, P., Rolla-Santos, A., Gomes, D., Marks, B., Perez-Montano, F., Rodriguez-Carvajal, M., *et al.* (2015). Regulatory nodD1 and nodD2 genes of *Rhizobium tropici* strain CIAT 899 and their roles in the early stages of molecular signaling and host-legume nodulation. *BMC Genomics* **16**, 251.
- Denarie, J., Debelle, F. and Rosenberg, C. (1992). Signaling and Host Range Variation in Nodulation. *Annual Review of Microbiology* **46**, 497-531.
- Denarie, J., Debelle, F., Truchet, G. and Prome, J.-C. (1993) *Rhizobium* and Legume Nodulation: A Molecular Dialogue. In *New Horizons in Nitrogen Fixation*, R. Palacios, J. Mora, W. Newton (eds.). Springer Netherlands, pp. 19-30.
- Dobert, R.C., Breil, B.T. and Triplett, E.W. (1994). DNA sequence of the common nodulation genes of *Bradyrhizobium elkanii* and their phylogenetic relationship to those of other nodulating bacteria. *Molecular plant-microbe interactions : MPMI* **7**, 564-572.
- Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792-1797.
- Emanuelsson, O., Brunak, S., von Heijne, G. and Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protocols* **2**, 953-971.
- Ezezika, O.C., Haddad, S., Neidle, E.L. and Momany, C. (2007). Oligomerization of BenM, a LysR-type transcriptional regulator: structural basis for the aggregation of proteins in this family. *Acta Crystallographica Section F* **63**, 361-368.
- Garcia, M., Dunlap, J., Loh, J. and Stacey, G. (1996). Phenotypic characterization and regulation of the *nolA* gene of *Bradyrhizobium japonicum*. *Molecular plant-microbe interactions : MPMI* **9**, 625-636.
- Garrity, G.M., Bell, J.A. and Lilburn, T.G. (2004). Taxonomic outline of the prokaryotes. *Bergey's manual of systematic bacteriology*. Springer, New York, Berlin, Heidelberg.
- Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D. and Bairoch, A. (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research* **31**, 3784-3788.
- Goethals, K., Van Montagu, M. and Holsters, M. (1992). Conserved motifs in a divergent nod box of *Azorhizobium caulinodans* ORS571 reveal a common structure in promoters regulated by LysR-type proteins. *Proceedings of the National Academy of Sciences* **89**, 1646-1650.
- Goodsell, D.S., Dutta, S., Zardecki, C., Voigt, M., Berman, H.M. and Burley, S.K. (2015). The RCSB PDB “Molecule of the Month”: Inspiring a Molecular View of Biology. *PLoS Biol* **13**, e1002140.
- Göttfert, M., Grob, P. and Hennecke, H. (1990). Proposed regulatory pathway encoded by the *nodV* and *nodW* genes, determinants of host specificity in *Bradyrhizobium japonicum*. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 2680-2684.
- Göttfert, M., Holzhäuser, D., Bäni, D. and Hennecke, H. (1992). Structural and functional analysis of two different *nodD* genes in *Bradyrhizobium japonicum* USDA110. *Mol. Plant-Microbe Interact* **5**, 257-265.
- Gottfert, M., Lamb, J.W., Gasser, R., Semenza, J. and Hennecke, H. (1989). Mutational analysis of the *Bradyrhizobium japonicum* common nod genes and further nod box-linked genomic DNA regions. *Molecular & general genetics : MGG* **215**, 407-415.

- Haas, J., Roth, S., Arnold, K., Kiefer, F., Schmidt, T., Bordoli, L. and Schwede, T. (2013). The Protein Model Portal—a comprehensive resource for protein structure and model information. *Database* **2013**.
- Hassan, S. and Mathesius, U. (2012). The role of flavonoids in root–rhizosphere signalling: opportunities and challenges for improving plant–microbe interactions. *Journal of Experimental Botany* **63**, 3429-3444.
- Henikoff, S., Haughn, G.W., Calvo, J.M. and Wallace, J.C. (1988). A large family of bacterial activator proteins. *Proceedings of the National Academy of Sciences* **85**, 6602-6606.
- Hollis, A.B., Kloos, W.E. and Elkan, G.H. (1981). DNA:DNA Hybridization Studies of *Rhizobium japonicum* and Related Rhizobiaceae. *Microbiology* **123**, 215-222.
- Hong, G.F., Burn, J.E. and Johnston, A.W. (1987). Evidence that DNA involved in the expression of nodulation (nod) genes in *Rhizobium* binds to the product of the regulatory gene nodD. *Nucleic Acids Research* **15**, 9677-9690.
- Honma, M.A., Asomaning, M. and Ausubel, F.M. (1990). *Rhizobium meliloti* nodD genes mediate host-specific activation of nodABC. *Journal of Bacteriology* **172**, 901-911.
- Horvath, B., Bachem, C.W., Schell, J. and Kondorosi, A. (1987). Host-specific regulation of nodulation genes in *Rhizobium* is mediated by a plant-signal, interacting with the nodD gene product. *The EMBO journal* **6**, 841-848.
- Hungria, M., Franchini, J.C., Campo, R.J., Crispino, C.C., Moraes, J.Z., Sibaldelli, R.N.R., *et al.* (2006). 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.
- Hunter, S., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., *et al.* (2009). InterPro: the integrative protein signature database. *Nucleic Acids Research* **37**, D211-D215.
- Jordan, D.C. (1982). NOTES: 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.
- Käll, L., Krogh, A. and Sonnhammer, E.L.L. (2004). A Combined Transmembrane Topology and Signal Peptide Prediction Method. *Journal of Molecular Biology* **338**, 1027-1036.
- Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T., Sasamoto, S., *et al.* (2002). Complete Genomic Sequence of Nitrogen-fixing Symbiotic Bacterium *Bradyrhizobium japonicum* USDA110. *DNA Research* **9**, 189-197.
- Kim, O.-S., Cho, Y.-J., Lee, K., Yoon, S.-H., Kim, M., Na, H., *et al.* (2012). Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. *International Journal of Systematic and Evolutionary Microbiology* **62**, 716-721.
- Kober, M.d.V., Sá, E.L.S.d., Freire, J.R.J. and Giongo, A. (2004). Characterization of variants of *Bradyrhizobium elkanii* and *B. japonicum* and symbiotic behaviour in soybeans. *Ciência Rural* **34**, 1459-1464.
- Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E.L.L. (2001). Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *Journal of Molecular Biology* **305**, 567-580.

- Loh, J., Garcia, M. and Stacey, G. (1997). NodV and NodW, a second flavonoid recognition system regulating nod gene expression in *Bradyrhizobium japonicum*. *Journal of Bacteriology* **179**, 3013-3020.
- Loh, J. and Stacey, G. (2003). Nodulation Gene Regulation in *Bradyrhizobium japonicum*: a Unique Integration of Global Regulatory Circuits. *Applied and Environmental Microbiology* **69**, 10-17.
- Loh, J., Stacey, M.G., Sadowsky, M.J. and Stacey, G. (1999). The *Bradyrhizobium japonicum* nolA Gene Encodes Three Functionally Distinct Proteins. *Journal of Bacteriology* **181**, 1544-1554.
- Luka, S., Sanjuan, J., Carlson, R.W. and Stacey, G. (1993). nolMNO genes of *Bradyrhizobium japonicum* are co-transcribed with nodYABCSUIJ, and nolO is involved in the synthesis of the lipo-oligosaccharide nodulation signals. *Journal of Biological Chemistry* **268**, 27053-27059.
- Maddocks, S.E. and Oyston, P.C.F. (2008). Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology* **154**, 3609-3623.
- Magrane, M. and Consortium, U. (2011). UniProt Knowledgebase: a hub of integrated protein data. *Database* **2011**.
- Melchiorre, M., de Luca, M., Gonzalez Anta, G., Suarez, P., Lopez, C., Lascano, R. and Racca, R. (2011). Evaluation of bradyrhizobia strains isolated from field-grown soybean plants in Argentina as improved inoculants. *Biol Fertil Soils* **47**, 81-89.
- Minamisawa, K. (1989). Comparison of Extracellular Polysaccharide Composition, Rhizobitoxine Production, and Hydrogenase Phenotype among Various Strains of *Bradyrhizobium japonicum*. *Plant and Cell Physiology* **30**, 877-884.
- Mulligan, J.T. and Long, S.R. (1989). A family of activator genes regulates expression of *Rhizobium meliloti* nodulation genes. *Genetics* **122**, 7-18.
- Nieuwkoop, A.J., Banfalvi, Z., Deshmane, N., Gerhold, D., Schell, M.G., Sirotkin, K.M. and Stacey, G. (1987). A locus encoding host range is linked to the common nodulation genes of *Bradyrhizobium japonicum*. *Journal of Bacteriology* **169**, 2631-2638.
- Peck, M.C., Fisher, R.F., Bliss, R. and Long, S.R. (2013). Isolation and Characterization of Mutant *Sinorhizobium meliloti* NodD1 Proteins with Altered Responses to Luteolin. *Journal of Bacteriology* **195**, 3714-3723.
- Petersen, T.N., Brunak, S., von Heijne, G. and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Meth* **8**, 785-786.
- Philips, S.J., Canalizo-Hernandez, M., Yildirim, I., Schatz, G.C., Mondragon, A. and O'Halloran, T.V. (2015). TRANSCRIPTION. Allosteric transcriptional regulation via changes in the overall topology of the core promoter. *Science* **349**, 877-881.
- Phillips, D.A. and Streit, W.R. (1995) Legume Signals to Rhizobial Symbionts: A New Approach for Defining Rhizosphere Colonization. In *Plant-Microbe Interactions*, G. Stacey, N.T. Keen (eds.). Springer Science & Business Media.
- Rahman, O., Cummings, S., Harrington, D. and Sutcliffe, I. (2008). Methods for the bioinformatic identification of bacterial lipoproteins encoded in the genomes of Gram-positive bacteria. *World Journal of Microbiology and Biotechnology* **24**, 2377-2382.
- Rivas, R., Martens, M., de Lajudie, P. and Willems, A. (2009). Multilocus sequence analysis of the genus *Bradyrhizobium*. *Systematic and Applied Microbiology* **32**, 101-110.

- Rostas, K., Kondorosi, E., Horvath, B., Simoncsits, A. and Kondorosi, A. (1986). Conservation of extended promoter regions of nodulation genes in *Rhizobium*. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 1757-1761.
- Sadowsky, M.J., Cregan, P.B., Gottfert, M., Sharma, A., Gerhold, D., Rodriguez-Quinones, F., *et al.* (1991). The Bradyrhizobium japonicum nodA gene and its involvement in the genotype-specific nodulation of soybeans. *Proc Natl Acad Sci U S A* **88**, 637-641.
- Salzberg, S.L., Delcher, A.L., Kasif, S. and White, O. (1998). Microbial gene identification using interpolated Markov models. *Nucleic Acids Research* **26**, 544-548.
- Sanjuan, J., Grob, P., Göttfert, M., Hennecke, H. and Stacey, G. (1994). NodW is essential for full expression of the common nodulation genes in Bradyrhizobium japonicum. *MPMI-Molecular Plant Microbe Interactions* **7**, 364-369.
- Schlaman, H.R., Phillips, D.A. and Kondorosi, E. (1998) Genetic Organization and Transcriptional Regulation of Rhizobial Nodulation Genes. In *The Rhizobiaceae: Molecular Biology of Model Plant-Associated Bacteria*, H.P. Spaink, A. Kondorosi, P. Hooykaas (eds.). Springer Netherlands, pp. 361-386.
- Schumpp, O. and Deakin, W.J. (2010). How inefficient rhizobia prolong their existence within nodules. *Trends in Plant Science* **15**, 189-195.
- Shindyalov, I.N. and Bourne, P.E. (1998). Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Engineering* **11**, 739-747.
- Sigrist, C.J.A., Cerutti, L., de Castro, E., Langendijk-Genevaux, P.S., Bulliard, V., Bairoch, A. and Hulo, N. (2010). PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Research* **38**, D161-D166.
- Somasegaran, P. and Hoben, H.J. (1994) Handbook for Rhizobia: methods in legume-rhizobium technology, Springer-Verlag.
- Spaink, H.P., Wijffelman, C.A., Pees, E., Okker, R.J.H. and Lugtenberg, B.J.J. (1987). Rhizobium nodulation gene nodD as a determinant of host specificity. *Nature* **328**, 337-340.
- Tamura, K., Stecher, G., Peterson, D., FilipSKI, A. and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution* **30**, 2725-2729.
- Teaney, G., III and Fuhrmann, J. (1993). Soybean response to nodulation by rhizobitoxine-producing bradyrhizobia as influenced by nitrate application. *Plant Soil* **154**, 219-225.
- Tóth, K. and Stacey, G. (2015). Does plant immunity play a critical role during initiation of the legume-rhizobium symbiosis? *Frontiers in Plant Science* **6**, 401.
- Vasilas, B.L. and Furhmann, J.J. (1993). Field Response of Soybean to Nodulation by a Rhizobitoxine-Producing Strain of Bradyrhizobium. *Agronomy Journal* **85**, 302-305.
- Vázquez, M., Dávalos, A., de las Peñas, A., Sánchez, F. and Quinto, C. (1991). Novel organization of the common nodulation genes in *Rhizobium leguminosarum* bv. phaseoli strains. *Journal of Bacteriology* **173**, 1250-1258.
- Wang, D., Yang, S., Tang, F. and Zhu, H. (2012). Symbiosis specificity in the legume – rhizobial mutualism. *Cellular Microbiology* **14**, 334-342.

- Wang, Q., Garrity, G.M., Tiedje, J.M. and Cole, J.R. (2007). Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Applied and Environmental Microbiology* **73**, 5261-5267.
- Wang, S.P. and Stacey, G. (1991). Studies of the Bradyrhizobium japonicum nodD1 promoter: a repeated structure for the nod box. *Journal of Bacteriology* **173**, 3356-3365.
- Zdobnov, E.M. and Apweiler, R. (2001). InterProScan – an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847-848.

Supplemental material

Supplementary Table 1. *Bradyrhizobium nod* operon/regulon gene inventory

Organism	Gene		nt	aa
	ID	Coordinates		
<i>Bradyrhizobium diazoefficiens</i> USDA 110	<i>nodW</i> *	BA000040.2_1878986_1879579	594	197
	<i>nodV</i> *	BA000040.2_1882335_1879666	2670	889
	<i>nolZ</i>	BA000040.2_2175594_2175821	228	75
	<i>nolY</i>	BA000040.2_2175928_2176656	729	242
	<i>nolA</i> *	BA000040.2_2179134_2179706	573	190
	<i>nodD2</i> *	BA000040.2_2180766_2181758	993	330
	<i>nodD1</i> *	BA000040.2_2182392_2183336	945	314
	<i>nodY</i>	BA000040.2_2183450_2184088	639	212
	<i>nodK</i>	-	-	-
	<i>nodA</i>	BA000040.2_2184130_2184762	633	210
	<i>nodB</i>	BA000040.2_2184963_2185416	453	151
	<i>nodC</i>	BA000040.2_2185433_2186890	1458	485
	<i>nodS</i>	BA000040.2_2186919_2187446	528	175
	<i>nodU</i>	BA000040.2_2187460_2189169	1710	569
	<i>nodI</i>	BA000040.2_2189186_2190091	906	301
	<i>nodJ</i>	BA000040.2_2190095_2190883	789	262
	<i>nolM</i>	BA000040.2_2190916_2191089	174	57
	<i>nolN</i>	BA000040.2_2191127_2191522	396	131
	<i>nolO</i>	BA000040.2_2191558_2193165	1647	548
	<i>nodZ</i>	BA000040.2_2193331_2194443	1113	370
<i>Bradyrhizobium elkanii</i> SEMIA 587	<i>nodW</i> *	AJJK01000027.1_118833_119536	704	233
	<i>nodV</i> *	AJJK01000029.1_119623_122295	2673	890
	<i>nolZ</i>	-	-	-
	<i>nolY</i>	AJJK01000096.1_326076_326309	234	77
	<i>nolA</i> *	AJJK01000095.1_321036_321522	487	162
	<i>nodD2</i> *	AJJK01002387.1_8523194_8524189	996	331
	<i>nodD1</i> *	AJJK01002387.1_8524830_8525774	945	314
	<i>nodY</i>	-	-	-
	<i>nodK</i>	AJJK01002387.1_8526308_8526585	278	91
	<i>nodA</i>	AJJK01002387.1_8526625_8527257	663	210
	<i>nodB</i>	AJJK01002387.1_8527458_8527913	456	151
	<i>nodC</i>	AJJK01002387.1_8527930_8529303	1374	457
	<i>nodS</i>	AJJK01002387.1_8529419_8529946	528	175
	<i>nodU</i>	AJJK01000365.1_1215479_1216615	1137	378
	<i>nodI</i>	AJJK01000616.1_2107854_2108629	776	257
	<i>nodJ</i>	AJJK01000616.1_2106915_2106706	792	263
	<i>nolM</i>	-	-	-
	<i>nolN</i>	-	-	-
	<i>nolO</i>	AJJK01000616.1_2104611_2106371	1761	586
	<i>nodZ</i>	AJJK01000616.1_2103328_2104320	993	330
<i>Bradyrhizobium elkanii</i> CCBAU 05737	<i>nodW</i> *	AJPV01000599.1_9181873_9182704	832	277
	<i>nodV</i> *	AJPV01000599.1_9178877_9181549	2673	890
	<i>nolZ</i>	-	-	-
	<i>nolY</i>	AJPV01000548.1_8647548_8647781	234	77
	<i>nolA</i> *	AJPV01000548.1_8642134_8642706	573	190
	<i>nodD2</i> *	AJPV01000548.1_8640072_8641067	996	331
	<i>nodD1</i> *	AJPV01000548.1_8638487_8639431	945	314
	<i>nodY</i>	-	-	-
	<i>nodK</i>	AJPV01000548.1_8637677_8637953	277	91
	<i>nodA</i>	AJPV01000548.1_8637004_8637636	633	210
	<i>nodB</i>	AJPV01000548.1_8636348_8636803	456	151
	<i>nodC</i>	AJPV01000548.1_8634958_8636331	1374	457
	<i>nodS</i>	AJPV01000548.1_8634315_8634842	528	175
	<i>nodU</i>	AJPV01000548.1_8632591_8633781	1191	396
	<i>nodI</i>	AJPV01000548.1_8631669_8632589	921	306
	<i>nodJ</i>	AJPV01000548.1_8630874_8631665	792	263
	<i>nolM</i>	-	-	-
	<i>nolN</i>	-	-	-
	<i>nolO</i>	AJPV01000548.1_8628570_8630597	2028	675
	<i>nodZ</i>	AJPV01000548.1_8627315_8628279	965	321

Supplementary Table 1. (continued)

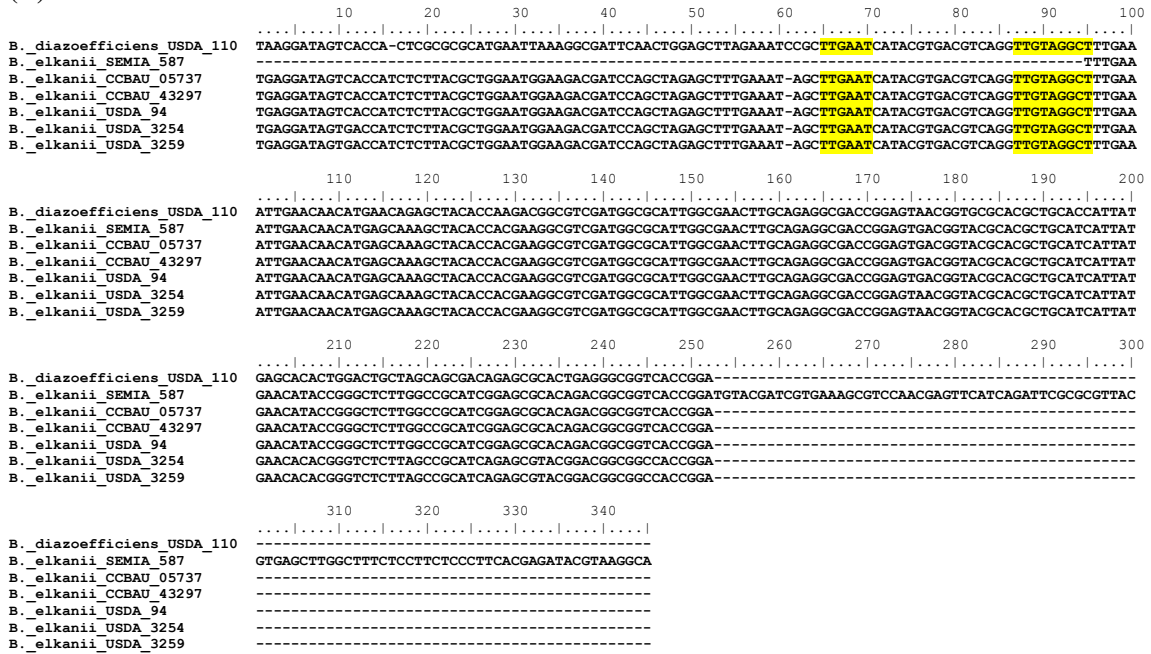
Organism	Gene		nt	aa
	ID	Coordinates		
<i>Bradyrhizobium elkanii</i> CCBAU 43297	<i>nodW*</i>	AJPW01000165.1_2853646_2854325	678	225
	<i>nodV*</i>	AJPW01000165.1_2850986_2853649	2664	887
	<i>nodZ</i>	-	-	-
	<i>nodY</i>	AJPW01000249.1_3991990_3992223	234	77
	<i>nodA*</i>	AJPW01000249.1_3997065_3997637	573	190
	<i>nodD2*</i>	AJPW01000249.1_3998704_3999699	996	331
	<i>nodD1*</i>	AJPW01000249.1_4000340_4001284	945	314
	<i>nodY</i>	-	-	-
	<i>nodK</i>	AJPW01000249.1_4001692_4002094	403	133
	<i>nodA</i>	AJPW01000249.1_4002135_4002767	633	210
	<i>nodB</i>	AJPW01000249.1_4002968_4003423	456	151
	<i>nodC</i>	AJPW01000249.1_4003440_4004813	1374	457
	<i>nodS</i>	AJPW01000249.1_4004929_4005456	528	175
	<i>nodU</i>	AJPW01000249.1_4005471_4007180	1710	569
	<i>nodI</i>	AJPW01000249.1_4007182_4008102	921	306
	<i>nodJ</i>	AJPW01000249.1_4008106_4008897	792	263
	<i>nodM</i>	-	-	-
	<i>nodN</i>	-	-	-
	<i>nodO</i>	AJPW01000249.1_4009174_4011201	2028	675
<i>nodZ</i>	AJPW01000249.1_4011492_4012485	994	330	
<i>Bradyrhizobium elkanii</i> USDA 94	<i>nodW*</i>	JAF01000092.1_6221570_6222249	678	225
	<i>nodV*</i>	JAF01000092.1_6218910_6221573	2464	887
	<i>nodZ</i>	-	-	-
	<i>nodY</i>	-	-	-
	<i>nodA*</i>	JAF01000072.1_8259046_8259618	573	190
	<i>nodD2*</i>	JAF01000072.1_8260685_8261680	996	331
	<i>nodD1*</i>	JAF01000072.1_8262321_8263265	945	314
	<i>nodY</i>	-	-	-
	<i>nodK</i>	JAF01000072.1_8263673_8264076	403	133
	<i>nodA</i>	JAF01000072.1_8264116_8264748	633	210
	<i>nodB</i>	JAF01000072.1_8264949_8265404	456	151
	<i>nodC</i>	JAF01000072.1_8265421_8266794	1374	457
	<i>nodS</i>	JAF01000072.1_8266910_8267437	528	175
	<i>nodU</i>	JAF01000072.1_8267971_8269161	1191	396
	<i>nodI</i>	JAF01000072.1_8269163_8270083	921	306
	<i>nodJ</i>	JAF01000072.1_8270087_8270878	792	263
	<i>nodM</i>	-	-	-
	<i>nodN</i>	-	-	-
	<i>nodO</i>	JAF01000072.1_8271155_8273182	2028	675
<i>nodZ</i>	JAF01000072.1_8273471_8274465	995	331	
<i>Bradyrhizobium elkanii</i> USDA 3254	<i>nodW*</i>	AXAH01000065.1_7903638_7903988	351	116
	<i>nodV*</i>	AXAH01000065.1_7900642_7903314	2673	890
	<i>nodZ</i>	-	-	-
	<i>nodY</i>	-	-	-
	<i>nodA*</i>	AXAH01000041.1_6058297_6058869	573	190
	<i>nodD2*</i>	AXAH01000041.1_6056239_6057228	990	329
	<i>nodD1*</i>	AXAH01000041.1_6054783_6055619	837	278
	<i>nodY</i>	-	-	-
	<i>nodK</i>	AXAH01000041.1_6054210_6054472	263	86
	<i>nodA</i>	AXAH01000041.1_6053183_6053815	633	210
	<i>nodB</i>	AXAH01000041.1_6052527_6052982	456	151
	<i>nodC</i>	AXAH01000041.1_6051120_6052509	1386	461
	<i>nodS</i>	AXAH01000041.1_6050494_6051021	528	175
	<i>nodU</i>	AXAH01000041.1_6048771_6050480	1710	569
	<i>nodI</i>	AXAH01000041.1_6047849_6048769	921	306
	<i>nodJ</i>	AXAH01000041.1_6047054_6047845	792	263
	<i>nodM</i>	-	-	-
	<i>nodN</i>	-	-	-
	<i>nodO</i>	AXAH01000041.1_6044746_6046773	2028	675
<i>nodZ</i>	AXAH01000041.1_6043471_6044465	995	331	

Supplementary Table 1. (continued)

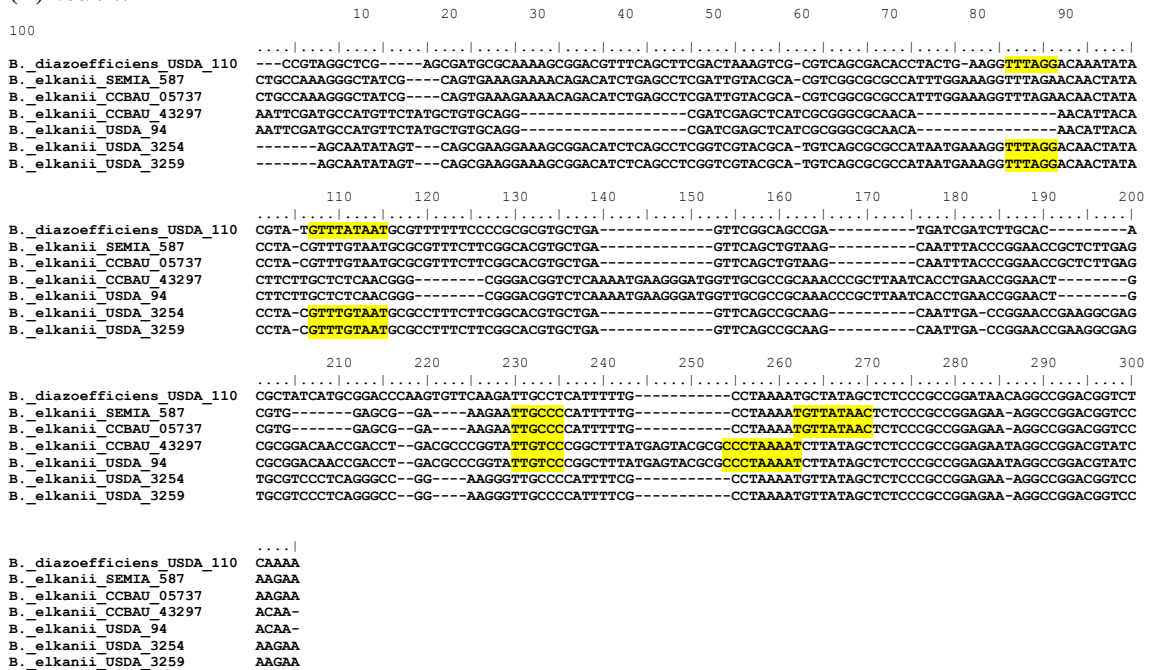
Organism	Gene			
	ID	Coordinates	nt	aa
<i>Bradyrhizobium elkanii</i> USDA 3259	<i>nodW</i> *	AXAW01000069.1_7664909_7665259	351	116
	<i>nodV</i> *	AXAW01000069.1_7661913_7664585	2673	890
	<i>nolZ</i>	-	-	-
	<i>nolY</i>	-	-	-
	<i>nolA</i> *	AXAW01000040.1_5585979_5586551	573	190
	<i>nodD2</i> *	AXAW01000040.1_5587618_5588609	992	329
	<i>nodD1</i> *	AXAW01000040.1_5589229_5590065	837	278
	<i>nodY</i>	-	-	-
	<i>nodK</i>	AXAW01000040.1_5590375_5590638	264	88
	<i>nodA</i>	AXAW01000040.1_5591033_5591665	633	210
	<i>nodB</i>	AXAW01000040.1_5591866_5592321	456	151
	<i>nodC</i>	AXAW01000040.1_5592339_5593728	1386	461
	<i>nodS</i>	AXAW01000040.1_5593827_5594354	528	175
	<i>nodU</i>	AXAW01000040.1_5594368_5596077	1710	569
	<i>nodI</i>	AXAW01000040.1_5596079_5596999	921	306
	<i>nodJ</i>	AXAW01000040.1_5597003_5597794	792	263
	<i>nolM</i>	-	-	-
	<i>nolN</i>	-	-	-
	<i>nolO</i>	AXAW01000040.1_5598075_5600102	2028	675
	<i>nodZ</i>	AXAW01000040.1_5600383_5601377	995	331

*regulatory *nod* genes*nodD₁* and *nodD₂*, the core regulatory genes in *nod* regulon are highlighted in grey

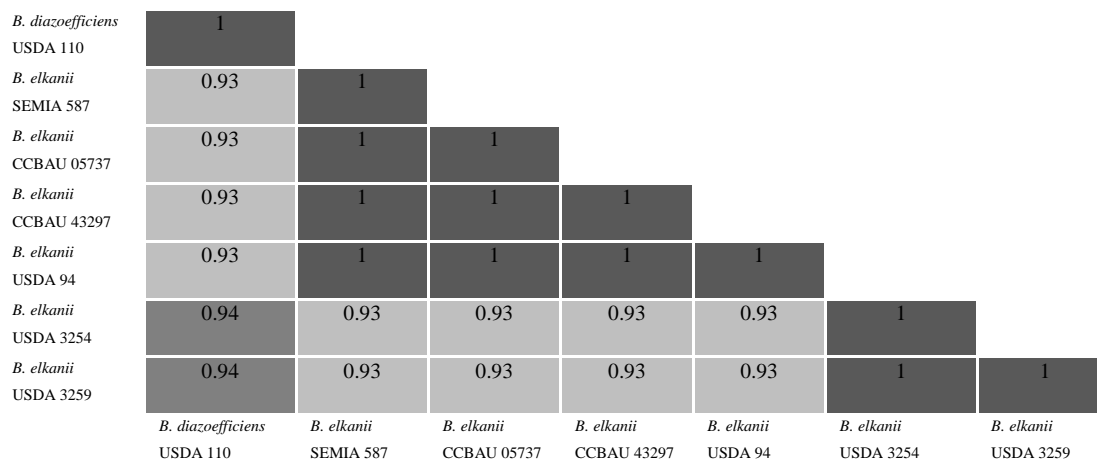
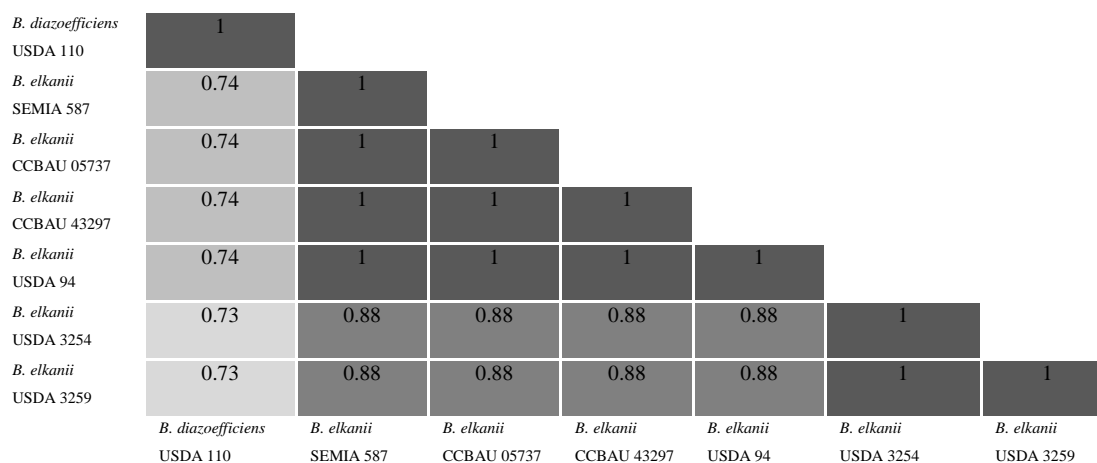
(A) *nolA*



(B) *nodVW*

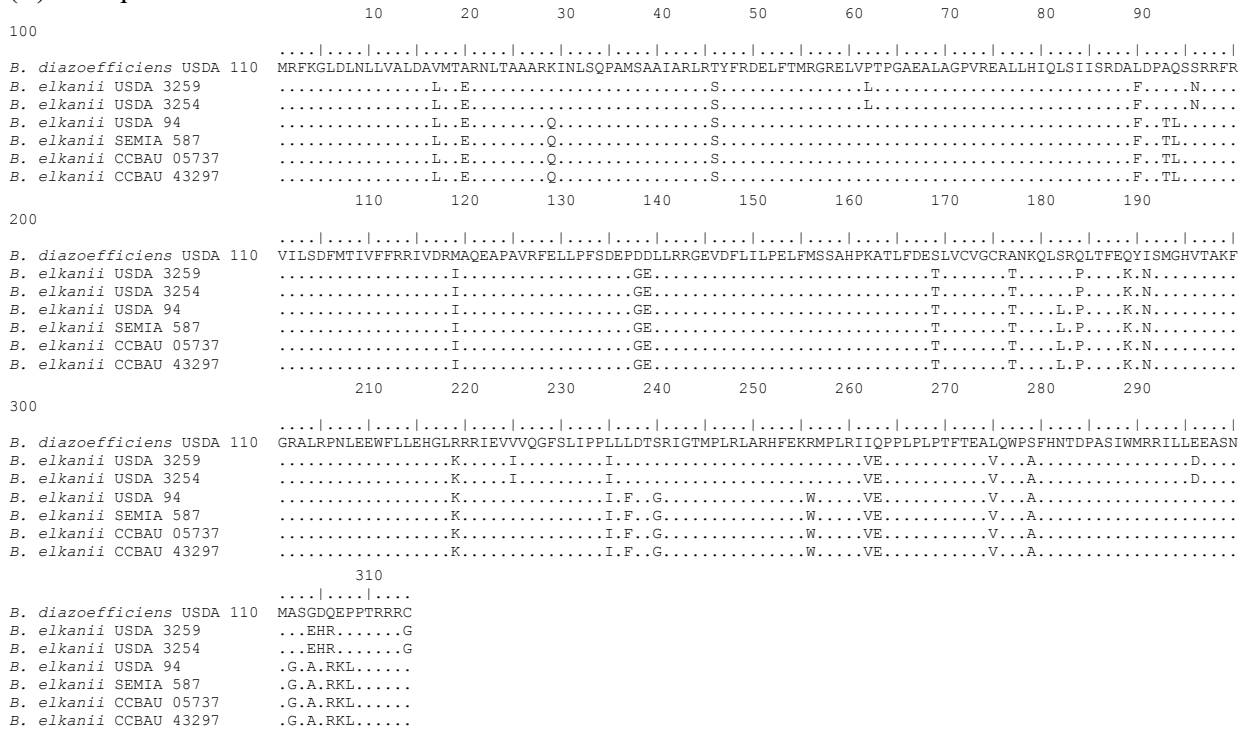


Supplementary Figure 1. Alignment of 250 bp upstream ORF region from *nolA* (A) and *nodVW* (B) genes. Yellow dashed sequences show a $-10/-35 \sigma^{70}$ potential promoter. Putative *nod* box motifs ATC-N₉-GAT were not found in these promoter regions. Last nucleotide in the sequence represents that immediately prior to ATG from CDS start site.

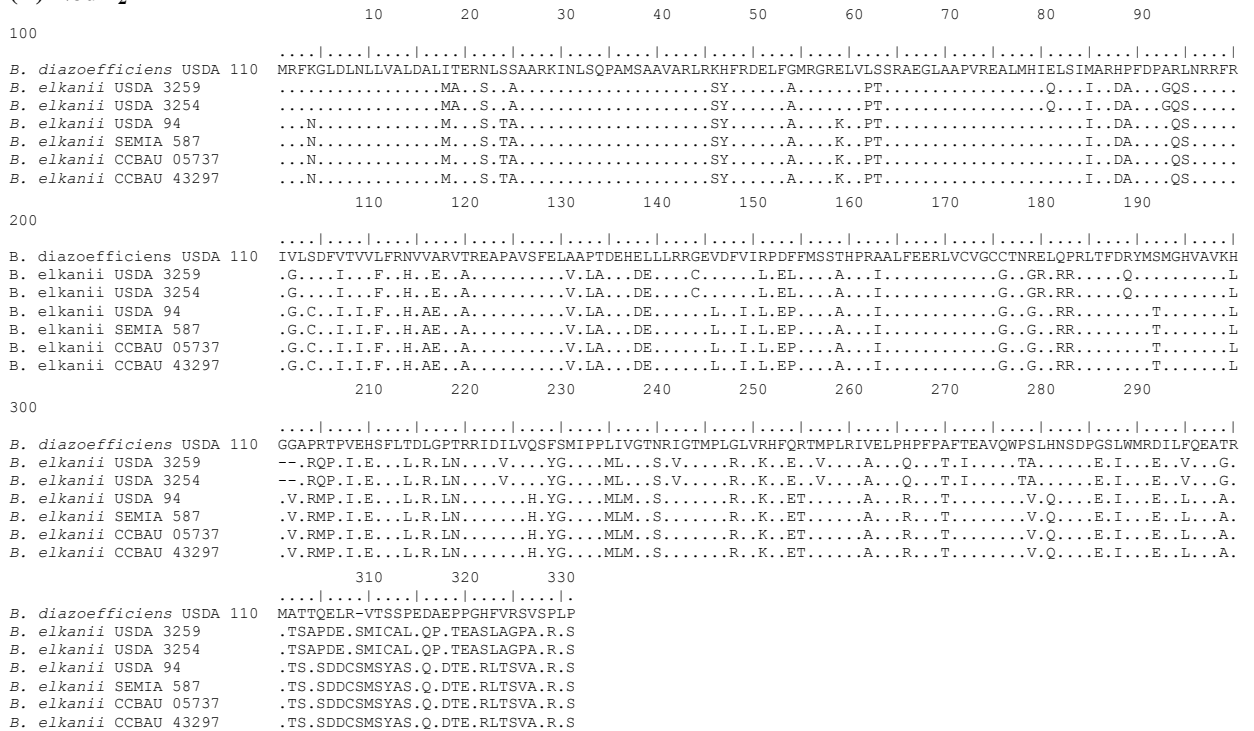
(A) NodD₁**(B) NodD₂**

Supplementary Figure 2. Global similarity of NodD regulator proteins in *B. elkanii* strains using Percent Accepted Mutations (PAM250) matrix. From dark- to light gray shaded boxes are the higher- to lower sequence similarities, respectively. Analysis was performed for NodD₁ (A), and NodD₂ (B) protein sequences.

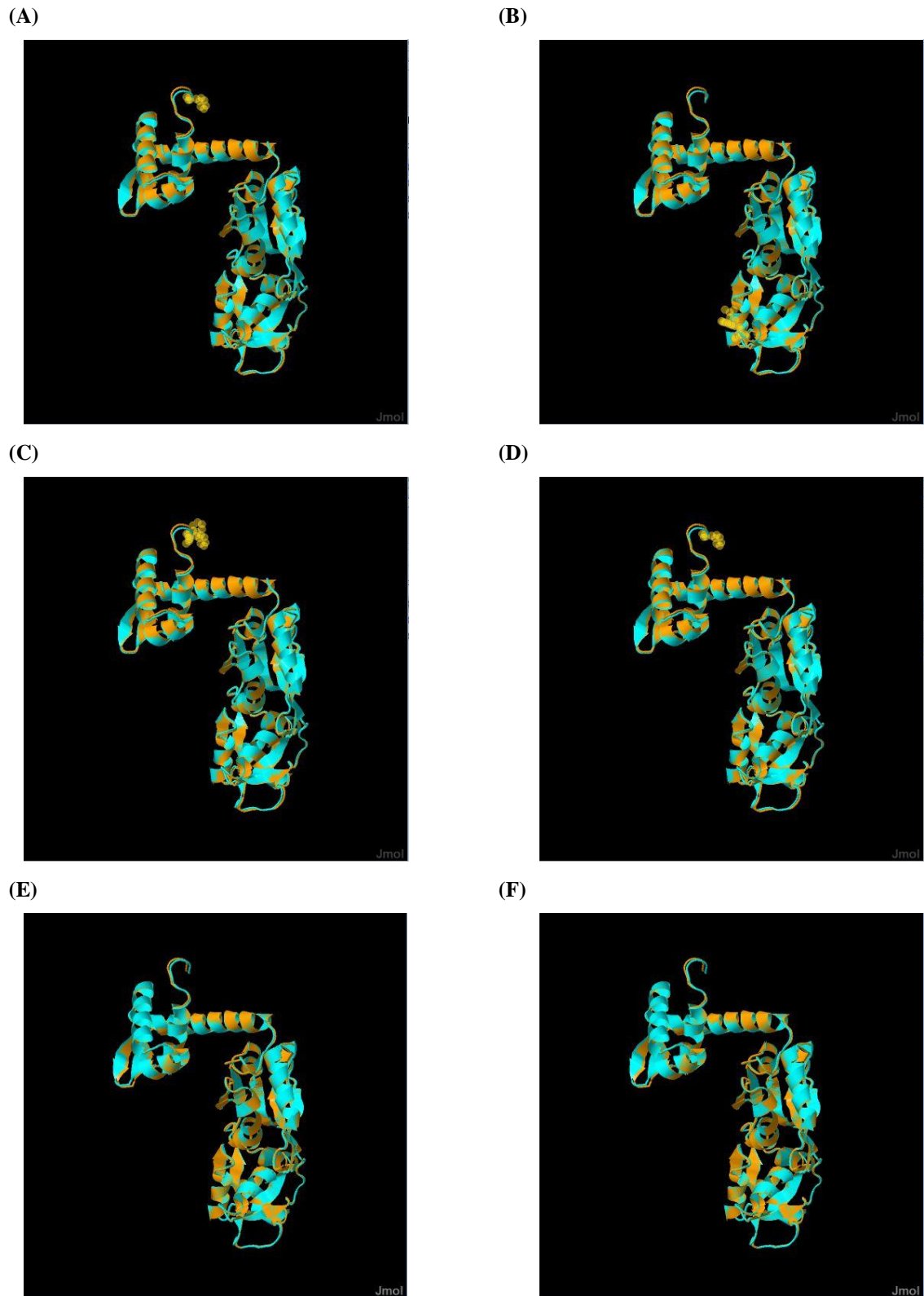
(A) NodD₁



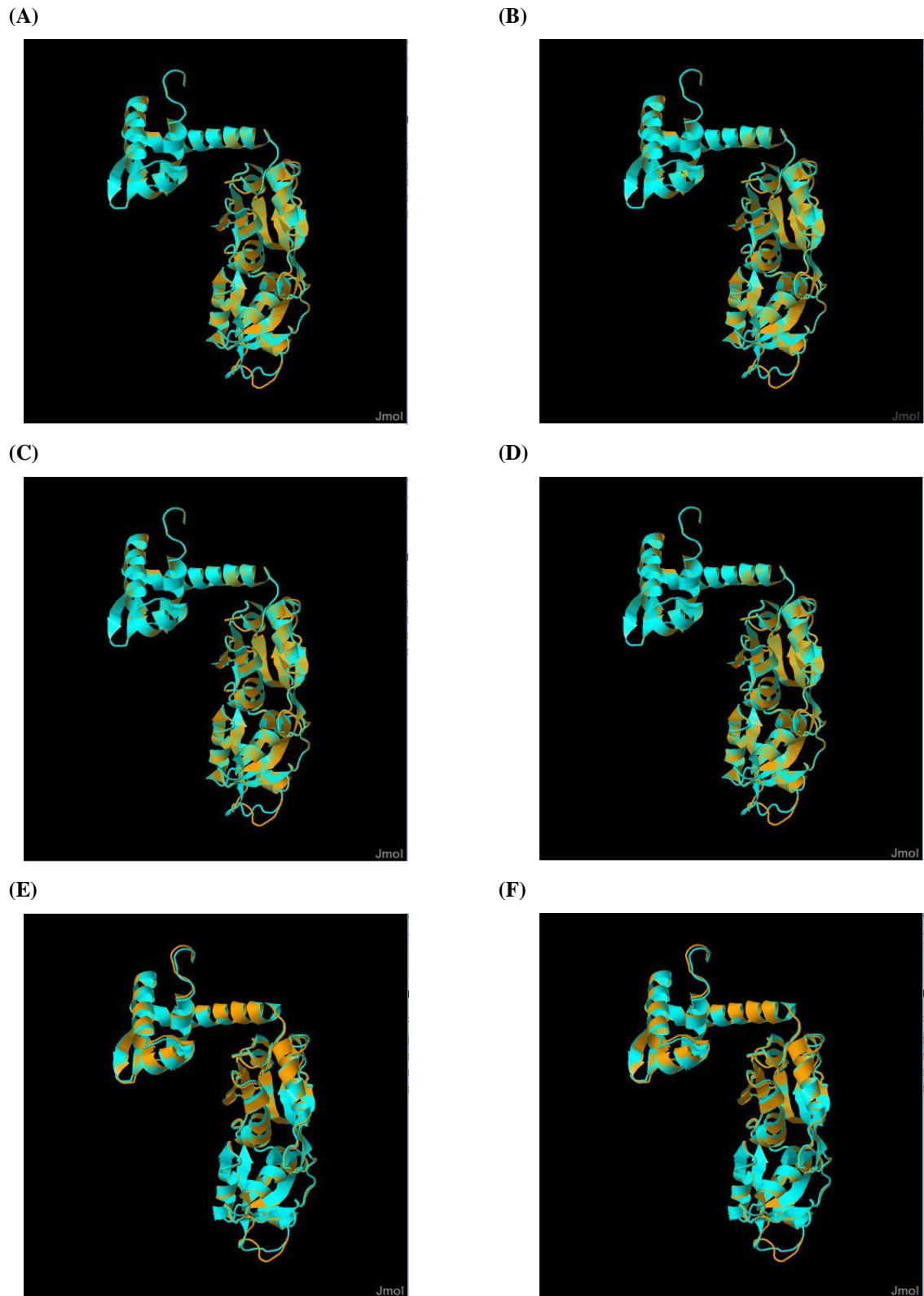
(B) NodD₂



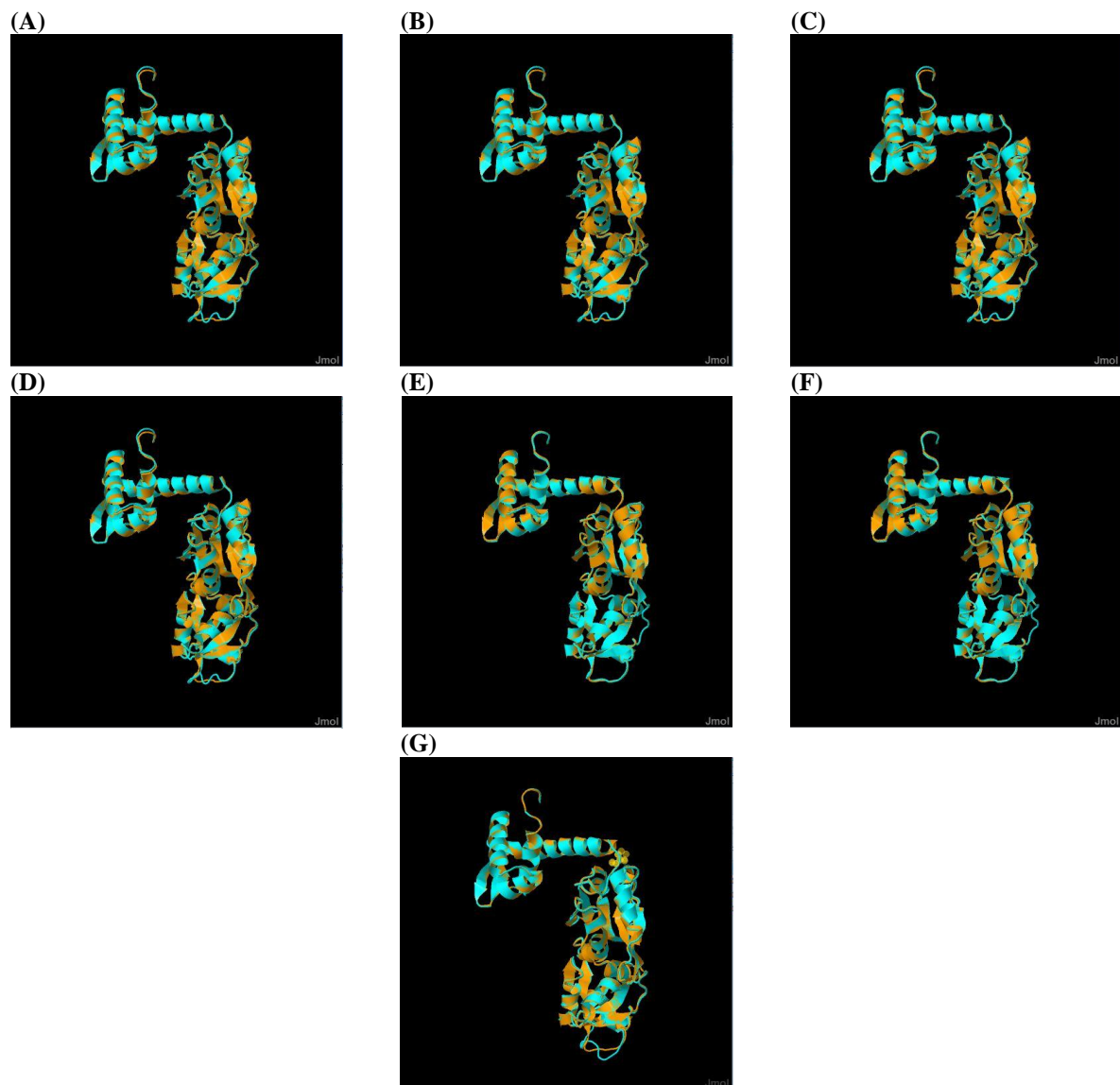
Supplementary Figure 3. Conservation plot from the multiple sequence alignment (MSA) of NodD regulator proteins from genomes of different *B. elkani* strains in contrast to reference *B. diazoefficiens* USDA 110. Global protein alignment was performed using MUltiple Sequence Comparison by Log-Expectation (MUSCLE) algorithm for NodD₁ (A), and NodD₂ (B) protein sequences.



Supplementary Figure 4. NodD₁ global protein structure alignment from *B. elkanii* strains against *B. diazoefficiens* USDA 110 using Combinatorial Extension (CE) algorithm. Orange colour residues represent NodD₁ from *B. diazoefficiens* USDA 110, superimposed by light-blue residues representing NodD₁ from *B. elkanii* SEMIA 587 (A), *B. elkanii* CCBAU 05737 (B), *B. elkanii* CCBAU 43297 (C), *B. elkanii* USDA 94 (D), *B. elkanii* USDA 3254 (E), and *B. elkanii* USDA 3259 (F).



Supplementary Figure 5. NodD₂ global protein structure alignment from *B. elkanii* strains against *B. diazoefficiens* USDA 110 using Combinatorial Extension (CE) algorithm. Orange colour residues represent NodD₂ from *B. diazoefficiens* USDA 110, superimposed by light-blue residues representing NodD₂ from *B. elkanii* SEMIA 587 (A), *B. elkanii* CCBAU 05737 (B), *B. elkanii* CCBAU 43297 (C), *B. elkanii* USDA 94 (D), *B. elkanii* USDA 3254 (E), and *B. elkanii* USDA 3259 (F).



Supplementary Figure 6. NodD₁-NodD₂ global protein structure alignment within each Bradyrhizobium genome using Combinatorial Extension (CE) algorithm. Orange colour residues represent NodD₁, superimposed by light-blue residues representing NodD₂. Genomes represented are *B. elkanii* SEMIA 587 (A), *B. elkanii* CCBAU 05737 (B), *B. elkanii* CCBAU 43297 (C), *B. elkanii* USDA 94 (D), *B. elkanii* USDA 3254 (E), *B. elkanii* USDA 3259 (F), and *B. diazoefficiens* USDA 110 (G).

4. CONSIDERAÇÕES FINAIS E PERSPECTIVAS

Como mencionado na Introdução desse trabalho, a imensa quantidade de informação gerada pelas metodologias de sequenciamento de alto rendimento tornou-se um desafio para os cientistas que buscam interpretar e dar um significado biológico para esse conjunto de dados. Para se ter uma ideia dessa quantidade, até o momento em que esse trabalho foi redigido cerca de 30.000 genomas bacterianos estavam disponíveis nos bancos de dados, seja na forma de genomas completos (fechados) ou de rascunhos (ou *drafts*, distribuído em vários *contigs*). Apenas em relação a espécies do gênero *Bradyrhizobium*, foco desse trabalho, 88 genomas podiam ser facilmente acessados. Entretanto, para a maioria desses genomas a anotação dos genes e das proteínas correspondentes é incompleta ou mesmo inexistente, o que dificulta imensamente análises do tipo das que foram realizadas na presente dissertação.

Mesmo que existam 10 genomas de linhagens de *B. elkanii* disponíveis nos bancos de dados, decidimos utilizar somente aqueles de linhagens que confirmadamente foram identificadas como pertencentes à espécie em estudo e que compartilhavam da mesma metodologia de montagem. Por isso, as análises foram executadas com a utilização de seis genomas, cujas anotações de seus genes e proteínas tiveram que ser realizadas, para a grande maioria, na sua totalidade. Esse trabalho, embora sendo uma etapa inicial, foi de fundamental importância para que pudéssemos elaborar a primeira figura do manuscrito (Figura 1), onde mostramos, pela primeira vez para a espécie de *B. elkanii*, uma comparação da organização dos genes envolvidos no processo de nodulação (genes *nod* e *nol*) para diferentes linhagens dessa espécie. Surpreendentemente constatamos que em vários genomas a organização encontrada na espécie usada como referência (*B. diazoefficiens* USDA 110) não foi mantida na sua totalidade, ainda que o principal operon *nod* tenha sido conservado. Inclusive alguns genes presentes na espécie referência foram perdidos nos genomas de *B. elkanii* (genes *nolZ*, *nolM* e *nolN*, por exemplo), ou mesmo substituídos, como foi o caso entre os genes *nodY* (em USDA 110) e *nodK* (nas linhagens de *B. elkanii*). Dentre os seis genomas de *B. elkanii* analisados, o da linhagem SEMIA 587 foi o que mais divergiu em relação à organização dos genes *nod* e *nol*, sugerindo que, no genoma dessa linhagem, vários eventos de recombinação gênica devem ter ocorrido e que culminaram em um maior espaçamento dos genes em questão no cromossoma. Essa

diferença na organização dos genes envolvidos no processo de nodulação pode sugerir uma diferença na eficiência desse processo, ou, então, uma mudança na discriminação da planta hospedeira. Experimentos que testem a especificidade a diferentes hospedeiros (diferentes variedades de soja ou outras espécies de leguminosas) e a eficiência na formação de nódulos poderão esclarecer essa questão. Também será importante realizarmos uma análise de sintenia entre os genomas das seis linhagens de *B. elkanii*, a fim de verificarmos a existência de rearranjos em outras regiões do genoma de SEMIA 587 ou se nessa linhagem ocorreu algum problema no momento da montagem do seu genoma.

Outra observação interessante foi a presença, em todos os seis genomas analisados, de duas cópias do gene *nodD*, *nodD₁* e *nodD₂*, com grande similaridade de sequência entre elas. Entretanto, apesar dessa similaridade, verificamos que os motivos de ligação ao DNA hélice-volta-hélice das proteínas correspondentes, NodD₁ e NodD₂, apresentam algumas variações na composição de aminoácidos, o que pode evidenciar uma diferença na eficiência de reconhecimento do *nod box*. Outras regiões também apresentam algumas diferenças que podem influir, por exemplo, na dimerização das subunidades para formar o tetrâmero funcional.

A regulação do processo de nodulação é extremamente complexa e, em *Bradyrhizobium*, envolve não só a ação das proteínas NodD₁ e NodD₂, mas também de NodA e NodVW. A fim de contribuirmos com um maior conhecimento desse processo em *B. elkanii*, resolvemos investigar as regiões à montante do início da tradução de cada um dos genes codificadores dessas proteínas (250 pb), buscando por regiões que fossem conservadas e por motivos de ligação de proteínas que pudessem atuar como ativadoras ou repressoras da transcrição dos mesmos. Como foi descrito no manuscrito, dois *nod boxes* foram identificados na região próxima ao início da região codificadora de NodD₁, sendo um deles com sobreposição à possível região utilizada para a ligação da RNA-polimerase complexada com o fator σ^{70} . De acordo com dados da literatura, esse *nod box* (representado na Figura 2A do manuscrito como Box1) controla a transcrição dos genes que estão presentes no operon *nod*, cuja transcrição tem orientação divergente à de *nodD₁*. O outro *nod box* (representado como Box2 na mesma figura) é descrito como um “possível” *nod box* em trabalhos anteriores e esse deve ser o responsável pela ativação da transcrição do gene *nodD₁*. A conservação de sequência na região de 250 pares de bases (pb) analisada entre os seis genomas de *B. elkanii* e com o genoma de *B. diazoefficiens*

USDA 110, também chamou nossa atenção, sugerindo que, nessas espécies, a regulação da transcrição de *nodD₁* possa ser semelhante.

Em relação às sequências reguladoras de *nodD₂*, não foi encontrado o mesmo grau de conservação que encontramos nessa mesma região de *nodD₁*. Para o gene *nodD₂*, as sequências das linhagens SEMIA 587, CCBAU 05737, CCBAU 43297 e USDA 94 apresentam um *nod box* parcial e uma sequência que pode ser utilizada pela RNA-polimerase complexada com o fator σ^{70} para iniciar a transcrição desse gene, além de serem bastante conservadas. Já para as linhagens USDA 3254 e USDA 3259 não foi possível identificarmos um *nod box* nem uma sequência que poderia ser usada como sítio para a ligação da RNA-polimerase complexada com o fator σ^{70} . Isso sugere que possa haver uma diferença na regulação da transcrição de *nodD₂* entre as linhagens de *B. elkanii* e que naquelas onde existe um *nod box* parcial, as proteínas NodD₁ e NodD₂ possam se ligar para estimular ou reprimir a transcrição de *nodD₂*, respectivamente. Embora mencionado na literatura a existência de um sítio de ligação para a proteína Nola na região promotora de *nodD₂*, tal região não foi identificada em nossas análises e, até o momento, não somos capazes de confirmar o papel ativador de Nola na transcrição de *nodD₂*.

As regiões reguladoras dos genes *nola* (altamente conservadas) e *nodVW* (não conservadas) não apresentaram *nod boxes*, indicando que outras proteínas, diferentes de NodD₁ e NodD₂, sejam as responsáveis pela regulação da transcrição desses genes.

Quando comparados os modelos propostos para as proteínas NodD₁ e NodD₂ constatamos que a maior similaridade e homologia se encontra na região que contém o domínio de ligação ao DNA hélice-volta-hélice. Entretanto, mesmo que exista uma grande similaridade entre essas proteínas dentre as espécies estudadas, as diferenças encontradas podem refletir uma maior ou menor afinidade pelo *nod box* ou entre as proteínas para a formação dos dímeros funcionais, como mencionado anteriormente. Essas pequenas diferenças podem resultar em uma maior ou menor ativação dos promotores e, em última instância, em uma maior ou menor produção dos fatores Nod, ou a produção de fatores Nod com cadeias laterais diferentes. Estudos dos perfis metabólicos, por espectrometria de massas, das linhagens de *B. elkanii* em comparação com a linhagem de *B. diazoefficiens* USDA 110, sob a indução de flavonoides específicos, poderão esclarecer essa questão.

Embora tenhamos proposto um modelo para a regulação da transcrição dos genes *nod* em *B. elkanii* com base nas observações desse trabalho e com resultados já descritos

na literatura, vários experimentos ainda devem ser realizados, a fim de comprovarmos a veracidade das hipóteses formuladas. Além dos já mencionados, experimentos de retardamento em gel serão necessários para comprovarmos a ligação das proteínas NodD₁ e NodD₂ nos *nod boxes* identificados nas regiões reguladoras dos respectivos genes, bem como a eficiência de cada ligação, e ensaios de duplo híbrido poderão mostrar se essas proteínas são capazes de formar dímeros funcionais. Também será interessante mapear os aminoácidos essenciais para a dimerização das subunidades e aqueles envolvidos na ativação ou repressão da transcrição. Enfim, esse trabalho é apenas o início de uma grande linha de investigação que poderá contribuir, de forma bastante significativa, para um maior conhecimento do processo de indução da nodulação de leguminosas que é estimulado pelos fatores Nod bacterianos.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- (1994) Biological Nitrogen Fixation: Research Challenges - A Review of Research Grants Funded by the U.S. Agency for International Development Washington, DC, The National Academies Press, pp. 59.
- Abby, S. and Daubin, V. (2007). Comparative genomics and the evolution of prokaryotes. *Trends in Microbiology* **15**, 135-141.
- Alföldi, J. and Lindblad-Toh, K. (2013). Comparative genomics as a tool to understand evolution and disease. *Genome Research* **23**, 1063-1068.
- Alm, E.J., Huang, K.H., Price, M.N., Koche, R.P., Keller, K., Dubchak, I.L. and Arkin, A.P. (2005). The MicrobesOnline Web site for comparative genomics. *Genome Research* **15**, 1015-1022.
- Andrade, D.S., Murphy, P.J. and Giller, K.E. (2002). Effects of liming and legume/cereal cropping on populations of indigenous rhizobia in an acid Brazilian Oxisol. *Soil Biology and Biochemistry* **34**, 477-485.
- Arnold, W., Rump, A., Klipp, W., Priefer, U.B. and Pühler, A. (1988). Nucleotide sequence of a 24,206-base-pair DNA fragment carrying the entire nitrogen fixation gene cluster of *Klebsiella pneumoniae*. *Journal of Molecular Biology* **203**, 715-738.
- Bais, H.P., Weir, T.L., Perry, L.G., Gilroy, S. and Vivanco, J.M. (2006). THE ROLE OF ROOT EXUDATES IN RHIZOSPHERE INTERACTIONS WITH PLANTS AND OTHER ORGANISMS. *Annual Review of Plant Biology* **57**, 233-266.
- Batista, L., Irisarri, P., Rebuffo, M., Cuitiño, M., Sanjuán, J. and Monza, J. (2015). Nodulation competitiveness as a requisite for improved rhizobial inoculants of *Trifolium pratense*. *Biol Fertil Soils* **51**, 11-20.
- Bonato, E.R., Bertagnolli, P.F., Lange, C.E. and Rubin, S.d.A.L. (2000). Teor de óleo e de proteína em genótipos de soja desenvolvidos após 1990. *Pesquisa Agropecuária Brasileira* **35**, 2391-2398.
- Brasil (2011) Instrução Normativa SARC N° 5, de 06 de Agosto de 2004. Ministério da Agricultura Pecuária e Abastecimento.
- Brockwell, J., Pilka, A. and Holliday, R. (1991). Soil pH is a major determinant of the numbers of naturally occurring *Rhizobium meliloti* in non-cultivated soils in central New South Wales. *Australian Journal of Experimental Agriculture* **31**, 211-219.
- Burn, J., Rossen, L. and Johnston, A.W.B. (1987). Four classes of mutations in the nodD gene of *Rhizobium leguminosarum* biovar. *viciae* that affect its ability to autoregulate and/or activate other nod genes in the presence of flavonoid inducers. *Genes & Development* **1**, 456-464.
- Callaham, D.A. and Torrey, J.G. (1981). The structural basis for infection of root hairs of *Trifolium repens* by *Rhizobium*. *Canadian Journal of Botany* **59**, 1647-1664.
- Campos, B.C., Hungria, M. and Tedesco, V. (2001). Eficiência da fixação biológica de N₂ por estirpes de *Bradyrhizobium* na soja em plantio direto. *Revista Brasileira de Ciência do Solo* **25**, 583-592.
- Carlson, R.W., Sanjuan, J., Bhat, U.R., Glushka, J., Spaink, H.P., Wijffjes, A.H., *et al.* (1993). The structures and biological activities of the lipo-oligosaccharide nodulation signals produced by type I and II strains of *Bradyrhizobium japonicum*. *Journal of Biological Chemistry* **268**, 18372-18381.

- Carrareto Alves, L., de Souza, J., Varani, A. and Lemos, E. (2014) The Family Rhizobiaceae. In *The Prokaryotes*, E. Rosenberg, E. DeLong, S. Lory, E. Stackebrandt, F. Thompson (eds.). Springer Berlin Heidelberg, pp. 419-437.
- Chen, X.-C., Feng, J., Hou, B.-H., Li, F.-Q., Li, Q. and Hong, G.-F. (2005). Modulating DNA bending affects NodD-mediated transcriptional control in *Rhizobium leguminosarum*. *Nucleic acids research* **33**, 2540-2548.
- Cowling, E.B. and Galloway, J.N. (2002). Challenges and opportunities facing animal agriculture: Optimizing nitrogen management in the atmosphere and biosphere of the Earth. *Journal of Animal Science* **80**, E157-E167.
- Denarie, J., Debelle, F. and Rosenberg, C. (1992). Signaling and Host Range Variation in Nodulation. *Annual Review of Microbiology* **46**, 497-531.
- Denarie, J., Debelle, F., Truchet, G. and Prome, J.-C. (1993) *Rhizobium* and Legume Nodulation: A Molecular Dialogue. In *New Horizons in Nitrogen Fixation*, R. Palacios, J. Mora, W. Newton (eds.). Springer Netherlands, pp. 19-30.
- Desbrosses, Guilhem J. and Stougaard, J. (2011). Root Nodulation: A Paradigm for How Plant-Microbe Symbiosis Influences Host Developmental Pathways. *Cell Host & Microbe* **10**, 348-358.
- Dixon, R. (1984). The genetic complexity of nitrogen fixation. *Gen. Microbiol* **130**, 2745-2755.
- Dixon, R. and Kahn, D. (2004). Genetic regulation of biological nitrogen fixation. *Nat Rev Micro* **2**, 621-631.
- Erisman, J.W., Sutton, M.A., Galloway, J., Klimont, Z. and Winiwarter, W. (2008). How a century of ammonia synthesis changed the world. *Nature Geosci* **1**, 636-639.
- FAO (2011) Current world fertilizer trends and outlook to 2015. In *Food and Agriculture Organization of the United Nations*. Rome, FAO, pp. 41.
- Fellay, R., Hanin, M., Montorzi, G., Frey, J., Freiberg, C., Golinowski, W., et al. (1998). nodD2 of *Rhizobium* sp. NGR234 is involved in the repression of the nodABC operon. *Molecular Microbiology* **27**, 1039-1050.
- Fisher, R.F., Egelhoff, T.T., Mulligan, J.T. and Long, S.R. (1988). Specific binding of proteins from *Rhizobium meliloti* cell-free extracts containing NodD to DNA sequences upstream of inducible nodulation genes. *Genes & Development* **2**, 282-293.
- Fraysse, N., Couderc, F. and Poinso, V. (2003). Surface polysaccharide involvement in establishing the rhizobium-legume symbiosis. *European Journal of Biochemistry* **270**, 1365-1380.
- Fu, H.A., Hartmann, A., Lowery, R.G., Fitzmaurice, W.P., Roberts, G.P. and Burris, R.H. (1989). Posttranslational regulatory system for nitrogenase activity in *Azospirillum* spp. *Journal of Bacteriology* **171**, 4679-4685.
- Gage, D.J. (2004). Infection and Invasion of Roots by Symbiotic, Nitrogen-Fixing Rhizobia during Nodulation of Temperate Legumes. *Microbiology and Molecular Biology Reviews* **68**, 280-300.
- Galloway, J.N. (1998). The global nitrogen cycle: changes and consequences. *Environmental Pollution* **102**, 15-24.
- Galloway, J.N. and Cowling, E.B. (2002). Reactive Nitrogen and The World: 200 Years of Change. *AMBIO: A Journal of the Human Environment* **31**, 64-71.
- Galloway, J.N., Dentener, F.J., Capone, D.G., Boyer, E.W., Howarth, R.W., Seitzinger, S.P., et al. (2004). Nitrogen Cycles: Past, Present, and Future. *Biogeochemistry* **70**, 153-226.

- Galloway, J.N., Townsend, A.R., Erismann, J.W., Bekunda, M., Cai, Z., Freney, J.R., *et al.* (2008). Transformation of the Nitrogen Cycle: Recent Trends, Questions, and Potential Solutions. *Science* **320**, 889-892.
- Galperin, M. and Koonin, E. (2014) Comparative Genomics Approaches to Identifying Functionally Related Genes. In *Algorithms for Computational Biology*, A.-H. Dediu, C. Martín-Vide, B. Truthe (eds.). Springer International Publishing, pp. 1-24.
- Geurts, R. and Bisseling, T. (2002). Rhizobium Nod Factor Perception and Signalling. *The Plant Cell* **14**, S239-S249.
- Giongo, A., Ambrosini, A., Vargas, L.K., Freire, J.R.J., Bodanese-Zanettini, M.H. and Passaglia, L.M.P. (2008). Evaluation of genetic diversity of bradyrhizobia strains nodulating soybean [*Glycine max* (L.) Merrill] isolated from South Brazilian fields. *Applied Soil Ecology* **38**, 261-269.
- Goethals, K., Van Montagu, M. and Holsters, M. (1992). Conserved motifs in a divergent nod box of *Azorhizobium caulinodans* ORS571 reveal a common structure in promoters regulated by LysR-type proteins. *Proceedings of the National Academy of Sciences* **89**, 1646-1650.
- Göttfert, M., Holzhäuser, D., Bäni, D. and Hennecke, H. (1992). Structural and functional analysis of two different nodD genes in *Bradyrhizobium japonicum* USDA110. *Mol. Plant-Microbe Interact* **5**, 257-265.
- Györgypal, Z., Kondorosi, E. and Kondorosi, A. (1991). Diverse signal sensitivity of NodD protein homologs from narrow and broad host range Rhizobia. *Molecular Plant-Microbe Interactions* **4**, 356-364.
- Haber, F. (2002). The synthesis of ammonia from its elements Nobel Lecture, June 2, 1920. *Reson* **7**, 86-94.
- Halbleib, C.M. and Ludden, P.W. (2000). Regulation of Biological Nitrogen Fixation. *The Journal of Nutrition* **130**, 1081-1084.
- Hardison, R.C. (2003). Comparative Genomics. *PLoS Biol* **1**, e58.
- Hardy, R.W.F., Burns, R.C., Hebert, R.R., Holsten, R.D. and Jackson, E.K. (1971). Biological nitrogen fixation: A key to world protein. *Plant Soil* **35**, 561-590.
- Henikoff, S., Haughn, G.W., Calvo, J.M. and Wallace, J.C. (1988). A large family of bacterial activator proteins. *Proceedings of the National Academy of Sciences* **85**, 6602-6606.
- Hungria, M., Campo, R.J. and Mendes, I.C. (2007) A importância do processo de fixação biológica do nitrogênio para a cultura da soja: componente essencial para a competitividade do produto brasileiro. In *Série Documentos (CNPQ)*. Londrina, Embrapa Soja, pp. 80 p.
- Hungria, M., Campo, R.J. and MENDES, I.d.C. (2001) Fixação biológica do nitrogênio na cultura da soja, Embrapa Soja; Brasília, DF: Embrapa Cerrados.
- Hymowitz, T., Dudley, J.W., Collins, F.I. and Brown, C.M. (1974). Estimations of Protein and Oil Concentration in Corn, Soybean, and Oat Seed by Near-Infrared Light Reflectance. *Crop Science* **14**, 713-715.
- Janczarek, M., Rachwał, K., Marzec, A., Grządziel, J. and Palusińska-Szys, M. (2015). Signal molecules and cell-surface components involved in early stages of the legume-rhizobium interactions. *Applied Soil Ecology* **85**, 94-113.
- Klug, W.S., Cummings, M.R., Spencer, C.A. and Palladino, M.A. (2009) Conceitos de Genética, Artmed Editora.

- Kober, M.d.V., Sá, E.L.S.d., Freire, J.R.J. and Giongo, A. (2004). Characterization of variants of *Bradyrhizobium elkanii* and *B. japonicum* and symbiotic behaviour in soybeans. *Ciência Rural* **34**, 1459-1464.
- Krause, A., Doerfel, A. and Göttfert, M. (2002). Mutational and Transcriptional Analysis of the Type III Secretion System of *Bradyrhizobium japonicum*. *Molecular Plant-Microbe Interactions* **15**, 1228-1235.
- Lima, S.C., Lopes, E.S. and Lemos, E.G.M. (1998). Caracterização de rizóbios (*Bradyrhizobium japonicum*) e produtividade da soja. *Scientia Agricola* **55**, 360-366.
- Lindström, K., Murwira, M., Willems, A. and Altier, N. (2010). The biodiversity of beneficial microbe-host mutualism: the case of rhizobia. *Research in Microbiology* **161**, 453-463.
- Loh, J., Garcia, M. and Stacey, G. (1997). NodV and NodW, a second flavonoid recognition system regulating nod gene expression in *Bradyrhizobium japonicum*. *Journal of Bacteriology* **179**, 3013-3020.
- Loh, J. and Stacey, G. (2003). Nodulation Gene Regulation in *Bradyrhizobium japonicum*: a Unique Integration of Global Regulatory Circuits. *Applied and Environmental Microbiology* **69**, 10-17.
- Loh, J., Stacey, M.G., Sadowsky, M.J. and Stacey, G. (1999). The *Bradyrhizobium japonicum* nolA Gene Encodes Three Functionally Distinct Proteins. *Journal of Bacteriology* **181**, 1544-1554.
- Loh, J.T. and Stacey, G. (2001). Feedback regulation of the *Bradyrhizobium japonicum* nodulation genes. *Molecular Microbiology* **41**, 1357-1364.
- Madsen, L.H., Tirichine, L., Jurkiewicz, A., Sullivan, J.T., Heckmann, A.B., Bek, A.S., *et al.* (2010). The molecular network governing nodule organogenesis and infection in the model legume *Lotus japonicus*. *Nat Commun* **1**, 10.
- McCoy, E. (1932). Infection by *Bact. Radicicola* in Relation to the Microchemistry of the Host's Cell Walls. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character* **110**, 514-533.
- Mergaert, P., Van Montagu, M. and Holsters, M. (1997). The nodulation gene nolK of *Azorhizobium caulinodans* is involved in the formation of GDP-fucose from GDP-mannose. *FEBS Letters* **409**, 312-316.
- Mulligan, J.T. and Long, S.R. (1985). Induction of *Rhizobium meliloti* nodC expression by plant exudate requires nodD. *Proceedings of the National Academy of Sciences of the United States of America* **82**, 6609-6613.
- Noisangiam, R., Teamtisong, K., Tittabutr, P., Boonkerd, N., Toshiki, U., Minamisawa, K. and Teaumroong, N. (2012). Genetic Diversity, Symbiotic Evolution, and Proposed Infection Process of *Bradyrhizobium* Strains Isolated from Root Nodules of *Aeschynomene americana* L. in Thailand. *Applied and Environmental Microbiology* **78**, 6236-6250.
- Olivares, J., Bedmar, E.J. and Sanjuán, J. (2013). Biological Nitrogen Fixation in the Context of Global Change. *Molecular Plant-Microbe Interactions* **26**, 486-494.
- Ozer, E., Allen, J. and Hauser, A. (2014). Characterization of the core and accessory genomes of *Pseudomonas aeruginosa* using bioinformatic tools Spine and AGent. *BMC Genomics* **15**, 737.
- Paffetti, D., Scotti, C., Gnocchi, S., Fancelli, S. and Bazzicalupo, M. (1996). Genetic diversity of an Italian *Rhizobium meliloti* population from different *Medicago sativa* varieties. *Applied and Environmental Microbiology* **62**, 2279-2285.

- Parker, M.A. (1999). Relationships of Bradyrhizobia from the Legumes *Apios americana* and *Desmodium glutinosum*. *Applied and Environmental Microbiology* **65**, 4914-4920.
- Peck, M.C., Fisher, R.F. and Long, S.R. (2006). Diverse Flavonoids Stimulate NodD1 Binding to nod Gene Promoters in *Sinorhizobium meliloti*. *Journal of Bacteriology* **188**, 5417-5427.
- Perret, X., Staehelin, C. and Broughton, W.J. (2000). Molecular Basis of Symbiotic Promiscuity. *Microbiology and Molecular Biology Reviews* **64**, 180-201.
- Pimentel, D. (2009). Energy Inputs in Food Crop Production in Developing and Developed Nations. *Energies* **2**, 1.
- Rangaraj, P., Rüttimann-Johnson, C., Shah, V.K. and Ludden, P.W. (2001). Accumulation of ⁵⁵Fe-Labeled Precursors of the Iron-Molybdenum Cofactor of Nitrogenase on NifH and NifX of *Azotobacter vinelandii*. *Journal of Biological Chemistry* **276**, 15968-15974.
- Recourt, K., van Brussel, A.A., Driessen, A.J. and Lugtenberg, B.J. (1989). Accumulation of a nod gene inducer, the flavonoid naringenin, in the cytoplasmic membrane of *Rhizobium leguminosarum* biovar *viciae* is caused by the pH-dependent hydrophobicity of naringenin. *Journal of Bacteriology* **171**, 4370-4377.
- Sadowsky, M., Graham, P. and Sugawara, M. (2013) Root and Stem Nodule Bacteria of Legumes. In *The Prokaryotes*, E. Rosenberg, E. DeLong, S. Lory, E. Stackebrandt, F. Thompson (eds.). Springer Berlin Heidelberg, pp. 401-425.
- Saleena, L., Loganathan, P., Rangarajan, S. and Nair, S. (2001). Genetic diversity and relationship between Bradyrhizobium strains isolated from blackgram and cowpea. *Biol Fertil Soils* **34**, 276-281.
- Sanjuan, J., Grob, P., Göttfert, M., Hennecke, H. and Stacey, G. (1994). NodW is essential for full expression of the common nodulation genes in *Bradyrhizobium japonicum*. *MPMI-Molecular Plant Microbe Interactions* **7**, 364-369.
- Schimel, D., Stephens, B.B. and Fisher, J.B. (2015). Effect of increasing CO₂ on the terrestrial carbon cycle. *Proceedings of the National Academy of Sciences* **112**, 436-441.
- Schlaman, H.R., Spaink, H.P., Okker, R.J. and Lugtenberg, B.J. (1989). Subcellular localization of the nodD gene product in *Rhizobium leguminosarum*. *Journal of Bacteriology* **171**, 4686-4693.
- Sharma, R.S., Mohammed, A., Mishra, V. and Babu, C.R. (2005). Diversity in a promiscuous group of rhizobia from three *Sesbania* spp. colonizing ecologically distinct habitats of the semi-arid Delhi region. *Research in Microbiology* **156**, 57-67.
- Shiro, S., Matsuura, S., Saiki, R., Sigua, G.C., Yamamoto, A., Umehara, Y., *et al.* (2013). Genetic Diversity and Geographical Distribution of Indigenous Soybean-Nodulating Bradyrhizobia in the United States. *Applied and Environmental Microbiology* **79**, 3610-3618.
- Smil, V. (1999). Detonator of the population explosion. *Nature* **400**, 415-415.
- Spaink, H.P., Wijffelman, C.A., Pees, E., Okker, R.J.H. and Lugtenberg, B.J.J. (1987). *Rhizobium* nodulation gene nodD as a determinant of host specificity. *Nature* **328**, 337-340.
- Stacey, G. (1995). *Bradyrhizobium japonicum* nodulation genetics. *FEMS Microbiology Letters* **127**, 1-9.

- Swanston, C., Homann, P., Caldwell, B., Myrold, D., Ganio, L. and Sollins, P. (2004). Long-term effects of elevated nitrogen on forest soil organic matter stability. *Biogeochemistry* **70**, 229-252.
- Sylvia, D.M. (2005) Principles and Applications of Soil Microbiology, Pearson Prentice Hall.
- Thornton, P.E., Lamarque, J.-F., Rosenbloom, N.A. and Mahowald, N.M. (2007). Influence of carbon-nitrogen cycle coupling on land model response to CO₂ fertilization and climate variability. *Global Biogeochemical Cycles* **21**, GB4018.
- Trentman, M., Dodds, W., Fencl, J., Gerber, K., Guarneri, J., Hitchman, S., *et al.* (2015). Quantifying ambient nitrogen uptake and functional relationships of uptake versus concentration in streams: a comparison of stable isotope, pulse, and plateau approaches. *Biogeochemistry* **125**, 65-79.
- Verma, D. (1992). Signals in Root Nodule Organogenesis and Endocytosis of Rhizobium. *The Plant Cell* **4**, 373-382.
- Verma, D., Miao, G.-H., Lee, N.-G., Cheon, C.-I. and Hong, Z. (1993) Root Nodule Organogenesis and Formation of the Peribacteroid Membrane Compartment. In *Advances in Molecular Genetics of Plant-Microbe Interactions, Vol. 2*, E. Nester, D. Verma (eds.). Springer Netherlands, pp. 343-352.
- Viprey, V., Del Greco, A., Golinowski, W., Broughton, W.J. and Perret, X. (1998). Symbiotic implications of type III protein secretion machinery in Rhizobium. *Molecular Microbiology* **28**, 1381-1389.
- Vitousek, P.M., Aber, J.D., Howarth, R.W., Likens, G.E., Matson, P.A., Schindler, D.W., *et al.* (1997). Technical Report: Human Alteration of the Global Nitrogen Cycle: Sources and Consequences. *Ecological Applications* **7**, 737-750.
- Wang, D., Yang, S., Tang, F. and Zhu, H. (2012). Symbiosis specificity in the legume – rhizobial mutualism. *Cellular Microbiology* **14**, 334-342.
- Wang, S.P. and Stacey, G. (1991). Studies of the Bradyrhizobium japonicum nodD1 promoter: a repeated structure for the nod box. *Journal of Bacteriology* **173**, 3356-3365.