



Bradyrhizobium elkanii nod regulon: insights through genomic analysis

Luciane M. P. Passaglia

Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil

Abstract

A successful symbiotic relationship between soybean [*Glycine max* (L.) Merr.] and *Bradyrhizobium* species requires expression of the bacterial structural *nod* genes that encode for the synthesis of lipochitooligosaccharide nodulation signal molecules, known as Nod factors (NFs). *Bradyrhizobium diazoefficiens* USDA 110 possesses a wide nodulation gene repertoire that allows NF assembly and modification, with transcription of the *nodYABCSUIJnoIMNodZ* operon depending upon specific activators, *i.e.*, products of regulatory *nod* genes that are responsive to signaling molecules such as flavonoid compounds exuded by host plant roots. Central to this regulatory circuit of *nod* gene expression are NodD proteins, members of the LysR-type regulator family. In this study, publicly available *Bradyrhizobium elkanii* sequenced genomes were compared with the closely related *B. diazoefficiens* USDA 110 reference genome to determine the similarities between those genomes, especially with regards to the *nod* operon and *nod* regulon. Bioinformatics analyses revealed a correlation between functional mechanisms and key elements that play an essential role in the regulation of *nod* gene expression. These analyses also revealed new genomic features that had not been clearly explored before, some of which were unique for some *B. elkanii* genomes.

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

Received: January 30, 2017; Accepted: February 16, 2017.

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 (Somasegaran and Hoben, 1994). Genetic and physiological information, including biochemical profile, DNA homology, and phylogenomics, is essential for clarifying the differences among isolates and for supporting the taxonomic classification of over 25 species in the genus *Bradyrhizobium* (Jordan, 1982; Rivas *et al.*, 2009). Among the species of this genus, considerable research efforts have focused on *B. japonicum* and *B. elkanii* because of their commercial use as a source for inoculant formulations, with many effective long-term programs for elite strain identification and selection being undertaken in different countries (Campos *et al.*, 2001; Kober *et al.*, 2004; Melchiorre *et al.*, 2011).

The characterization of symbiosis itself is not an easy task because of the high degree of specialization involved in this phenomenon, particularly since a restricted 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 initial bacterial contact with the plant root to late nodule development and nitrogen fixation, resulting in biological nitrogen fixation that varies according to the host-microsymbiont combination (Schumpp and Deakin, 2010; Wang *et al.*, 2012). In view of this, 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. The establishment of a successful mutualistic relationship initially requires products of the bacterial structural *nod* genes that encode for the synthesis of lipochitooligosaccharide (LCO) nodulation signaling molecules, also known as Nod factors (NFs). Subsequently, the recognition of symbiotic NFs by the plant triggers a signaling cascade that ultimately allows bacterial infection and induces *de novo* organogenesis of the nodule to accommodate the symbiont and further support nitrogen fixation (Tóth and Stacey, 2015). *Bradyrhizobium diazoefficiens* USDA 110 (formerly *B. japonicum* USDA 110) has a wide nodulation gene repertoire involved in NF assembly and modification.

In several species of rhizobia *nod* genes are frequently organized in an operon (*nod* operon), which suggests that regulation of their expression involves common mech-

Send correspondence to Luciane M. P. Passaglia. Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul (UFRGS), Av. Bento Gonçalves, 9500, 91501-970, Porto Alegre, RS, Brazil. E-mail address: luciane.passaglia@ufrgs.br

anisms (Denarie *et al.*, 1993). Indeed, transcription of the *nodYABCUSIJnolMNO* operon in *B. diazoefficiens* USDA 110 depends upon transcriptional activators, *i.e.*, products of regulatory *nod* genes responsive to signaling molecules, such as flavonoid compounds, exuded by host plant roots (Luka *et al.*, 1993; Loh and Stacey, 2003). Central to this regulatory circuit of *nod* gene expression are NodD proteins, members of the LysR-type regulator family. Upon activation by a particular flavonoid ligand NodD proteins can bind to specific DNA motifs upstream of the *nod* operon, the so called *nod* boxes, and selectively control the expression of structural *nod* genes in the early stages of plant-bacteria interaction (Hong *et al.*, 1987; Henikoff *et al.*, 1988; Goethals *et al.*, 1992). Although multiple isoforms of NodD proteins have been identified in distinct rhizobial species, perhaps indicative of a role in expanding the plant host spectrum of these symbionts, only two were found in the *B. diazoefficiens* USDA 110 genome, namely, NodD₁ and NodD₂, products of the *nodD*₁ and *nodD*₂ genes, respectively (Gottfert *et al.*, 1989). These two proteins show distinct expression patterns and play different functional roles in regulating the expression of structural *nod* genes. Active NodD₁, *i.e.*, in the presence of a flavonoid ligand molecule such as genistein in soybean root exudates, operates as a positive transcriptional regulator of the *nod* operon in *B. diazoefficiens*. Unique to 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, the efficiency of this process depends on appropriate spatial and temporal expression of these genes. Hence, it is not surprising that, in addition to positive transcriptional regulators, *nod* genes are also controlled by repressor elements, as is the case for NodD₂ protein, that acts as a negative regulator of the *nod* operon (Loh and Stacey, 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, the roles of Nola, a MerR-type regulator encoded by *nola*, and NodVW, the product of *nodVW*, that form a two-component regulatory system are particularly noteworthy. Initially identified as a soybean genotype-specific nodulation factor, Nola was later shown to be an activator of *nodD*₂ involved in negative regulation with NodD₂, with both involved in the feedback and quorum regulation of *nod* genes (Sadowsky *et al.*, 1991; Garcia *et al.*, 1996; Loh and Stacey, 2003). On the other hand, the NodVW two-component regulatory system provides 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 the soybean inoculant industry, strains of *B. japonicum* and *B. elkanii* differ markedly

in their physiology and in their competitive fitness (Minamisawa, 1989; Vasilas and Furhmann, 1993). Such differences indicate the need to identify and characterize the genetic nature of specificity in the symbiotic relationship as a crucial step in developing guided strategies to enhance the effectiveness of soybean inoculants, given that biological nitrogen fixation ultimately provides a resource for more sustainable agricultural systems.

While there is a considerable amount of knowledge regarding the genetics and molecular mechanisms of the soybean symbiont *B. diazoefficiens*, including the complete genome sequence, information on the genetics of *B. elkanii* is restricted mainly to research focused on specific features. Although genomic data and a few drafts of the genome are available for *B. elkanii*, comparative genomic analyses of these two species have not been reported. In this study, publicly available *B. elkanii* genomes were compared with the closely related *B. diazoefficiens* USDA 110 reference genome (Kaneko *et al.*, 2002) to gain some insights into the mechanisms of *nod* gene expression in *B. elkanii*, especially for the *nod* operon and *nod* regulon. The results of this analysis should provide a more comprehensive understanding of the 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 the complete genome of *B. diazoefficiens* USDA 110 (reference genome) and draft genomes of *B. elkanii* strains SEMIA 587, CCBAU 05737, CCBAU 43297, USDA 94, USDA 3254 and USDA 3259 were obtained from the publicly available database of The National Center for Biotechnology Information (NCBI; <http://www.ncbi.nlm.nih.gov>). Genome features and accession numbers are described in Table 1. The type-strain taxonomy of each organisms genome was confirmed by 16S rRNA sequence analysis using the 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 the EzTaxon server database (Kim *et al.*, 2012).

Bioinformatics analyses

Functional annotation was done with the Rapid Annotation using Subsystem Technology (RAST) server (Aziz *et al.*, 2008), with Glimmer set for gene calling (Salzberg *et al.*, 1998), frameshift correction, backfilling of gaps and automatic fixing errors. Assigned functional features were triple-checked with InterProScan (Zdobnov and Apweiler, 2001) by the signature-recognition method in the InterPro database (Hunter *et al.*, 2009), ScanProsite (de Castro *et al.*, 2006) for protein signature matches in the PROSITE database (Sigrist *et al.*, 2010), and BLASTp

Table 1 - Characteristics of the genomes analysed.

| | <i>B. diazoefficiens</i> | <i>B. elkanii</i> | <i>B. elkanii</i> | <i>B. elkanii</i> | <i>B. elkanii</i> | <i>B. elkanii</i> | <i>B. elkanii</i> |
|-------------------------|--------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| | USDA 110 | SEMIA 587 | CCBAU 05737 | CCBAU 43297 | USDA 94 | USDA 3254 | 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 of 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 of 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 |

against the UniProtKB database (Magrane and Consortium, 2011). An inventory of genes involved in nodulation (structural- and regulatory-*nod/nol* genes) for each genome in this study is provided in Table S1. The *B. diazoefficiens* USDA 110 genome (Kaneko *et al.*, 2002) was used as the reference genome and missing *nod/nol* genes were searched for in other genomes with BLASTn using homologous nucleotide sequences of the closely-related reference species. Possible frameshift annotation errors in assigned genes were corrected and Open Reading Frames (ORFs) were checked with the Expert Protein Analysis System (ExpASY) translate tool (Gasteiger *et al.*, 2003) by comparing with the respective reference.

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

All bioinformatics analyses used the default parameters of the respective software.

Phylogenetic and sequence analyses

Multiple sequence alignment of NodD proteins was done by distance estimation using *kmer* counting and progressive alignment with log-expectation scores, followed by refinement using tree-dependent restricted partitioning implemented by MUSCLE (Edgar, 2004). The algorithm was implemented in the Molecular Evolutionary Genetics Analysis - MEGA 6.0 package (Tamura *et al.*, 2013) and the parameters were set for the 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 done using the Neighbor-Joining method in the same package, with the molecular distances of the aligned sequences computed based on p-distance parameters and 1,000 bootstrap replicates and pairwise deletion treatment for gaps. Point accepted mutation (PAM) 250 calculations were used as a substitution matrix model for scoring sequence alignments (Dayhoff and Schwartz, 1978).

Protein structure prediction and alignment

Protein 3D structure prediction was done using the SWISS-MODEL web server (Biasini *et al.*, 2014) based on evolutionarily-related structures, amino acid sequences and protein structure homologies. The technique uses hidden Markov model-sensitive searches run against the SWISS-MODEL template library (SMTL) to generate a structural model of the protein of interest. A QMEAN potential to assess model quality is then generated with an independent accuracy evaluation by the Continuous Automated Model EvaluatiOn project – CAMEO (Haas *et al.*, 2013) based on target sequences pre-released by the Protein Data Bank (PDB) (Berman *et al.*, 2007). The predicted protein structures were aligned with a combinatorial extension (CE) algorithm (Shindyalov and Bourne, 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). The parameters were set as 30 for the 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

Annotation of the *B. elkanii* genome followed by manual curation revealed a conserved operon structure and organization of the *nodKABC**SUIJnolOnodZ* genes in strains CCBAU 05737, CCBAU 43297, USDA 94, USDA 3254 and USDA 3259 similar to that in the *B. diazoefficiens* USDA 110 *nod* operon, with the substitution of *nodY* for the corresponding *nodK* and the lack of *nolMN* genes. Curiously, in *B. elkanii* SEMIA 587, a distinct pattern for this “canonical” gene organization was observed, with these genes scattered in the chromosome and organized as *nodKABC**S*, a second sparse block containing *nodI*, *nodJ*, *nolO* and *nodZ*, and then *nodU*, which was separate from the other genes of this operon. As in the case of *nolMN*, *nolZ* was not identified in the genomes of any *B. elkanii* strain; likewise, *nolY* was not detected in the genomes of strains USDA 94, USDA 3254, and USDA 3259 (Figure 1).

Annotation also highlighted the presence of *nodD*₁ and *nodD*₂ regulatory genes in all six *B. elkanii* genomes and in the *B. diazoefficiens* USDA 110 genome. The organization of these two genes in the *B. elkanii* genomes followed the pattern observed for *B. diazoefficiens* USDA 110, *i.e.*, they were positioned close to each other and close to the *nod* operon, although in opposite orientation. Similarly, the presence of *nodVW* that coded for the two-component regulatory elements was ubiquitous in all six *B. elkanii* genomes, with its location relative to the *nod* operon varying according to each genome. Conversely, the *nolA* gene showed a conserved location among the *B. elkanii* genomes, close to *nodD*₂; the exception was for strain

SEMIA 587, in which this gene was located at a position distant from *nodD*₂ (Figure 1).

Analysis of *nodD* promoter regions

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

On the other hand, analysis of the 250-bp region upstream of *nodD*₂ revealed that only the *B. elkanii* SEMIA 587, CCBAU 05737, CCBAU 43297 and USDA 94 genomes exhibited a -10/-35 σ^{70} potential promoter with the sequence TTGTCG-N₁₃-CTGTAAAAG, along with a partial *nod* box-like sequence located 152 bp upstream from the putative σ^{70} promoter. In the *B. diazoefficiens* USDA 110 and *B. elkanii* USDA 3254 and USDA 3259 genomes, neither of these elements could be identified in the *nodD*₂ promoter region, *i.e.*, they apparently had no potential -10/-35 σ^{70} promoters nor a partial (or complete) *nod* box-like sequence (Figure 2B).

The *nolA* promoter regions (250 bp) of all genomes 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, the *nolA* promoter regions displayed a high degree of sequence similarity among genomes (Figure S1A). Potential -10/-35 σ^{70} promoter sequences for the *nodVW* regulatory gene were found in all genomes, although there were differences in the structure and location of these promoters. In contrast to *nolA*, the promoter regions of the *nodVW* genes differed considerably in sequence conservation among genomes (Figure S1B). A search for the *nod* box in the *nolA* and *nodVW* promoter regions revealed the absence of this regulatory motif in the 250-bp upstream gene sequences in all genomes (Figure S1A,B).

Sequence analysis of regulatory Nod proteins

All the *B. elkanii* genomes displayed NodD₁ ORFs of 314 amino acids, in agreement with the size of the NodD₁ protein of *B. diazoefficiens* USDA 110. In contrast, there was discrete variation in the size of the NodD₂ ORFs among genomes, with ORFs of 330 amino acids in *B.*

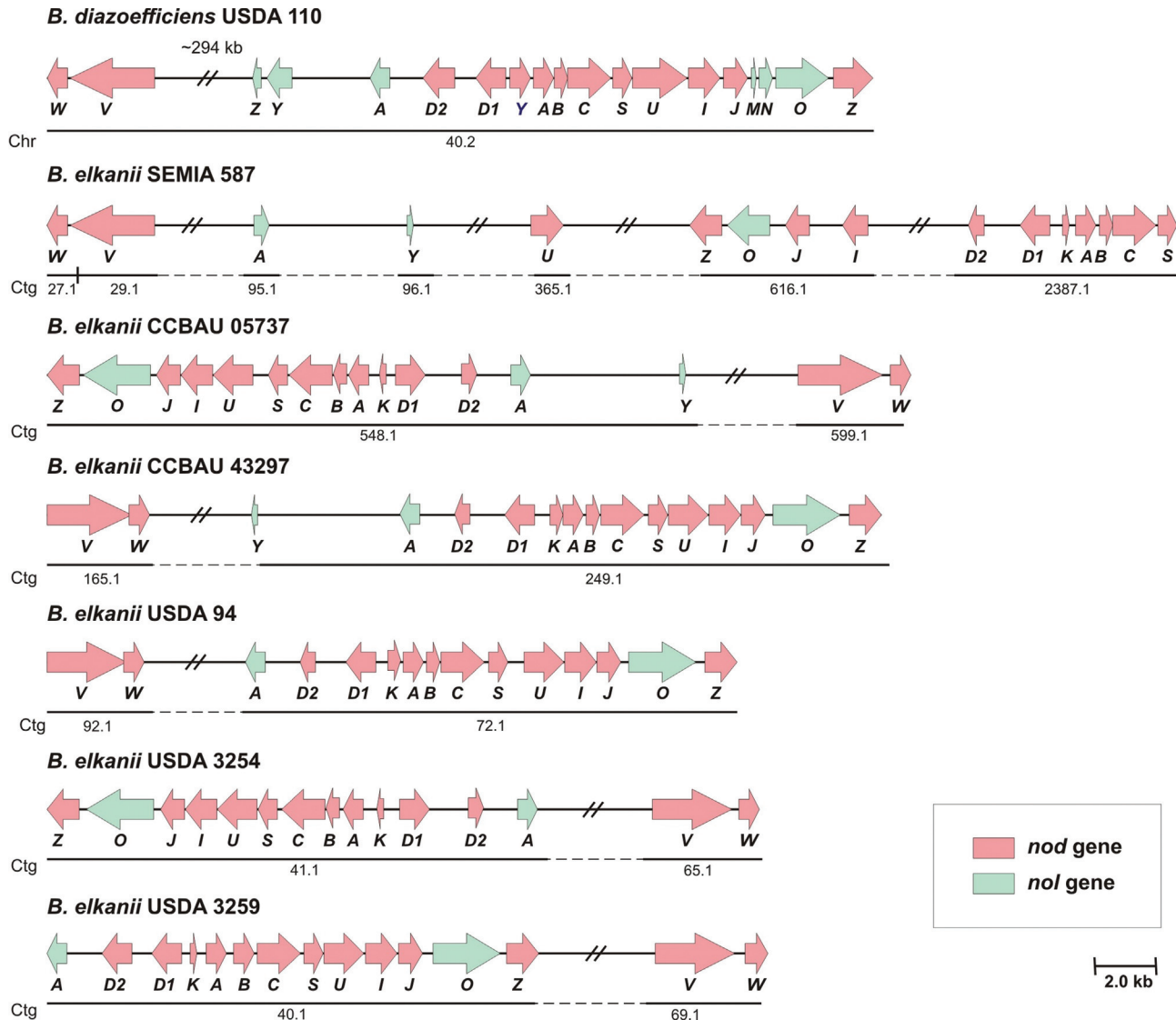


Figure 1 - Genetic map showing organization of the *nod/nol* genes in *B. diazoefficiens* USDA 110 (reference genome) and six *B. elkanii* genomes. Letters indicate the name of each gene in the genome. Gene classes are color coded and transcriptional orientation is indicated by the arrows. Relative gene distances are indicated or follow the scale presented. Gene location within the genome is represented by a straight line below the gene map of the respective chromosome (*B. diazoefficiens*) or contig (*B. elkanii*) ID.

diazoefficiens USDA 110, 331 amino acids 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 (Table S1).

NodD₁ and NodD₂ showed a relatively high degree of conservation, although some dissimilarity among organisms ultimately clustered each *B. elkanii* protein into one of two groups (Figure S2). Indeed, a PAM250 matrix showed that *B. elkanii* strains SEMIA 587, CCBAU 05737, CCBAU 43297 and USDA 94 shared 100% global sequence similarity among themselves for both NodD₁ and NodD₂; consequently, each protein set for these strains grouped together. The same situation occurred with USDA 3254 and USDA 3259, although sequence similarity decreased by 7% for NodD₁ and by 12% for NodD₂ compared

to their respective homologs from other strains within the same species. NodD regulator proteins from the *B. elkanii* strains were still conserved when compared to *B. diazoefficiens* USDA 110 orthologs, especially NodD₁, which exhibited at least 93% sequence similarity, while for NodD₂ no less than 73% similarity was observed (Figure S3A-B). Additionally, conservation plots for NodD₁ and NodD₂ proteins from the *B. elkanii* genomes revealed lower conservation at the carboxy-terminus region of the proteins, especially for NodD₂, when compared to the reference genome of *B. diazoefficiens* USDA 110 (Figure S4A,B).

All the sequences analysed belonged to the LysR family of transcriptional regulators, and exhibited a 57-amino acid helix-turn-helix (HTH) LysR-type domain lo-

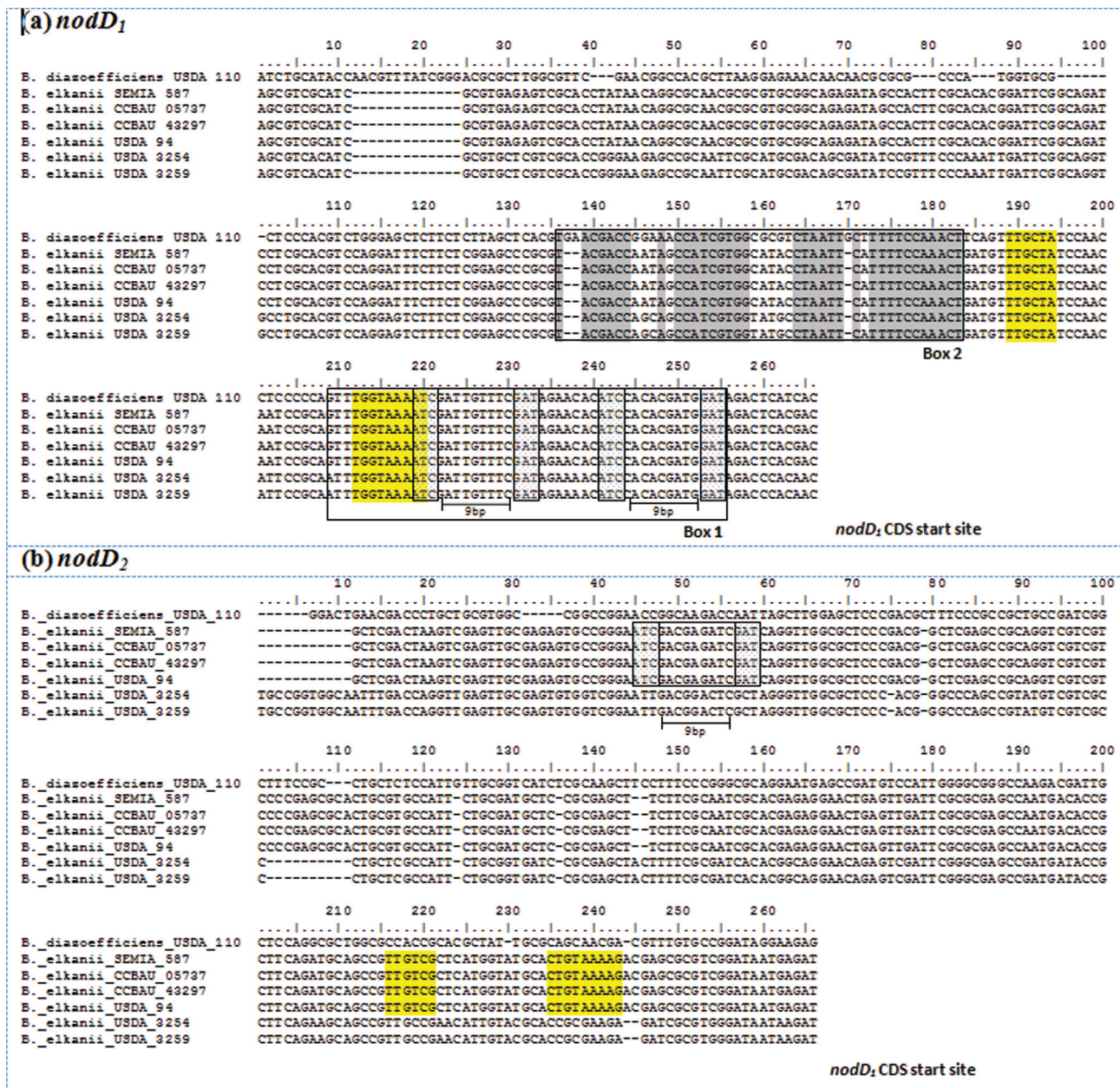


Figure 2 - Alignment of the ~250-bp region upstream from *nodD* ORFs in *Bradyrhizobium* genomes. (a) The *nodD*₁ *nod* box sequences are boxed and show a consensus *nod* box sequence near the *nodD*₁ ORF start codon (Box 1) and a presumptive *nod* box-like sequence (Box 2) upstream, with conserved residues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

cated at position 6-63. These domains also contained a 20-amino acid HTH DNA-binding motif from residues 23-42 in the amino-terminus region of NodD₁ and NodD₂ proteins. The NodD₁ DNA-binding motif from *B. elkanii* strains SEMIA 587, CCBAU 05737, CCBAU 43297 and USDA 94 was highly conserved and showed 100% identity with the sequence LTAAARQINLSQPAMSAIA (Figure 3A). Curiously, *B. elkanii* USDA 3254 and USDA 3259 showed 100% identity with the *B. diazoefficiens* USDA

110 NodD₁ DNA-binding motif LTAAARKINLSQPAMSAIA, with a single amino acid substitution occurring at position 29, in which glutamine (Q) was replaced by lysine (K) (Figure 3A).

The *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 LTAAARKINLSQPAMSAIA, while for strains USDA 3254 and USDA 3259 a serine (S) replaced

the threonine (T) at position 24 of NodD₂, resulting in the DNA-binding motif LSAAARKINLSQPAMSAAVA (Figure 3B). A comparison of NodD₂ among *Bradyrhizobium* species showed that *B. diazoefficiens* USDA 110 shared higher similarity with *B. elkanii* strains USDA 3254 and USDA 3259, although an additional amino acid residue divergence at position 25 was observed, with replacement of an alanine (A) in *B. elkanii* for a serine (S) in *B. diazoefficiens*, providing the sequence LSSAARKINLSQPAMSAAVA (Figure 3B).

In silico analysis of the subcellular location of NodD₁ and NodD₂ proteins did not predict the presence of any significant ordinary signal peptide cleavage site in their amino acid sequences or any twin-arginine signal peptide cleavage site. Lipoprotein signal peptides and non-classical protein secretory patterns were also not identified. Additional protein topology and signal peptide examination did not identify any transmembrane helices. Finally, calculations using an integrative approach algorithm indicated that NodD₁ and NodD₂ were more likely to be cytoplasmic components rather than membrane-bound proteins.

Structural analysis of regulatory Nod proteins

Global protein structure alignment of NodD₁ from the *B. elkanii* strains revealed high identity and similarity to NodD₁ protein from the reference genome of *B. diazoefficiens* USDA 110. In particular, the genomes of *B. elkanii* strains USDA 3254 and USDA 3259 showed 92.62% and 97.65% structural identity and similarity for NodD₁ compared to the orthologous protein from *B. diazoefficiens* USDA 110, respectively. Although slightly lower, the corresponding values for NodD₁ structural identity and similarity from *B. elkanii* strains SEMIA 587, CCBAU 05737, CCBAU 43297 and USDA 94 were 91.61% and 95.97%, respectively. Analysis of the NodD₁ structural alignments indicated that the higher identity and similarity of *B. elkanii* strains USDA 3254 and USDA 3259 with *B. diazoefficiens* USDA 110 resided mainly in the HTH motif of the DNA-binding domain (Figure S5A-F). In contrast, NodD₂ global protein structure alignments for the *B. elkanii* SEMIA 587, CCBAU 05737, CCBAU 43297 and USDA 94 genomes showed higher identity (75.17%) and similarity (84.56%) with the NodD₂ ortholog from *B.*

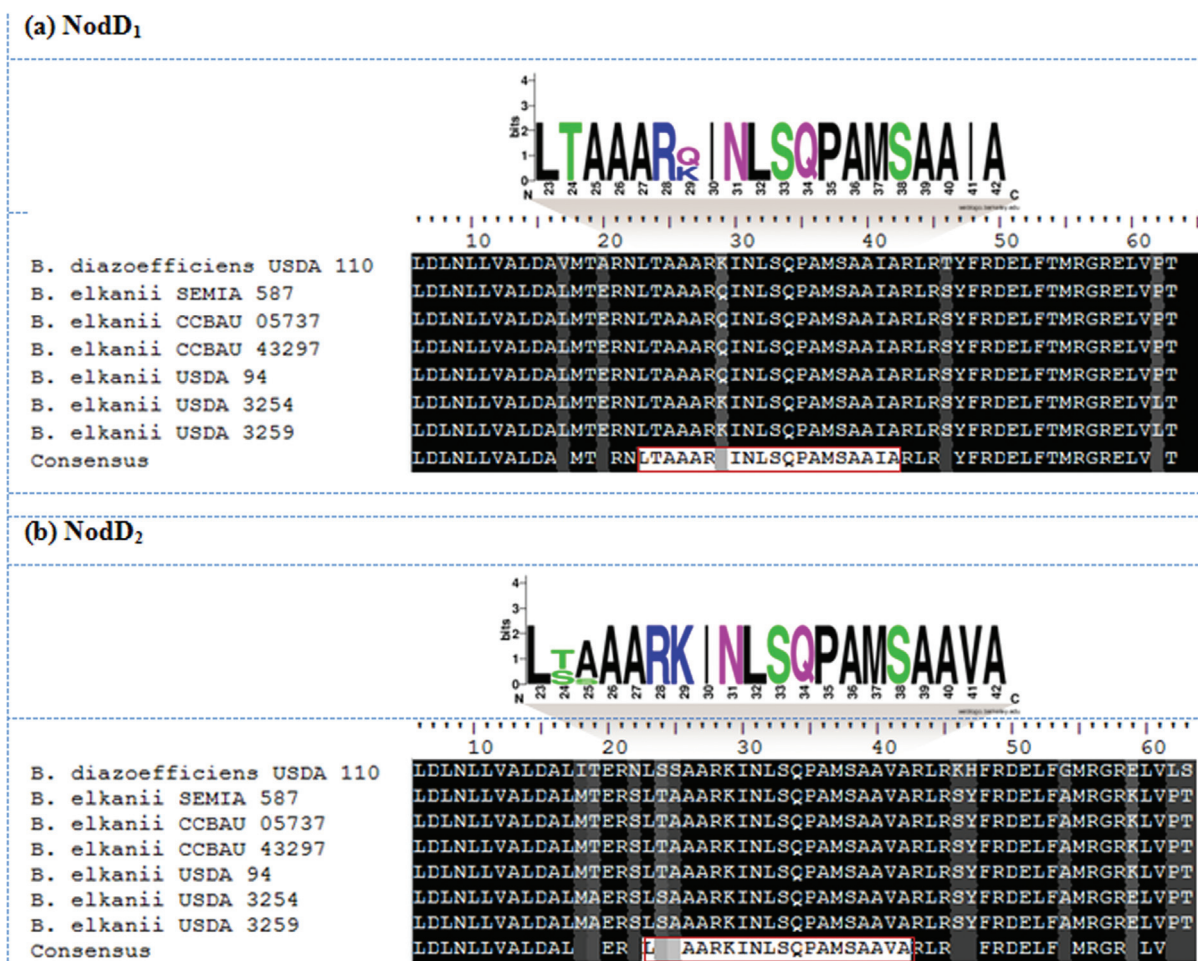


Figure 3 - Multiple sequence alignment (MSA) of the *B. elkanii* NodD HTH LysR-type domain and sequence logo showing conservation of the HTH DNA-binding motif. The HTH DNA-binding sequence conserved among organisms is boxed in red. The sequences are for the (a) NodD₁ and (b) NodD₂ domains.

diazoefficiens USDA 110 than with the *B. elkanii* USDA 3254 and USDA 3259 genomes, for which the identity and similarity were no higher than 74.16% and 82.89%, respectively (Figure S6A-F). Additionally, even though substantial identity and similarity were observed in the structural alignment of NodD₂ between *Bradyrhizobium* species, the values were lower compared to the structural alignments for NodD₁ proteins.

When the global structural alignment of NodD₁ vs. NodD₂ proteins was compared within each genome, the identity and similarity were still relatively significant. In the *B. elkanii* USDA 3254 and USDA 3259 genomes for example, the NodD₁ vs. NodD₂ structural alignment showed 70.47% identity and 81.88% similarity, while for the *B. elkanii* SEMIA 587, CCBAU 05737, CCBAU 43297 and USDA 94 genomes, the identity and similarity reached values of 69.46% and 81.54%, respectively. The *B. diazoefficiens* USDA 110 reference genome showed the most dissimilar structural alignment of NodD₁ vs. NodD₂ and consequently had the lowest identity (63.76%) and similarity (80.54%) values (Figure S7A-F). However, considering only the 57-residue protein segment corresponding to the DNA-binding domain containing the HTH motif, the NodD₁ vs. NodD₂ structural alignment values were higher in all genomes, being identical in the *B. elkanii* USDA 3254 and USDA 3259 genomes, *i.e.*, 100% identity and similarity. The *B. elkanii* SEMIA 587, CCBAU 05737, CCBAU 43297 and USDA 94 genomes also showed better structural alignment of the amino-terminus region that accommodated the DNA-binding domain for the NodD₁ vs. NodD₂ comparison, with 91.38% identity and 98.28% similarity, compared to *B. diazoefficiens* USDA 110, which showed

values of 81.03% and 93.10% for identity and similarity, respectively (Figure 4).

Discussion

nod operon/regulon in *Bradyrhizobium* genomes

Lipochitooligosaccharides (LCOs) produced by the action of bacterial nodulation (*nod*, *nol*, *noe*) gene products are key signaling molecules in establishing the *Bradyrhizobium*-legume symbiosis that triggers the formation of a new organ (the nodule) in which biological nitrogen fixation occurs (Day *et al.*, 2000). *Bradyrhizobium diazoefficiens* and *B. elkanii* can nodulate soybean plants. Despite this common ability, detailed genetic and biochemical studies involving different *Bradyrhizobium* species have shown distinct physiological profiles that ultimately 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 the *B. diazoefficiens* USDA 110 and *B. elkanii* genomes. Moreover, the presence of the structural *nodY/KABCSUIJnolOnodZ* genes, arranged in an operon in the *B. diazoefficiens* and most *B. elkanii* genomes, may indicate that similar patterns of LCO formation can be expected for these organisms, depending primarily upon a particular operon regulation in each genome.

The organization of *nod* genes in clusters/operons is a common and often conserved feature among various rhizobia genomes, many of them with a chromosomal location, as in the case of *B. diazoefficiens*, thereby facilitating coordination of the expression of these respective genes (Schlaman *et al.*, 1998). Although *nod* genes in *Bradyrhizobium*

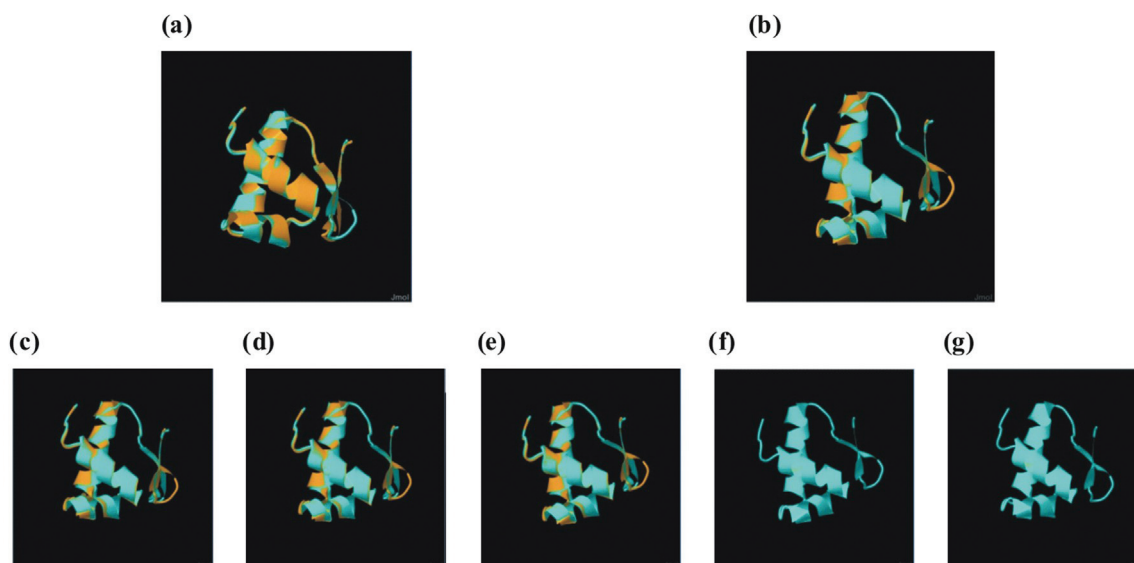


Figure 4 - Structural alignment of the NodD₁ and NodD₂ 57-residue HTH domain within each genome using the Combinatorial Extension (CE) algorithm. The alignments are for *B. diazoefficiens* USDA 110 (a), *B. elkanii* SEMIA 587 (b), *B. elkanii* CCBAU 05737 (c), *B. elkanii* CCBAU 43297 (d), *B. elkanii* USDA 94 (e), *B. elkanii* USDA 3254 (f) and *B. elkanii* USDA 3259 (g). Orange residues represent NodD₁ superimposed on light-blue residues representing NodD₂.

are frequently organized in operons, this is not a fixed pattern, as seen in the *B. elkanii* SEMIA 587 genome which showed a reduced *nodKABCS* operon with additional *nod/nol* genes scattered along the chromosome. Interestingly, separation of the genes from this operon may result in the differential regulation of *nod* genes, ultimately providing an altered pattern of LCO formation for this strain compared to other *B. elkanii* genomes, even though this is not mandatory for divergent *nod* gene arrangement (Vázquez *et al.*, 1991). As in most rhizobia, expression of the *nod/nol* operon in *Bradyrhizobium* is dependent on distinct flavonoid compounds produced and secreted by each leguminous host plant, with the quantity and spectrum of these molecules varying according to the age and physiological state of the plant (Phillips and Streit, 1996; Loh and Stacey, 2003; Hassan and Mathesius, 2012).

Regulation of the *nod/nol* operon in *B. diazoefficiens* and *B. elkanii* is essentially under control of the same elements, namely, the *nod* regulon that consists of *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, both of which were invariably located close to each other and in opposite orientation to the *nod* operon in *Bradyrhizobium* genomes, reinforced their importance as key components in the core regulatory mechanism of operonic *nod* gene expression. Indeed, the products of *nodD* genes are assumed to be fundamental in coordinating structural *nod* genes in many symbionts, including *Bradyrhizobium* species (Mulligan and Long, 1989; Honma *et al.*, 1990; Loh and Stacey, 2003; del Cerro *et al.*, 2015). For instance, in *B. diazoefficiens* products of the *nodD*₁ and *nodD*₂ regulatory genes orchestrate *nodYABCSUIJnolMNOnodZ* operon expression by activation or repression, respectively (Loh and Stacey, 2003). Such a regulatory system suggests a common mechanism that may apply for *B. elkanii* genomes with conserved *nod* gene organization.

Besides their divergent transcriptional orientation, components of the LysR-type transcriptional regulator (LTTR) family, such as NodD₁ and NodD₂, also exhibit another characteristic feature, namely, autoregulation (Maddocks and Oyston, 2008), which ultimately indicates that expression of the associated operon largely depends on the regulation of these regulatory genes themselves. Since *B. diazoefficiens nodD* mutants retain a large marginal ability to nodulate soybean plants, alternative transcriptional activators were proposed to take part in this process. Indeed, the identification of a two-component regulatory system consisting of NodV and NodW, products of *nodVW*, sheds some light on this paradox and provides a suitable explanation. In such a system, the sensor kinase component NodV can detect the environmental stimulus (such as 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, suggests a host-specific role for these proteins, probably through recognition of specific flavonoid inducers produced by the host plants in response to the correlate sensor component NodV (Göttfert *et al.*, 1990; Sanjuan *et al.*, 1994). As shown here, we identified this same system in *B. elkanii* genomes, suggesting that both species possess an alternative pathway for *nod* operon activation and possibly a strategy for broadening their respective host ranges.

Whereas successful soybean nodulation depends on timely expression of the structural *nod* gene at the right place and in suitable amounts, it is reasonable to assume that negative regulation also occurs, in addition to activation. Although this negative control in *B. diazoefficiens* is mediated by the *nodD*₂ product that operates as a *nod* operon repressor as mentioned before, this process is assisted by the *nolA* gene product, which establishes 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 a member of the MerR family of transcriptional regulators that activates transcription upon binding to specific DNA motifs and consequently induces DNA binding leading to its appropriate alignment for RNA polymerase positioning and subsequent transcriptional activation (Ansari *et al.*, 1992; Philips *et al.*, 2015). Moreover, *nolA* has the peculiar feature of encoding for the three proteins *NolA*₁, *NolA*₂ and *NolA*₃ that originate from three in-frame ATG start codons, with *NolA*₁ controlling the expression of the other two and being involved in the activation of *nodD*₂ (Loh *et al.*, 1999). This mechanism is also shared with high homology by the *B. elkanii* genomes and, in most cases, shows a very similar gene organization and transcription to *nodD*₁ and *nodD*₂, as previously demonstrated (Dobert *et al.*, 1994). This finding indicates tight *nod* gene regulation in this 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 of *B. diazoefficiens* and their marked similarity with those in the *B. elkanii* genomes, both in terms of structure and organization, we hypothesized that major phenotypic differences in LCO production and secretory patterns in these two species could occur at some transcriptional or post-transcriptional regulatory level. Accordingly, in the presence of the corresponding flavonoid compounds, activated NodD proteins would specifically bind to conserved *cis*-regulatory elements on bacterial DNA, namely, 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 and are located upstream to the *nod* operon and control its expression (Rostas *et al.*, 1986; Nieuwkoop *et al.*, 1987; Goethals *et al.*, 1992). The conservation of such *cis*-regulatory elements and their essentiality in many rhizobia species studied so far strongly suggests that NodD₁ is central in *nod* operon gene expression in these species, including *Bradyrhizobium* (Spaink *et al.*, 1987). The *nod* operon activation by NodD₁ in *B. elkanii* genomes is no exception to this rule and seems to be under a similar, if not identical, regulatory mechanism since the same genetic features, highly conserved in relation to *B. diazoefficiens* USDA 110, are observed in this context.

Besides the activation of *nod* gene expression, an autoregulatory function unique to *Bradyrhizobium* is assigned to *nodD*₁ by its own product NodD₁ upon binding the same isoflavone molecules that activate the *nod* operon. Due to divergent transcriptional orientation, the *nod* box sequences are located upstream from the *nod* operon containing the structural *nod* genes and the *nodD*₁/*nodD*₂ regulatory genes. This location of *nod* boxes in between the *nod* operon and *nodD*₁/*nodD*₂ within the genome complicates the analysis of their regulation. Although autoregulation by NodD₁ is also accomplished by a DNA-binding mechanism, this is reportedly 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 and Stacey, 1991). Indeed, the conservation of this regulatory sequence in *B. elkanii* genomes reinforces its proposed function, a conclusion 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 among the presumed *nod* box-like sequences in distinct *Bradyrhizobium* species, *e.g.*, the slightly smaller size and some nucleotide divergence in *B. elkanii* genomes compared to *B. diazoefficiens*, that ultimately may affect the DNA-binding process itself.

Based on the *nodD*₁ and *nodD*₂ sequence-homology and their resulting description as members of LysR-type regulators, it seems reasonable to also consider a potential autoregulatory role for NodD₂, akin to what can be observed for NodD₁. Previous screening of *Rhizobium japonicum* USDA 191 DNA sequences for the presence of a *nod* box revealed that only the *nodD*₁ promoter region showed high sequence homology 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 were observed (Appelbaum *et al.*, 1988). Although no complete *nod* box has been found in the promoter region of *nodD*₂ in *Bradyrhizobium* genomes, a remarkably 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, the same genomes that contained this

palindromic sequence also displayed a potential -10/-35 σ^{70} binding site. Since this relationship has not been reported before, the precise biological significance of this findings is still unclear and may represent another mechanism (in addition to Nola) by which *nodD*₂ transcription is regulated in this species; this question deserves further investigation.

The current model for *nodD*₂ expression in *B. diazoefficiens* essentially considers the induction of transcription by Nola upon binding to a putative Nola binding-site present upstream to *nodD*₂ (Garcia *et al.*, 1996; Loh and Stacey, 2003). Sequence homology has shown that Nola is a member of the MerR transcriptional regulator protein family that 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 in relation to the respective promoter (Brown *et al.*, 2003). Surprisingly, as shown here, preliminary screening for a potential Nola binding site upstream to *nodD*₂ found no corresponding conserved *cis*-element in this region, contrary to current literature.

NodD protein sequences and structure conservation

In addition to *cis*-regulatory regions in the bacterial genome, *trans*-acting factors can also affect transcriptional regulation. The high degree of protein conservation observed for NodD₁ and NodD₂ among *Bradyrhizobium* genomes highlighted the evolutionary importance of the mechanism by these regulators operate. Indeed, the NodD amino acid sequence showed extensive conservation that ranged from > 90% sequence similarity among NodD proteins from organisms of the same genus to < 50% for NodD sequences from distantly related organisms (Göttfert *et al.*, 1992). Although the global sequence similarity for NodD can vary considerably among different organisms, a general feature is the high-to-low conservation in the amino acid sequence within the protein from the amino-terminus to the carboxy-terminus of the polypeptide. This structural pattern occurs in many rhizobia species and is consistent with the function attributed to each domain of the protein (Burn *et al.*, 1987, 1989; Horvath *et al.*, 1987).

The amino-terminal region, the most conserved part of the protein, anchors the HTH DNA-binding motif responsible for recognition of the *nod* box *cis*-elements in bacterial DNA. Despite this higher conservation in sequence similarity, subtle amino acid substitutions nevertheless occur among *B. elkanii* genomes and may ultimately modify the affinity of this motif for the DNA, thereby altering the pattern of binding and consequent activity of these regulators. On the other hand, the NodD carboxy-terminus region is believed to be primarily involved in protein oligomerization, an important aspect since it is generally accepted that NodD₁ is active in a tetrameric form that arises from the cognate homodimers (Peck *et al.*, 2013). Multiple weak interactions occur among non-specific amino acids in this region of NodD proteins and may contribute to oligo-

merization (Ezezika *et al.*, 2007; Peck *et al.*, 2013). As observed in other rhizobia, the NodD carboxy-terminus region was clearly less conserved in the *B. elkanii* genomes. Even though this lower sequence similarity did not markedly affect protein structure, it may play a role in oligomerization.

Based on the information available for the *B. diazoefficiens* USDA 110 reference genome, 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 features (Figure 5). 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 *nodD₂* suggests a putative NodD₁ binding site in this region, with a possible regulatory effect on *nodD₂* as a transcriptional activation. Such a mechanism may possibly function under higher concentrations of NodD₁ as a consequence of the autore-

gulation of *nodD₁* expression by its respective product. Under these conditions, a surplus of NodD₁ may be “trapped” by this partial *nod* box sequence; NodD₁ may show lower affinity for the regulator protein relative to the conserved consensus *nod* box sequence and ultimately activate transcription of the repressor protein NodD₂ to balance expression of the *nod* operon.

The prediction of highly conserved structures between NodD₁ and NodD₂ and their structural alignment suggested the possible formation of NodD₁-NodD₂ heterodimers since their protein structures are very similar, but no specific amino acid residues responsible for oligomerization have yet been identified. If NodD₁-NodD₂ heterodimers are formed in *Bradyrhizobium*, they may well affect the NodD₁-driven regulation in this symbiont, in addition to interfering with *nod* operon gene expression. Although the data are suggestive of these events, there is still no conclusive evidence for this hypothesis and experimental confirmation of such mechanisms is required to prove the validity of this model.

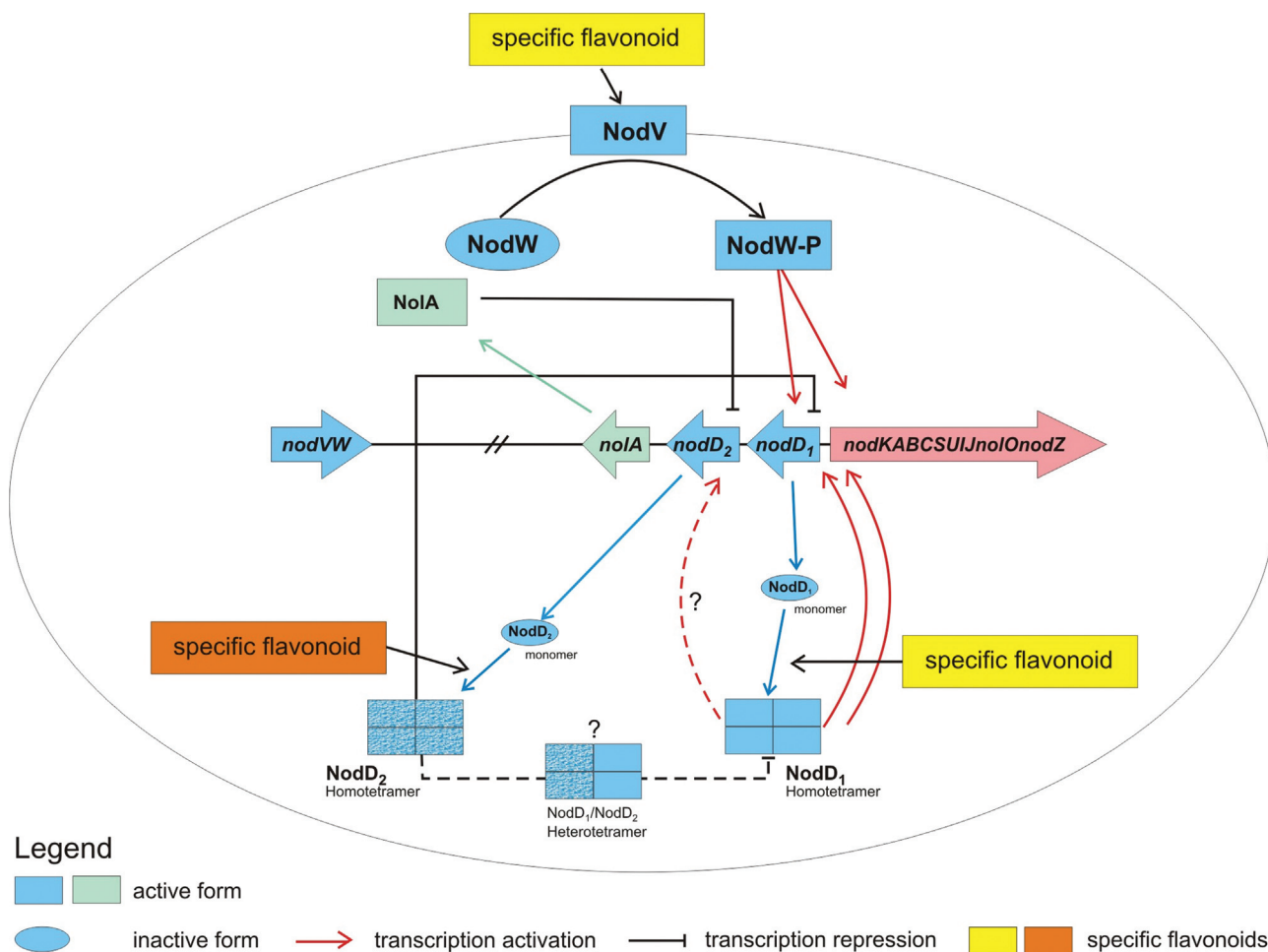


Figure 5 - A model illustrating the modulation of *nod* gene expression proposed for *B. elkanii*. Expression of the *nod* operon containing the structural *nod* genes is regulated by regulatory *nod* genes (*nod* regulon) in the presence of the respective flavonoid inducer. Transcriptional activation of the *nod* operon is mediated by NodD₁ and NodVW, resulting in the biosynthesis of lipochitooligosaccharides. Negative regulation of the *nod* operon is due to the action of NodD₂ and indirectly by NolA. Based on Loh and Stacey (2003).

Our knowledge of the genetic control of nodulation and our understanding of the mechanisms involved in the early events of a new symbiotic relationship have undoubtedly expanded considerably in the last two decades. Although a general framework for nodulation and symbiosis has not yet been developed for all diazotrophic microorganisms, given their genetic and phenotypic diversity, comparative studies may nevertheless help to clarify these processes and provide some useful preliminary evidence for future investigations. As microbial genome sequencing projects are quickly delivering an ever-increasing amount of data, comparative analyses of *cis*-elements and their cognate *trans*-acting factors could provide new insights into the activation and/or repression of *nod* gene expression.

This work has identified several similarities between *B. elkanii* genomes and the closely related *B. diazoefficiens*. Based on these similarities, it was possible to identify and correlate functional mechanisms and key elements that play an essential role in regulating *nod* gene expression. In addition, new genomic features that had not been clearly explored before, some of which were unique to certain *B. elkanii* genomes, have raised new questions for future research.

Acknowledgments

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 AZ, Chael ML and O'Halloran TV (1992) Allosteric underwinding of DNA is a critical step in positive control of transcription by Hg-MerR. *Nature* 355:87-89.
- Appelbaum ER, Thompson DV, 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 RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass E, Kubal M, *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*. *Mol Gen Genet* 214:420-424.
- Bendtsen JD, Jensen LJ, Blom N, von Heijne G and Brunak S (2004) Feature-based prediction of non-classical and leaderless protein secretion. *Prot Eng Des Select* 17:349-356.
- Bendtsen J, Nielsen H, Widdick D, Palmer T and Brunak S (2005) Prediction of twin-arginine signal peptides. *BMC Bioinformatics* 6:167.
- Berman H, Henrick K, Nakamura H and Markley JL (2007) The worldwide Protein Data Bank (wwPDB): Ensuring a single, uniform archive of PDB data. *Nucleic Acids Res* 35:301-303.
- Bernstein FC, Koetzle TF, Williams GJB, Meyer Jr EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T and Tasumi M (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, Kiefer F, Cassarino TG, Bertoni M, Bordoli L, *et al.* (2014) SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* 42:252-258.
- Brown NL, Stoyanov JV, Kidd SP and Hobman JL (2003) The MerR family of transcriptional regulators. *FEMS Microb Rev* 27:145-163.
- Burn J, Rossen L and Johnston AWB (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 Dev* 1:456-464.
- Burn JE, Hamilton WD, Wootton JC and Johnston AW (1989) Single and multiple mutations affecting properties of the regulatory gene *nodD* of *Rhizobium*. *Mol Microbiol* 3:1567-1577.
- Campos BC, Hungria M and Tedesco V (2001) Eficiência da fixação biológica de N₂ por estirpes de *Bradyrhizobium* na soja em plantio direto. *Rev Bras Ciênc Sol* 25:583-592.
- Day RB, Loh J, Cohn JR and Stacey G (2000) Signal exchange involved in the establishment of the *Bradyrhizobium*-legume symbiosis. In: Triplett E (ed) *Prokaryotic Nitrogen Fixation, a Model System for the Analysis of a Biological Process*. Horizon Scientific Press, Norfolk, pp 385-414.
- Dayhoff MO and Schwartz RM (1978) A model of evolutionary change in proteins. In: Dayhoff MO (ed) *Atlas of Protein Sequence and Structure*. Natl Biomed Res Found, Washington DC, pp 345-352.
- de Castro E, Sigrist CA, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A and Hulol N (2006) ScanProsite: Detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res* 34:362-365.
- del Cerro P, Rolla-Santos AAP, Gomes DF, Marks BB, Pérez-Montano F, Rodríguez-Carvajal MA, Nakatani AS, Gil-Serrano A, Megías M, Ollero FJ, *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, Truchet G and Prome J-C (1993) *Rhizobium* and legume nodulation: A molecular dialogue. In: Palacios R, Mora J and Newton W (eds) *New Horizons in Nitrogen Fixation*. Vol 17. Current Plant Science and Biotechnology in Agriculture. Springer, Dordrecht, pp 19-30.
- Doberst RC, Breil BT and Triplett EW (1994) DNA sequence of the common nodulation genes of *Bradyrhizobium elkanii* and their phylogenetic relationship to those of other nodulating bacteria. *Mol Plant Microbe Interact* 7:564-572.
- Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 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.
- Ezekika OC, Haddad S, Neidle EL and Momany C (2007) Oligomerization of BenM, a LysR-type transcriptional regulator:

- Structural basis for the aggregation of proteins in this family. *Acta Crystal Sect 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*. *Mol Plant Microbe Interact* 9:625-636.
- Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD and Bairoch A (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 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. *Proc Natl Acad Sci U S A* 89:1646-1650.
- Goodsell DS, Dutta S, Zardecki C, Voigt M, Berman HM and Burley SK (2015) The RCSB PDB "Molecule of the Month": Inspiring a molecular view of biology. *PLoS Biol* 13:e1002140.
- Gottfert M, Lamb JW, 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. *Mol Gen Genet* 215:407-415.
- 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*. *Proc Natl Acad Sci U S A* 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.
- 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. doi:10.1093/database/bat031.
- Hassan S and Mathesius U (2012) The role of flavonoids in root-rhizosphere signalling: Opportunities and challenges for improving plant-microbe interactions. *J Exp Bot* 63:3429-3444.
- Henikoff S, Haughn GW, Calvo JM and Wallace JC (1988) A large family of bacterial activator proteins. *Proc Natl Acad Sci U S A* 85:6602-6606.
- Hong GF, Burn JE and Johnston AW (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 Res* 15:9677-9690.
- Honma MA, Asomaning M and Ausubel FM (1990) *Rhizobium meliloti nodD* genes mediate host-specific activation of *nodABC*. *J Bacteriol* 172:901-911.
- Horvath B, Bachem CW, 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. *EMBO J* 6:841-848.
- Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, et al. (2009) InterPro: The integrative protein signature database. *Nucleic Acids Res* 37:211-215.
- Jordan DC (1982) Transfer of *Rhizobium japonicum* Buchanan 1980 to *Bradyrhizobium* gen. nov., a genus of slow-growing, root nodule bacteria from leguminous plants. *Int J Sys Evol Microb* 32:136-139.
- Käll L, Krogh A and Sonnhammer ELL (2004) A combined transmembrane topology and signal peptide prediction. *Meth J Mol Biol* 338:1027-1036.
- Kaneko T, Nakamura Y, Sato S, Minamisawa K, Uchiumi T, Sasamoto S, Watanabe A, Idesawa K, Iriguchi M, Kawashima K, et al. (2002) Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA Res* 9:189-197.
- Kim O-S, Cho Y-J, Lee K, Yoon S-H, Kim M, Na H, Park S-C, Jeon YS, Lee J-H, Yi H, et al. (2012) Introducing EzTaxon-e: A prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. *Int J Sys Evol Microb* 62:716-721.
- Kober MdV, de Sá ELS, Freire JRJ and Giongo A (2004) Characterization of variants of *Bradyrhizobium elkanii* and *B. japonicum* and symbiotic behaviour in soybeans. *Ciênc Rural* 34:1459-1464.
- Krogh A, Larsson B, von Heijne G and Sonnhammer ELL (2001) Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J Mol Biol* 305:567-580.
- Loh J and Stacey G (2003) Nodulation gene regulation in *Bradyrhizobium japonicum*: A unique integration of global regulatory circuits. *Appl Environ Microbiol* 69:10-17.
- Loh J, Garcia M and Stacey G (1997) NodV and NodW, a second flavonoid recognition system regulating nod gene expression in *Bradyrhizobium japonicum*. *J Bacteriol* 179:3013-3020.
- Loh J, Stacey MG, Sadowsky MJ and Stacey G (1999) The *Bradyrhizobium japonicum nolA* gene encodes three functionally distinct proteins. *J Bacteriol* 181:1544-1554.
- Luka S, Sanjuan J, Carlson RW and Stacey G (1993) *nolMNO* genes of *Bradyrhizobium japonicum* are co-transcribed with *nodYABCSTUIJ*, and *nolO* is involved in the synthesis of the lipo-oligosaccharide nodulation signals. *J Biol Chem* 268:27053-27059.
- Maddocks SE and Oyston PCF (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 (Oxford) 2011:bar009.
- 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 Cell Physiol* 30:877-884.
- Mulligan JT and Long SR (1989) A family of activator genes regulates expression of *Rhizobium meliloti* nodulation genes. *Genetics* 122:7-18.
- Nieuwkoop AJ, Banfalvi Z, Deshmane N, Gerhold D, Schell MG, Sirotkin KM and Stacey G (1987) A locus encoding host range is linked to the common nodulation genes of *Bradyrhizobium japonicum*. *J Bacteriol* 169:2631-2638.
- Peck MC, Fisher RF, Bliss R and Long SR (2013) Isolation and characterization of mutant *Sinorhizobium meliloti* NodD1 proteins with altered responses to luteolin. *J Bacteriol* 195:3714-3723.

- Petersen TN, Brunak S, von Heijne G and Nielsen H (2011) SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nat Methods* 8:785-786.
- Philips SJ, Canalizo-Hernandez M, Yildirim I, Schatz GC, Mondragon A and O'Halloran TV (2015) TRANSCRIPTION. Allosteric transcriptional regulation via changes in the overall topology of the core promoter. *Science* 349:877-881.
- Phillips DA and Streit WR (1996) Legume signals to rhizobial symbionts: A new approach for defining rhizosphere colonization. In: Stacey G and Keen NT (eds) *Plant-Microbe Interactions*. Vol 1. Springer Science & Business Media, Springer, Berlin, pp 236-271.
- 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 J Microbiol Biotechnol* 24:2377-2382.
- Rivas R, Martens M, de Lajudie P and Willems A (2009) Multi-locus sequence analysis of the genus *Bradyrhizobium*. *Syst Appl Microbiol* 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*. *Proc Natl Acad Sci U S A* 83:1757-1761.
- Sadowsky MJ, Cregan PB, Gottfert M, Sharma A, Gerhold D, Rodriguez-Quinones F, Keyser HH, Hennecke H and Stacey G (1991) The *Bradyrhizobium japonicum nola* gene and its involvement in the genotype-specific nodulation of soybeans. *Proc Natl Acad Sci U S A* 88:637-641.
- Salzberg SL, Delcher AL, Kasif S and White O (1998) Microbial gene identification using interpolated Markov models. *Nucleic Acids Res* 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*. *Mol Plant Microbe Interact* 7:364-369.
- Schlaman HR, Phillips DA and Kondorosi E (1998) Genetic organization and transcriptional regulation of rhizobial nodulation genes. In: Spaik HP, Kondorosi A and Hooykaas P (eds) *The Rhizobiaceae: Molecular Biology of Model Plant-Associated Bacteria*. Springer, Dordrecht, pp 361-386.
- Schumpp O and Deakin WJ (2010) How inefficient rhizobia prolong their existence within nodules. *Trends Plant Sci* 15:189-195.
- Shindyalov IN and Bourne PE (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng* 11:739-747.
- Sigrist CJA, Cerutti L, de Castro E, Langendijk-Genevaux PS, Bulliard V, Bairoch A and Hulo N (2010) PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Res* 38:161-166.
- Somasegaran P and Hoben HJ (1994) *Handbook for Rhizobia: Methods in Legume-Rhizobium Technology*. Springer-Verlag, New York, 450 p.
- Spaik HP, Wijffelman CA, Pees E, Okker RJH and Lugtenberg BJJ (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. *Mol Biol Evol* 30:2725-2729.
- Tóth K and Stacey G (2015) Does plant immunity play a critical role during initiation of the legume-rhizobium symbiosis? *Front Plant Sci* 6:401.
- Vasilas BL and Furhmann JJ (1993) Field response of soybean to nodulation by a rhizobitoxine-producing strain of *Bradyrhizobium*. *Agron J* 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. *J Bacteriol* 173:1250-1258.
- Wang D, Yang S, Tang F and Zhu H (2012) Symbiosis specificity in the legume – Rhizobial mutualism. *Cell Microbiol* 14:334-342.
- Wang Q, Garrity GM, Tiedje JM and Cole JR (2007) Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 73:5261-5267.
- Wang SP and Stacey G (1991) Studies of the *Bradyrhizobium japonicum nodD1* promoter: A repeated structure for the *nod* box. *J Bacteriol* 173:3356-3365.
- Zdobnov EM and Apweiler R (2001) InterProScan – An integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17:847-848.

Supplementary material

The following online material is available for this article:

Table S1 - *Bradyrhizobium nod* operon/regulon gene inventory.

Figure S1 - Alignment of the ~250-bp region upstream from *nola* and *nodVW* ORFs.

Figure S2 - Evolutionary relationship of the *nod* regulatory proteins NodD₁ and NodD₂.

Figure S3 - Global similarity of NodD regulator proteins in *B. elkanii* strains.

Figure S4 - Conservation plot from the multiple sequence alignment of NodD regulator proteins.

Figure S5 - NodD₁ global protein structure alignment for *B. elkanii* strains against *B. diazoefficiens* USDA 110 using the Combinatorial Extension (CE) algorithm.

Figure S6 - NodD₂ global protein structure alignment for *B. elkanii* strains against *B. diazoefficiens* USDA 110 using the Combinatorial Extension (CE) algorithm.

Figure S7 - NodD₁-NodD₂ global protein structure alignment within each *Bradyrhizobium* genome using the Combinatorial Extension (CE) algorithm.

Associate Editor: Ana Tereza R. Vasconcelos

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License (type CC-BY), which permits unrestricted use, distribution and reproduction in any medium, provided the original article is properly cited.