FEMS Microbiology Letters, 366, 2018, fnz046

doi: 10.1093/femsle/fnz046
Advance Access Publication Date: 12 March 2019
Research Letter

RESEARCH LETTER – Environmental Microbiology

Brazilian-adapted soybean *Bradyrhizobium* strains uncover IS elements with potential impact on biological nitrogen fixation

Gesiele Almeida Barros-Carvalho[1,2], Mariangela Hungria[3], Fabrício Martins Lopes[4] and Marie-Anne Van Sluys[1,*,†]

[1]GaTE Lab; Departamento de Botânica - Instituto de Biociências, Universidade de São Paulo, 277 Matão Street, 05508-090, São Paulo, SP, Brazil, [2]Instituto de Matemática e Estatística, Universidade de São Paulo, 1010 Matão Street, 05508-090, São Paulo, SP, Brazil, [3]Embrapa Soja, C.P. 231, 86001–970, Londrina, Pr, Brazil and [4]Universidade Tecnológica Federal do Paraná, 1640 Alberto Carazzai Avenue, 86300-000, Cornélio Procópio, Pr, Brazil

*Corresponding author: Departamento de Botânica – USP, Rua do Matão 277, São Paulo 05508-090, SP, Brazil. Tel: +55-11-3091-7724; E-mail: mavsluys@usp.br

One sentence summary: The contribution of insertion sequences to the diversity of soybean-nodulating *Bradyrhizobium* strains in Brazil.

Editor: Euan James

†Marie-Anne Van Sluys, http://orcid.org/0000-0002-6506-2734

ABSTRACT

Bradyrhizobium diazoefficiens CPAC 7 and *Bradyrhizobium japonicum* CPAC 15 are broadly used in commercial inoculants in Brazil, contributing to most of the nitrogen required by the soybean crop. These strains differ in their symbiotic properties: CPAC 7 is more efficient in fixing nitrogen, whereas CPAC 15 is more competitive. Comparative genomics revealed many transposases close to genes associated with symbiosis in the symbiotic island of these strains. Given the importance that insertion sequences (IS) elements have to bacterial genomes, we focused on identifying the local impact of these elements in the genomes of these and other related *Bradyrhizobium* strains to further understand their phenotypic differences. Analyses were performed using bioinformatics approaches. We found IS elements disrupting and inserted at regulatory regions of genes involved in symbiosis. Further comparative analyses with 21 *Bradyrhizobium* genomes revealed insertional polymorphism with distinguishing patterns between *B. diazoefficiens* and *B. japonicum* lineages. Finally, 13 of these potentially impacted genes are differentially expressed under symbiotic conditions in *B. diazoefficiens* USDA 110. Thus, IS elements are associated with the diversity of *Bradyrhizobium*, possibly by providing mechanisms for natural variation of symbiotic effectiveness.

Keywords: bioinformatics; N-fixing bacteria; symbiotic genomic island; transposable elements

INTRODUCTION

The contribution of biological nitrogen fixation (BNF) to the soybean crop in Brazil is outstanding, as suggested by research done for more than half a century (Hungria *et al.* 2006; Herridge, Peoples and Boddey 2008). Brazilian soils are void of compatible rhizobia and the first *Bradyrhizobium* strains were brought from other countries in the 1950s and 1960s. Nowadays, the two most important strains used in commercial inoculants are *B. japonicum* CPAC 15 (= SEMIA 5079) and *B. diazoefficiens* CPAC 7

Received: 7 November 2018; **Accepted:** 15 May 2019

(= SEMIA 5080). CPAC 15 is a natural variant from strain SEMIA 566, obtained in the 1960s from a nodule of a soybean inoculated with a North American inoculant and belongs to the same serogroup as USDA 123; it was selected for being adapted to the Brazilian Cerrado biome. CPAC 7 is a natural variant of CB 1809, a strain that came from Australia, and was also selected for adaptation to the Brazilian Cerrado biome. These strains are used in commercial inoculants in most of the 30 million hectares cropped with soybean in the country today, reducing input costs, decreasing the contamination of water reservoirs and mitigating the emission of greenhouse gases (Hungria and Mendes 2015).

Symbiotic nitrogen fixation involves a set of specific genes such as those responsible for a complex exchange of molecular signals between the host plant and the bacterium (Van Sluys *et al*. 2002; Oldroyd and Downie 2008; Oldroyd *et al*. 2011), nodulation genes, type III secretion systems and N_2-fixation process *per se* (Loh and Stacey 2003). In *Bradyrhizobium* genomes, the majority of these genes are located within a symbiotic island (Göttfert *et al*. 2001; Kaneko *et al*. 2002; Kaneko *et al*. 2011; Siqueira *et al*. 2014; Iida *et al*. 2015).

Differences in symbiotic performance have been reported between CPAC 7 and CPAC 15, and the two strains are considered to be distinct species within *Bradyrhizobium* (Delamuta *et al*. 2013). CPAC 7 is known for its high efficiency in fixing N_2, while CPAC 15 shows high competitiveness in the ability of establishing nodules and saprophytic capacity in the soil (Mendes, Hungria and Vargas 2004; Batista *et al*. 2007). The genomes of these strains were sequenced and compared to identify the factors that could contribute to such symbiotic differences (Siqueira *et al*. 2014). From this study, one shared feature of the symbiotic island was revealed, the large proportion of transposases (14 to 16%), similar to what has been described for the more closely related genome strains of *B. diazoefficiens* USDA 110, *B. diazoefficiens* NK6 and *B. japonicum* USDA 6.

In bacteria, insertion sequences (IS) elements are the most common transposable elements (TEs). These TEs correspond to DNA sequences with a simple genetic organization and usually encode proteins necessary to their transposition carrying one open reading frame (ORF) corresponding to a transposase (Siguier, Gourbeyre and Chandler 2014). Transposases are often flanked by short sequences of inverted repeats (IRs) with a length of up to 40 base pairs (bps) that are recognized by transposase enzymatic activity to promote transposition. Depending on the transposase family, after the transposition event, direct repeats (DRs) of 2 to 14 bp may be generated. Each family is characterized by the transposase domain, the length of the IR and the flanking DR produced (Mahillon and Chandler 1998).

IS elements are important modulators of genomes and are often considered as agents of evolution by influencing the adaptation of the organism. Indeed, they can promote various types of rearrangements, including insertion, deletion, inversion and replicons junction as revealed by comparative genomics (Monteiro-Vitorello *et al*. 2005). Insertion events can lead to the integration of gene sets with specialized functions such as resistance to antibiotics, virulence and symbiosis, among others. In addition, IS elements can disrupt coding regions or be inserted into regulatory regions, affecting the expression of neighboring genes (Cerveau *et al*. 2011; Casacuberta and González 2013; Siguier, Gourbeyre and Chandler 2014). No information on the IS content is available for these strains. Thus, we aimed to define the IS element gene set and investigated their impact in differentiating the Brazilian-adapted *B. diazoefficiens* CPAC 7 and *B. japonicum* CPAC 15 strains, mainly on their symbiotic capacity.

MATERIALS AND METHODS

The CPAC 7 strain genome was retrieved from the genome database of the National Center for Biotechnology Information (NCBI) and has 9085 533 bp available in 13 contigs (NCBI ID: ADOU02). Similarly, the genome of CPAC 15 is available as a scaffold with 9583 027 bp (NCBI ID: CP007569).

IS elements identification and annotation

ISsaga (Varani, Siguier and Gourbeyre 2011) and ISfinder (Siguier *et al*. 2006) were used to identify IS elements in the two genomes. This analysis was performed in two steps: automated analysis and manual curation. The automated step was performed by submitting the whole genome in FASTA format to the ISsaga web tool using default parameters. Then the output was combined with the ISfinder database, the main repository of prokaryotic TEs, which was used as reference in the prediction. After the computational prediction, the annotation was completed by manual inspection. If one specific IS element was already deposited in the ISfinder database, the original annotation was retrieved. New ISs were annotated manually. The FASTA sequence was inspected for the presence of IRs, their size, the presence of one or more ORFs, the formation of DRs in insertion sites and the IS element status (complete or partial). The family characteristics were considered in the analysis. As additional tools, BLAST (Altschul *et al*. 1990, 1997) using the RefSeq database (Pruitt, Tatusova and Maglott 2005), BLAST2Seq (Tatusova and Madden 1999) and in-house Perl scripts were used to further perform sequence analyses.

The IS-isoforms were considered following ISsaga criteria. Otherwise, new IS elements assigned in *B. diazoefficiens* CPAC 7 were named ISBd followed by a number according to the annotation order, and in *B. japonicum* CPAC 15 were named ISBja following ISfinder recommendations.

The new ISs were deposited in the ISfinder database.

Finally, the present study focused on complete IS elements confirmed by manual annotation. An IS was considered complete if both 5'and 3' IRs were identified. Moreover, a complete IS element was considered intact when the nucleotide ORF sequence originated a protein without premature stop codon.

Genomic context of full length IS elements

Complete IS elements of each genome [complete IS: ORF(s) and IRs] were mapped and their genomic surrounding analyzed. Taking the coordinates of each IS, the flanking ORFs were retrieved and the distance (in bases) between the IS element and their ORFs was defined using in-house Perl scripts. Each ORF was manually inspected, also verifying the annotation of proteins defined as hypothetical *a priori*, using RefSeq, UniProt (Uniprot Consortium 2014) and Interpro (Hunter *et al*. 2009) databases.

Each complete IS element received a classification according to its genomic context: (i) **inter_ORF** refers to IS elements inserted into intergenic regions, distant from ORFs; (ii) **ORF_disruption** refers to IS inserted into coding regions; (iii) **Promoter_region** refers to IS elements present in promoter regions where distances of up to 100 bases from the start codon were considered; (iv) **Terminator_region** refers to IS elements inserted into terminator regions (distance of up to 25 bases from stop codon) and (v) **modified_ORF** refers to IS elements overlapping genes when in the opposite DNA strand or to IS elements and ORF sharing sequences when in the same sense of the DNA strand. Fig. S1 is a representation of a modified ORF.

Table 1. Comparison of the copy numbers of each full length IS element present in the genomes and those found in the symbiotic island.

		Bd CPAC 7		Bd USDA 110		Bd NK6		**Bj CPAC 15**		Bj E109		Bj USDA 6			
IS element family	**Full IS elements**	**Total**	**SI**	**Total**	**SI**	**Total**	**SI**	**Total**	**SI**	**Total**	**SI**	**Total**	**SI**	**IS reference copy (pb)**	**Reference**
IS110 ssgr IS1111	ISBd10	1	1	1	1	0	0	1	0	0	0	0	0	1,410	This study
IS110 ssgr IS1111	ISBd11	1	1	1	1	1	1	1	1	1	1	1	1	1,339	This study
IS110 ssgr IS1111	ISBd12	1	1	0	0	1	1	0	0	0	0	0	0	1,357	This study
IS110 ssgr IS1111	ISBd13	1	1	1	1	0	0	0	0	0	0	0	0	1,263	This study
IS110 ssgr IS1111	ISBj4	3	0	6	0	6	2	4	0	2	0	1	0	1,576	Isfinder database
IS110 ssgr IS1111	ISBja3	1	1	1	1	1	1	2	1	1	1	1	1	1,406	This study
IS110 ssgr IS1111	ISBja4	1	0	0	0	1	0	4	0	0	0	0	0	1,370	This study
IS110 ssgr IS1111	ISBja5	0	0	0	0	1	0	1	0	0	0	0	0	1,352	This study
IS110 ssgr IS1111	ISBja7	0	0	1	0	0	0	1	0	0	0	0	0	1,154	This study
IS1182	ISBd4	1	0	0	0	2	0	0	0	0	0	0	0	1,697	This study
IS1182	ISBja2	0	0	0	0	0	0	3	0	0	0	0	0	1,564	This study
IS1380	ISBdi2	12	4	1	1	39	16	1	1	0	0	0	0	1,622	Isfinder database
IS21	ISFK1	2	1	3	3	0	0	3	2	5	2	2	2	2,532	Isfinder database
IS21	ISBj11	1	1	4	2	0	0	0	0	1	1	1	1	2,606	Isfinder database
IS21	ISBd7	4	2	1	0	0	0	1	0	1	0	1	0	2,479	This study
IS21	ISBd8	1	1	1	1	0	0	0	0	0	0	0	0	2,245	This study
IS256	IS1632	2	2	2	2	22	12	4	3	2	2	2	2	1,395	Isfinder database
IS256	ISBd1	1	1	1	1	1	1	1	1	1	1	1	1	1,303	This study
IS3 ssgr IS150	ISBj3	1	1	1	1	0	0	0	0	0	0	0	0	1,284	Isfinder database
IS3 ssgr IS150	ISRj2	7	6	10	5	40	14	7	4	4	1	4	1	1,364	Isfinder database
IS3 ssgr IS3	ISBd25	1	1	1	1	0	0	1	1	1	1	1	1	643	This study
IS3 ssgr IS407	ISBja6	0	0	0	0	0	0	2	1	0	0	0	0	1,268	This study
IS481	ISBd2	1	1	1	1	0	0	3	2	1	1	1	1	1,240	This study
IS481	ISBja1	0	0	0	0	0	0	3	0	3	0	3	0	1,271	This study
IS5 ssgr IS427	ISBj2	8	7	9	8	25	12	14	8	8	6	10	8	832	Isfinder database
IS5 ssgr IS5	ISBd5	1	0	0	0	0	0	0	0	0	0	0	0	1,535	This study
IS5 ssgr IS5	ISBj2_B	2	2	2	1	2	2	2	2	2	2	2	2	1,509	Isfinder database
IS5 ssgr IS5	ISBj5_B	2	0	7	3	0	0	2	0	0	0	0	0	1,524	Isfinder database
IS6	ISBj7	1	1	3	3	7	3	4	4	4	4	4	4	816	Isfinder database
IS630	ISBj5	2	0	4	0	6	3	2	0	3	0	4	0	1,185	Isfinder database
IS630	ISBd14	2	0	0	0	2	0	0	0	0	0	0	0	1,044	This study
IS630	ISRj1	9	7	15	8	9	0	10	8	9	7	9	7	1,118	Isfinder database
IS66	ISBja8	1	1	1	1	0	0	1	1	1	1	1	1	1,584	This study
IS701	ISBj6_B	1	1	1	1	0	0	0	0	1	1	1	1	1,448	Isfinder database
IS91	ISBd6	1	0	1	0	1	0	0	0	0	0	0	0	1,283	This study
IS91	ISBd9	1	0	1	0	1	0	0	0	0	0	0	0	1,103	This study
ISL3	ISBd3	1	1	0	0	0	0	2	0	0	0	0	0	1,654	This study
ISNCY ssgr ISLbi1	ISBj12	2	1	2	1	0	0	1	0	4	0	2	0	1,743	Isfinder database

SI: Symbiotic Island; Bd: *Bradyrhizobium diazoefficiens*; Bj: *Bradyrhizobium japonicum*

To detect insertion events in CPAC 7 and CPAC 15 genomes that could be affecting genes in the symbiotic island, comparisons were performed with 21 other genomes of *Bradyrhizobium* (Table S1). This analysis was performed with BLASTn adopting the default parameters and the queries were the nucleotide sequence encompassing the gene and the complete IS element.

Expression analysis of impacted genes

We also investigated whether the impacted genes could be identified as differentially expressed under symbiotic conditions in *B. diazoefficiens* USDA 110 [Transcriptome – 21 days after inoculation (DAI)] (Pessi *et al.* 2007), Proteome – 21 DAI (Delmotte *et al.* 2010) and Transcriptome – 28 DAI (Franck *et al.* 2014). To identify the homologues (≥90% identity), we mapped our targets using BLASTn against genes of USDA 110 presented in these three studies.

RESULTS

Defining full length IS element population in the genomes of Brazilian-adapted *Bradyrhizobium* strains

We identified 77 complete IS elements in the genome of *B. diazoefficiens* CPAC 7 belonging to 15 IS families and 81 IS in the genome of *B. japonicum* CPAC 15 belonging to 13 families (Tables S2 and S3). Among the 38 different complete IS elements (not considering isoforms), 15 were already available in the ISfinder database and 23 new IS elements were annotated in this study (Table 1).

Table 1 summarizes the structural features of the complete IS elements and their respective copy numbers in CPAC 7 and CPAC 15 compared with their most closely related strains. *B. diazoefficiens* USDA 110 and *B. diazoefficiens* NK6 are closely related to *B. diazoefficiens* CPAC 7, whereas *B. japonicum* E109 and *B. japonicum* USDA 6 are closely related to *B. japonicum* CPAC 15. Among these bacteria, *B. diazoefficiens* NK6 has the highest number of IS

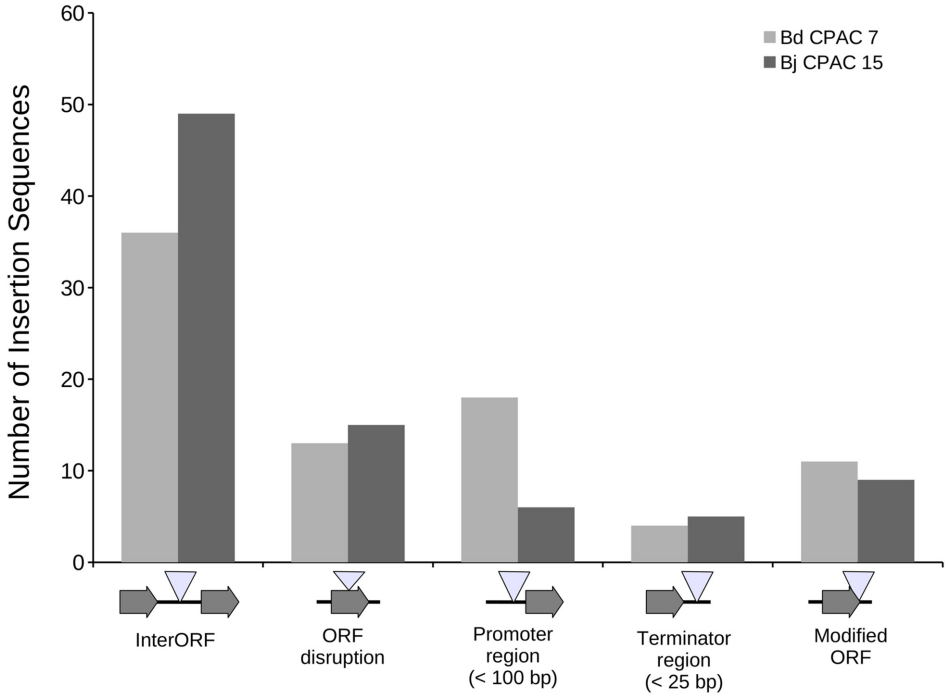

Figure 1. Genomic context of complete IS elements identified in *Bradyrhizobium diazoefficiens* CPAC 7 and *Bradyrhizobium japonicum* CPAC15. InterORF: intergenic region; ORF disruption: IS inserted into coding regions; Promoter region: IS inserted into promoter region (distance < 100 bp from start codon); Terminator region: IS inserted into terminator region (distance < 25 bp from stop codon); Modified ORF: IS overlapping gene when in opposite strand or IS and ORF sharing sequences when in the same sense of the DNA strand.

elements (168) annotated in this study, while *B. japonicum* USDA 6 has the lowest number (52). We also noted that CPAC 7 and CPAC 15 strains share 22 IS. Emphasis should be given to ISBdi2, which has 12 copies in *B. diazoefficiens* CPAC 7 and only one copy in *B. japonicum* CPAC 15. Conversely, *B. diazoefficiens* CPAC 7 presents eight copies of ISBj2, whereas *B. japonicum* CPAC 15 contains 14 copies of this IS in its genome.

Comparing these closely related strains, ISBja2 is unique to CPAC 15 with three copies, whereas ISBd5 is present only in CPAC 7 (one copy). ISBd11 is present as a single copy and same insertion in all of the compared strains. On the other hand, ISBj7 has four copies in the symbiotic island of *B. japonicum* strains, whereas it varies from one to seven insertions in *B. diazoefficiens* strains. All three insertions of ISBja1 are shared by the three *B. japonicum* strains, while ISBd6 and ISBd9 are present only in *B. diazoefficiens* strains (Table 1).

Local impact of complete IS elements

After the annotation process, all coordinates for complete IS elements were retrieved from the genomes and their location in relation to a nearby ORF were investigated. Fig. 1 shows that the event found most frequently is **inter_ORF** insertions. In addition, CPAC 7 and CPAC 15 share similar **ORF_disruption**, **Terminator_region** and **Modified_ORF** insertions. However, *B. diazoefficiens* CPAC 7 has more IS in the promoter region than *B. japonicum* CPAC 15. Conversely, CPAC 15 is more frequently associated with interORF positioning. The full list of the impacted genes is available in Table 2. Considering all complete IS elements, about ~61% of *B. diazoefficiens* CPAC 7 and 50% of *B. japonicum* CPAC 15 IS elements are clustered at the symbiotic island.

IS-impacted genes

As a result, from this detailed analysis, a collection of impacted genes in CPAC 7 and CPAC 15 are revealed (Table 2), of which only genes related to the symbiosis are described below. *B. diazoefficiens* CPAC 7 has 13 ORFs disrupted by complete IS elements, while *B. japonicum* CPAC 15 has 15 disrupted ORFs. Among these, two genes previously described as relevant to the symbiosis are: (i) a methylthiotransferase gene that is disrupted in both CPAC 7 (BJA5080_0 7602) and CPAC 15 (BJS_0 8822), and (ii) an ABC transporter permease that is only disrupted in CPAC 15 (BJS_0 8733). Methylthiotransferase is a single copy gene of 1980 bp. Distinct IS elements (ISBj2 and ISRj1) caused the disruption event. A representation of this insertion is shown in Fig. 2a. Methylthiotransferase in the CPAC 7 strain had its coordinate corrected (BJA5080_0 7602, Cont24: Part I 1 595 750:1 595 779 and Part II 1 596 900:1 598 849). The gene was disrupted by an intact ISRj1 element (Cont24–1 595 783:1 596 900); however, according to the Interpro search, the functional domain (Part II) of the methylthiotransferase gene remains intact (Fig. 2a). In contrast, its functional domain in CPAC 15 is disrupted (Part I BJS_0 8822–224 174:224 900 and Part II BJS_0 8163–218 724:219 979) by a more complex insertion event. It is composed of a partial non-identified IS (BJS_0 8711), a group II intron (BJS_0 8710–219 992:221 903) and a defective ISBj2 (223 343:224 173), as shown in Fig. 2a. Also, in the symbiotic island of CPAC 15, we found a component of an ABC transporter operon (BJS_0 8733–Part I 9 436 782:9 436 851 and Part II 9 438 256:9 438 848) disrupted by a complete IS1632 (9 436 852:9 438 246).

As presented in Table 2, *B. diazoefficiens* CPAC 7 has more complete IS element upstream genes. Among the 23 genes with a complete IS element inserted into the promoter region, 10 are located in the symbiotic island. These are hypothetical

Table 2. Impacted genes by complete IS elements in Brazilian strains, some of these being homologous genes in *Bradyrhizobium diazoefficiens* USDA110 differentially expressed in symbiosis conditions.

	IS element	Gene	CPAC 7	CPAC 15	Distance between IS and ORF (bp)	Symbiotic island?	Expressed in symbiotic condition in USDA 110
			Locus Tag	Locus Tag			
Disruption	ISBdj12	Transposase	BJA5080_07124	BJS_05811	-	No	-
	IS1632	ABC-transporter – permease	-	BJS_08733	-	Yes	blr1679[d]
	ISBd2	Transposase	BJA5080_08164	BJS_08315	-	Yes	bsl1856[u]
	ISBd2	RNA-directed DNA polymerase	-	BJS_08833	-	Yes	-
	ISBdi2	Hypothetical	BJA5080_06299	-	-	No	-
	ISBd3	Phosphoglycolate phosphatase	-	BJS_07033	-	No	-
	ISFK1	Transposase	BJA5080_08268	BJS_08831	-	No	-
	ISBj11	Hypothetical	BJA5080_08345	-	-	Yes	-
	ISBd7	trbI-conjugal transfer	BJA5080_08064	-	-	Yes	-
	ISBd7	DEAD-like helicases superfamily	BJA5080_06102	-	-	No	-
	ISBj7	Transposase	-	BJS_08948	-	Yes	-
	ISBd11	Transposase	BJA5080_08069	BJS_08415	-	Yes	-
	ISBj2	trbI-conjugal transfer	BJA5080_08366	BJS_08864	-	Yes	-
	ISBj2	Transposase	-	BJS_08412	-	Yes	-
	ISBj2 ISRj1	Methylthiotransferase/radical SAM-type	BJA5080_07602	BJS_08822	- -	Yes	bll1977[u]
	ISRj1	traG-conjugal transfer	BJA5080_08181	BJS_08775	-	Yes	-
	ISRj1	Oxidoreductase	BJA5080_00853	-	-	No	-
	ISRj1	Hypothetical	-	BJS_07667	-	Yes	bll1979[u]
	ISBja3	Transposase	BJA5080_07660	BJS_08809	-	Yes	-
	ISBja6	Transposase	-	BJS_08844	-	Yes	-
Promoter region	ISBdi2	Hypothetical	BJA5080_08087	-	95	Yes	-
	ISBdi2	Hypothetical	BJA5080_00952	-	61	No	-
	ISBdi2	Hypothetical	BJA5080_04252	-	16	No	-
	ISBdi2	Type IV pilus assembly pilZ	BJA5080_08401	-	67	No	blr5568[d]
	ISBdi2	Hypothetical	BJA5080_08259	-	91	No	-
	ISBd6	priA-Primosomal N	BJA5080_01519	-	18	No	-
	ISBd7	ATPase	BJA5080_06101	-	71	No	-
	ISBd12	Major facilitator transporter	BJA5080_07988	-	81	Yes	-
	ISBj2	Hypothetical	BJA5080_07587	-	97	Yes	-
	ISBj2	Hypothetical	BJA5080_07732	-	44	Yes	-
	ISBj2	Lytic murein transglucosyllase	-	BJS_07214	48	No	bll2447[d]
	ISBj2	Multi-sensor signal transduction	-	BJS_07014	19	No	-
	ISBj2	Glucosylltransferase	-	BJS_02329	35	No	-
	ISBj4	Hypothetical	BJA5080_00921	-	80	No	-
	ISBja2	Putative signal peptide	-	BJS_06214	81	No	-
	ISBja3	Signal peptide	-	BJS_00582	9	No	-
	ISBja8	Hypothetical	BJA5080_08031	-	59	Yes	-
	ISBja8	Hypothetical	BJA5080_08028	-	50	Yes	-
	ISRj1	Type III effector nopAN	N/A	BJS_08797	7	Yes	blr1912[u]
	ISRj1	Amidohydrolase	BJA5080_06025	-	40	No	-
	ISRj2	PilA2 – pilus assembly	N/A	-	15	Yes	-
	ISRj2	Exonuclease sbcD	BJA5080_08362	-	24	Yes	-
	ISRj2	Peptidase C58	BJA5080_07710	-	24	Yes	blr2118[u]

Table 2. Continued

Terminator region	ISBdi2	Type III effector nopAG	BJA5080_08310	BJS_09091	1	Yes	bll1862[u]
	ISBdi2	Sel1 domain	BJA5080_08421	-	1	No	-
	ISBdi2	Transglutaminase	BJA5080_08400	-	1	No	blr5569[u]
	ISBd7	Hypothetical	-	BJS_08586	2	No	-
	ISBj2	rhcU - type III secretion	BJA5080_08074	BJS_08411	1	Yes	-
	ISBj2	Adenylate/guanilate cyclase	-	BJS_01536	1	No	-
	ISBj4	HTH- type- lysR	-	BJS_04020	7	No	-
	IS element	**Gene**	**Locus Tag**	**Locus Tag**	**Shared or overlapping sequence (bp)**	**Symbiotic island?**	**Expressed in symbiotic condition in USDA 110**
Modified ORF	IS1632	modC-Molybdenum ABC transport	BJA5080_08107	BJS_08743	48	Yes	-
	ISBd1	Hypothetical	BJA5080_07730	BJS_07527	24	Yes	-
	ISBdi2	Hypothetical	BJA5080_08396	-	40	No	-
	ISBdi2	Hypothetical	BJA5080_08161	BJS_08313	18-15	Yes	-
	ISFK1	Hypothetical	BJA5080_08296	BJS_08483	3	Yes	-
	ISBd6	Hypothetical	BJA5080_01521	-	121	No	blr0447[d]
	ISBj7	Hypothetical	BJA5080_07526	BJS_08509	4	Yes	-
	ISBd10	Hypothetical	BJA5080_02655	-	28	Yes	blr1546[u]
	ISBj5	Putative transcriptional regulator	BJA5080_06196	BJS_07670	23	No	-
	ISBj2	Multi-sensor signal transducer	-	BJS_08694	16	No	-
	ISRj1	Lysine 2,3-aminomutase	-	BJS_08612	69	No	bll8232[u]
	ISRj2	hypB-hydrogenase nickel incorporation	BJA5080_08448	BJS_08967	50	Yes	blr1732[u]
	ISRj2	Exonuclease – subunit sbcD	BJA5080_08346	-	42	Yes	bll1924[d]

N/A: Not annotated in the genome. / u: up-regulated / d: down-regulated

genes, ATPase, a lytic murein transglucosyllase, glucosylltransferase, signal peptides, genes related to pilus assembly, a type III effector *nopAN* and peptidase C58. The ISRj1 is a single insertion event shared by both Brazilian strains that was found seven nucleotides upstream of the type III effector *nopAN* gene. Its representation is displayed in Fig. 2b. ISRj1 is intact in the *B. japonicum* CPAC 15 strain but not in *B. diazoefficiens* CPAC 7 (Cont20–113 141:113 977). The ISRj2 was found inserted 24 bp from the start codon of peptidase C58 in *B. diazoefficiens* CPAC 7 (Table 2).

We found IS in the terminator region of ORFs in seven cases, but only two are in the symbiotic island (Table 2). These two genes correspond to the type III secretion genes *nopAG* and *rhcU* and have an IS element in the stop codon, both events shared by CPAC 7, CPAC 15 and in all of their close relatives evaluated herein. *nopAG* in CPAC 7 and CPAC 15 were previously annotated as hypothetical. Another case of type III secretion terminator_ORF, *rhcU*, has ISBj2 (IS5 ssgr IS427) inserted in its stop codon (Table 2). Among the specific insertions in the terminator region outside the symbiotic island, we highlight the ISBdi2 insertion in the transglutaminase gene as it is present only in CPAC 7 and was found expressed in the study of *B. diazoefficiens* USDA 110 transcriptome at 21 DAI.

Table 2 shows the number of bases being shared or overlapped in Modified_ORF. Among 13 genes, eight are found in the symbiotic island and encompass hypothetical genes, a putative transcriptional regulator, a signal transducer, a lysine 2,3-aminomutase, an exonuclease, *modC* and *hypB*. Interestingly, *modC* shares 48 nucleotides with IS1632 (family IS256). The shared IS element sequence predicted is present in the encoded *modC* protein when *in silico* translated (Fig. S2). The transposase of the IS1632 in CPAC 7 shows an *indel* and causes a change in the reading frame that leads to a premature stop codon. On the other hand, the transposase from CPAC 15 is intact and therefore potentially active.

hypB is located in the symbiotic island and shares 50 bp with ISRj2. Besides occurring in CPAC 7 and CPAC 15, this insertion event was found in the strains USDA 110, USDA 6 and E109 (Fig. 3), and that in USDA 110 was identified as upregulated in symbiotic condition (Table 2).

Comparative study with other *Bradyrhizobium* strains and expression analyses

In order to understand whether some of these insertion events are unique to the Brazilian *Bradyrhizobium* strain genomes, we compared the most relevant insertions with 21 other *Bradyrhizobium* genomes (Fig. 3). Among the 34 insertion events highlighted, 22 are present in more than one genome of *Bradyrhizobium* analyzed and 28 occur in the symbiotic island (Fig. 3). Fourteen of these genes are differentially expressed (*) in symbiotic conditions (21 and 28 DAI) in the transcriptome and proteome studies of *B. diazoefficiens* USDA 110. In addition, we verified which insertion events are located in the symbiotic island.

Twenty-one insertion events identified in *B. diazoefficiens* CPAC 7 and *B. japonicum* CPAC 15 are shared by other *Bradyrhizobium* genomes, such as the strains USDA 110 and NK6 of *B. diazoefficiens*, and the strains E109, USDA 122 and USDA 6 of

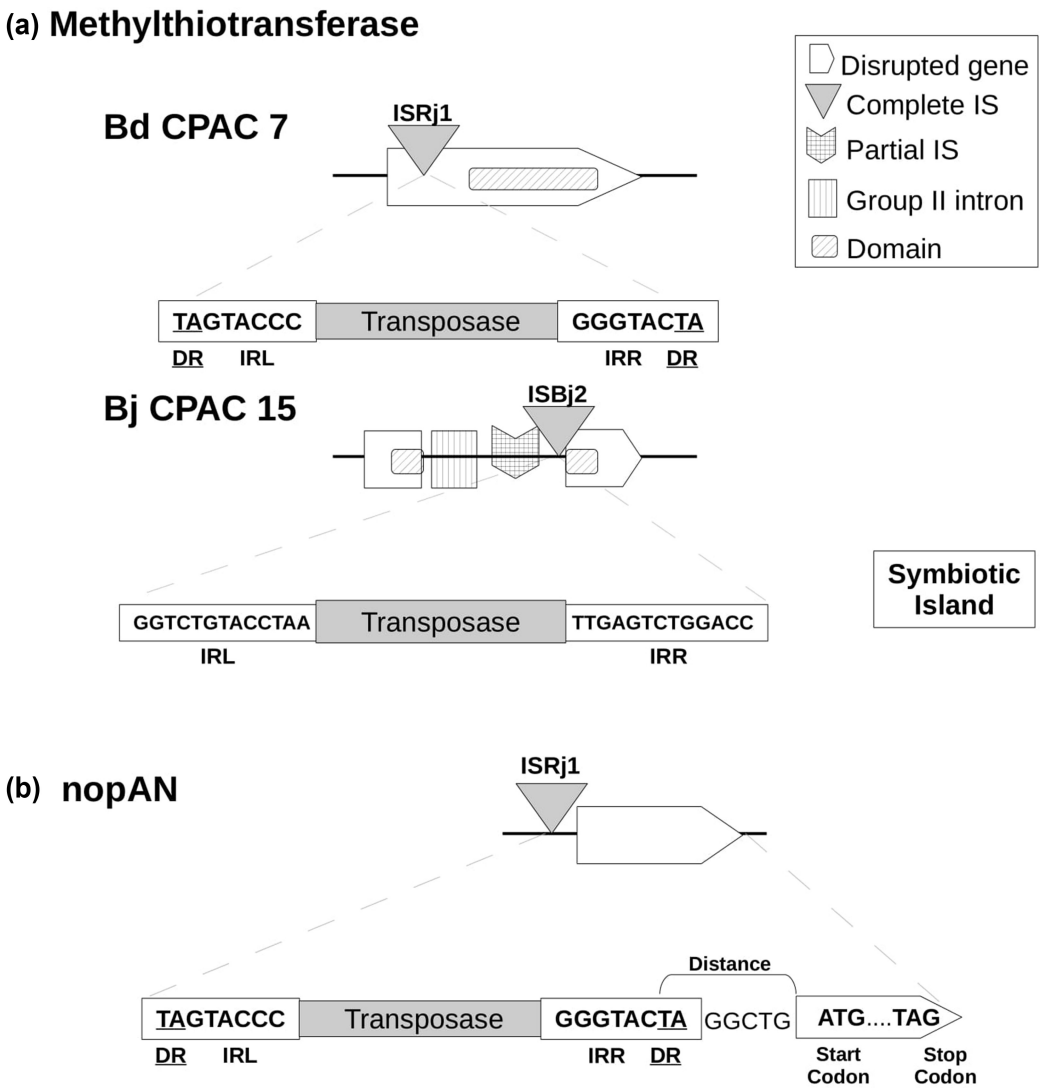

Figure 2. Representation of genes important for symbiosis located on a symbiotic island and impacted by IS elements. (**A**) Gene methylthiotransferase disrupted by different insertion events in *Bradyrhizobium diazoefficiens* CPAC 7 and *Bradyrhizobium japonicum* CPAC 15. (**B**) Gene *nopAN* (nodulation outer protein) containing ISRj1 (Family IS630) inside its promoter region. DR, direct repeat; IRL, inverted repeat left; IRR, inverted repeat right.

B. japonicum, but are absent in most of the other genomes (Fig. 3). With the exception of the methylthiotransferase disruption event produced by different IS elements, when the insertion event is detected it is shared by several strains.

DISCUSSION

Our analyses describe 23 new IS elements found in *B. diazoefficiens* CPAC 7 and *B. japonicum* CPAC 15. Also, we confirmed that these elements concentrate in the symbiotic island of both strains, as suggested by Siqueira *et al.* (2014). The annotation of these new IS and their submission to the ISfinder database should expand the current knowledge about these elements in *Bradyrhizobium* and related bacteria.

The predominance of IS in the symbiotic islands has been reported in other rhizobial genomes such as *Mesorhizobium loti* MAFF303099 (Kaneko *et al.* 2000), *B. diazoefficiens* USDA 110 (Kaneko *et al.* 2002), *B. japonicum* USDA 6 (Kaneko *et al.* 2011) and *B. diazoefficiens* NK6 (Iida *et al.* 2015). Current knowledge assumes that the symbiotic island is derived from a symbiotic plasmid (*pSym*) that has been integrated to the chromosome via horizontal transfer (Sullivan and Ronson 1998; Göttfert *et al.* 2001; Finan 2002; Okubo *et al.* 2016). This would be a possible explanation for the high concentration of IS in the symbiotic island because, according to Siguier, Filée and Chandler (2006a), the presence of IS elements in plasmids assists in the inclusion of new genes, increasing the repertoire of non-essential genes for its maintenance. Thus, *pSym* may have been brought with the IS at the moment of the integration, because as presented in the *pSym* of *Rhizobium* sp. NGR234, such elements are present (Freiberg *et al.* 1997).

Uchiumi *et al.* (2004) analyzed the expression of genes in the symbiotic island of bacteroids of *M. loti* MAFF303099 and found that most transposases and symbiotic genes are highly upregulated during symbiosis (e.g. *nif* and *fix*), while most genes outside the island are downregulated. Similar results were obtained with USDA 110 (Wei *et al.* 2008). These studies strengthened the hypothesis that TEs may have important roles in the BNF process.

Disruption of a methylthiotransferase, involved, among other functions, in providing the radical S-adenosylmethionine

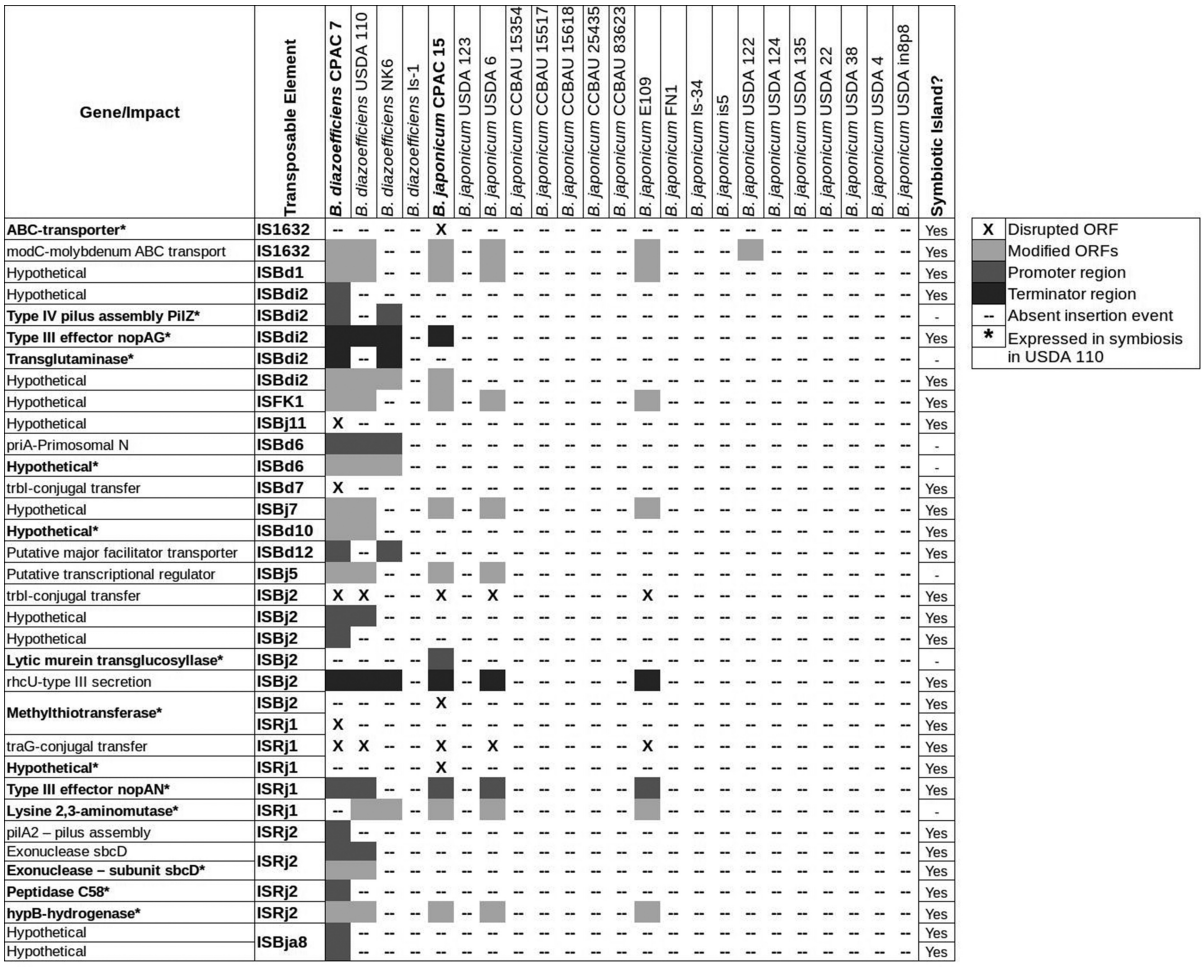

Figure 3. Main insertion events that might be affecting genes of *Bradyrhizobium*.

to the nitrogenase formation (Wiig *et al.* 2012; Hu and Ribbe 2013; Ribbe *et al.* 2014), is common to CPAC 7 and CPAC 15 but is the result of two independent events produced by two distinct IS elements. By contrast, this gene remains intact in *B. japonicum* USDA 6 (BJ6T_78 720) and *B. diazoefficiens* USDA 110 (bll1977) (Fig. 3). Indeed, *B. diazoefficiens* USDA 110, the intact homologue (bll1977) was differentially expressed at 21 DAI (Delmotte *et al.* 2010). We speculate that the intact functional domain of the methylthiotransferase of *B. diazoefficiens* CPAC 7 supports the higher BNF efficiency of this bacterium in comparison with *B. japonicum* CPAC 15, in which the protein functional domain is disrupted by a complex insertion (Fig. 2). However, further research will be needed to test this hypothesis.

Another gene in the symbiotic island that may explain differences in BNF capacities between CPAC 15 and CPAC 7 is the ABC transporter permease because it is disrupted only in CPAC 15 and its homologue (blr1679) is expressed in symbiosis at 21 DAI. Indeed, we observed that this gene is intact in the other *Bradyrhizobium* genomes (Fig. 3).

Nodulation outer proteins, encoded by *nop* genes, are classified as type III secretion systems essential to the infection process during the establishment of symbiosis, and are also required for the nodulation of some tropical legumes (Dale *et al.* 2002; Alfano and Collmer 2004; Kimbrel *et al.* 2013; Tsukui *et al.* 2013). Both *nopAN* and *nopAG* are affected by the insertions of ISRj1 and ISBdi2 in both CPAC strains. We found that the ISRj1 insertion upstream to the *nopAN* is conserved in five of the nitrogen-fixing *Bradyrhizobium* strains investigated (CPAC 7, USDA 110, CPAC 15, USDA 6 and E109) (Fig. 3). Besides, in USDA 110 (blr1912), *nopAN* was expressed at 21 and 28 DAI. However, it remains to be demonstrated whether this insertion event is indeed beneficial, beyond the fact of being conserved in the *Bradyrhizobium* strains studied herein (Fig. 3). As already mentioned, the *nopAG* gene has the ISBdi2 in its stop codon, but this element appears to be inactive once it has a deletion of 470 nucleotides when compared with ISBdi2 reference. It is interesting that both ISBdi2 and *nopAG* had their homologues (bll1861 and bll1862, respectively) expressed in symbiotic conditions in *B. diazoefficiens* USDA 110 (Pessi *et al.* 2007).

Peptidase C58 is widely reported in pathogenic and root-nodulating bacteria. This protease family plays a key role after the invasion of the host cell and may be involved in resistance mechanisms (Hotson and Mudgett 2004). We identified an insertion event of ISRj2 at 24 bp from its start codon unique to CPAC 7 (Fig. 3). ISRj2 may interfere with the performance of *B. diazoefficiens* CPAC 7 in symbiosis since its *B. diazoefficiens* USDA 110 homologue, with no upstream IS element, is upregulated in N_2-fixing bacteroids (Delmotte *et al.* 2010).

The presence of IS downstream to genes is also relevant because they can interfere with the transcription termination process (Peters, Vangeloff and Landick 2011). *B. diazoefficiens* CPAC 7 and *B. japonicum* CPAC 15 share only two insertion events

in terminator regions, both located in symbiotic islands and involving a gene of the type III secretion system (*nopAG* and *rhcU*) (Table 2). These insertions are conserved in some other *Bradyrhizobium* genomes, as observed in Fig. 3.

The *modC* modified_ORF event occurs in six *Bradyrhizobium* genomes; *modC* encodes a protein that participates in the ABC transport complex, being responsible, together with genes *modA* and *modB,* for the transport of molybdenum into the bacterial cell. Given that molybdenum is an essential co-factor for the nitrogenase activity, *modC* activity is expected to be essential (Delgado *et al.* 2006). However, gene redundancy is observed for CPAC 15, CPAC 7 and USDA110. Therefore, experimental studies should be conducted to investigate the impact of this insertion in the *modC* gene.

Similarly, *hypB* shares 50 nucleotides with ISRj2, a conserved event in five *Bradyrhizobium* strains: CPAC 7, USDA 110, CPAC 15, E109 and USDA 6 (Fig. 3). This gene is part of the operon *hypABFCDE* (*hyp:* hydrogenase pleiotrophy) present in symbiotic islands and responsible for nickel incorporation into hydrogenases (Hansel *et al.* 2001). ISRj2 is inserted between *hypB* and *hypF* genes. Siqueira *et al.* (2014) showed that hydrogen-uptake system of Brazilian-adapted *Bradyrhizobium* strains located in symbiotic islands appears to be non-functional. This is probably due to the presence of a transposase in the operon in *B. diazoefficiens* USDA110, the *hypB* (blr1732) gene, also present in CPAC 7 and CPAC 15, which was expressed in symbiotic condition 21 DAI, evidence that the operon *hypABFCDE* is active (Pessi *et al.*2007). Previous studies showed that *B. diazoefficiens* CPAC 7 has the Hup^+ phenotype whereas the *B. japonicum* CPAC 15 has the Hup^- phenotype (Boddey and Hungria 1997). The *hup* genes are responsible for expression of hydrogenase activity and that in turn increases BNF efficiency in soybean (Albrecht *et al.* 1979). Thus, the phenotypic differences (Hup^+ and Hup^-) may be because while *B. diazoefficiens* strains have different hydrogenase genes in other chromosomal loci (Göttfert *et al.* 2001; Loh and Stacey 2003), this does not occur in *B. japonicum* strains (Siqueira *et al.* 2014).

In conclusion, manual curation and annotation of full length IS elements revealed a distinctive contribution of these elements to the Brazilian-adapted *Bradyrhizobium* species. Also revealed are ancestral insertions in the symbiotic island of multiple strains of *B. diazoefficiens* and *B. japonicum*. Unique insertions in CPAC 7 and CPAC 15 suggest recent transposition activity. Finally, most of the sequenced *B. japonicum* strains do not display insertions in the symbiotic island. Taken together, the results emphasize the role of IS elements in genome plasticity.

SUPPLEMENTARY DATA

Supplementary data are available at *FEMSLE* online.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Almiro Pisetta, Dra. Lenita R. Esteves and Dra. Tatiana C. S. Corrêa for suggestions to the manuscript and to Dra Andrea Kauffmann-Zeh for English revision and suggestions to the manuscript. GABC was recipient of a Coordenação de Aperfeiçoamento do Pessoal de Nível Superior (CAPES) PhD scholarship—Finance Code 001*. MAVS supported the research with grants from Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP 2008/52074–0) and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Conflicts of interest. None declared.

REFERENCES

Albrecht SL, Maier RJ, Hanus FJ *et al.* Hydrogenase in *Rhizobium japonicum* increases nitrogen fixation by nodulated soybeans. *Science* 1979;**203**:1255–7.

Alfano JR, Collmer A. Type III secretion system effector proteins: double agents in bacterial disease and plant defense. *Annu Rev Phytopathol* 2004;**42**:385–414.

Altschul SF, Gish W, Miller W *et al.* Basic local alignment search tool. *J Mol Biol* 1990;**215**:403–10.

Altschul SF, Madden TL, Schäffer AA *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;**25**:3389–402.

Batista JS, Hungria M, Barcellos F *et al.* Variability in *Bradyrhizobium japonicum* and *B. elkanii* seven years after introduction of both the exotic microsymbiont and the soybean host in a cerrados soil. *Microb Ecol* 2007;**53**:270–84.

Boddey LH, Hungria M. Phenotypic grouping of Brazilian *Bradyrhizobium* strains which nodulate soybean. *Biol Fertil Soils* 1997;**25**:407–15.

Casacuberta E, González J. The impact of transposable elements in environmental adaptation. *Mol Ecol* 2013;**22**:1503–17.

Cerveau N, Leclercq S, Bouchon D *et al.* Evolutionary dynamics and genomic impact of prokaryote transposable elements. In: Pontarotti P (ed). *Evolutionary Biology – Concepts, Biodiversity, Macroevolution and Genome Evolution*. Berlin: Springer, 2011, 291–312.

Dale C, Plague GR, Wang B *et al.* Type III secretion systems and the evolution of mutualistic endosymbiosis. *Proc Natl Acad Sci USA* 2002;**99**:12397–402.

Delamuta JR, Ribeiro RA, Ormeño-Orrillo E *et al.* Polyphasic evidence supporting the reclassification of *Bradyrhizobium japonicum* group Ia strains as *Bradyrhizobium diazoefficiens* sp. nov. *Int J Syst Evol Microbiol.* 2013;**63**:3342–51.

Delgado MJ, Tresierra-Ayala A, Talbi C *et al.* Functional characterization of the *Bradyrhizobium japonicum modA* and *modB* genes involved in molybdenum transport. *Microbiology* 2006;**152**:199–207.

Delmotte N, Ahrens CH, Knief C *et al.* An integrated proteomics and transcriptomics reference data set provides new insights into the *Bradyrhizobium japonicum* bacteroid metabolism in soybean root nodules. *Proteomics* 2010;**10**:1391–400.

Finan T. Evolving insights: symbiosis islands and horizontal gene transfer. *J Bacteriol* 2002;**184**:2855–6.

Franck S, Franck WL, Birke SR *et al.* Comparative transcriptomic analysis of symbiotic *Bradyrhizobium japonicum*. *Symbiosis* 2014;**63**:123–35.

Freiberg C, Fellay R, Bairoch A *et al.* Molecular basis of symbiosis between *Rhizobium* and legumes. *Nature* 1997;**387**:394–401.

Göttfert M, Röthlisberger S, Kündig C *et al.* Potential symbiosis specific genes uncovered by sequencing a 410-kilobase DNA region of the Bradyrhizobium japonicum chromosome. *J Bacteriol* 2001;**183**:1405–12.

Hansel A, Axelsson R, Lindberg P *et al.* Cloning and characterisation of a *hyp* gene cluster in the filamentous *Cyanobacterium nostoc* sp. Strain PCC 73102. *FEMS Microbiol Lett* 2001;**201**:59–64.

Herridge D, Peoples M, Boddey R. Global inputs of biological nitrogen fixation in agricultural systems. *Plant Soil* 2008;**311**:1–18.

Hotson A, Mudgett MB. Cysteine proteases in phytopathogenic bacteria: identification of plant targets and activation of innate immunity. *Curr Opin Plant Biol* 2004; **7**:384–90.

Hu Y, Ribbe MW. Nitrogenase assembly. *Biochim Biophys Acta* 2013;**1827**:1112–22.

Hungria M, Franchini J, Campo R *et al*. Nitrogen nutrition of soybean in Brazil: Contributions of biological N_2 fixation and n fertilizer to grain yield. *Can J Plant Sci* 2006;**86**:927–39.

Hungria M, Mendes IC. Nitrogen fixation with soybean: the perfect symbiosis?In: Bruijn FJ (ed). *Biological Nitrogen Fixation*. New Jersey: John Wiley & Sons, Inc, 2015, 1005–19.

Hunter S, Apweiler R, Attwood TK *et al*. InterPro: the integrative protein signature database. *Nucleic Acids Res* 2009; **37**:D211–5.

Iida T, Itakura M, Anda M *et al*. Symbiosis island shuffling with abundant insertion sequences in the genomes of extra-slow-growing strains of soybean bradyrhizobia. *Appl Environ Microbiol* 2015;**81**:4143–54.

Kaneko T, Maita T, Hirakawa H *et al*. Complete genome sequence of the soybean symbiont *Bradyrhizobium japonicum* strain USDA6T. *Genes* 2011;**2**:763–87.

Kaneko T, Nakamura Y, Sato S *et al*. Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA Res* 2002;**9**:189–97.

Kaneko T, Nakamura Y, Sato S *et al*. Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. *DNA Res* 2000;**7**:331–8.

Kimbrel JA, Thomas WJ, Jiang Y *et al*. Mutualistic co-evolution of type III effector genes in *Sinorhizobium fredii* and *Bradyrhizobium japonicum*. *PLOS Pathog* 2013;**9**:e1003204.

Loh J, Stacey G. Nodulation gene regulation in *Bradyrhizobium japonicum*: a unique integration of global regularory circuits. *Appl Environ Microbiol* 2003;**69**:10–17.

Mahillon J, Chandler M. Insertion sequences. *Microbiol Mol Biol Rev* 1998;**62**:725–74.

Mendes IC, Hungria M, Vargas MAT. Establishment of *Bradyrhizobium japonicum* and *B. elkanii* strains in a Brazilian cerrado oxisol. *Biol Fertil Soils* 2004;**40**:28–35.

Monteiro-Vitorello CB, de Oliveira MC, Zerillo MM *et al*. *Xylella* and *Xanthomonas* Mobil'omics. *OMICS* 2005;**9**:146–59.

Okubo T, Piromyou P, Tittabutr P *et al*. Origin and evolution of nitrogen fixation genes on symbiosis islands and plasmid in *Bradyrhizobium*. *Microbes Environ* 2016;**31**:260–7.

Oldroyd G, Downie J. Coordinating nodule morphogenesis with rhizobial infection in legumes. *Annu Rev Plant Biol* 2008;**59**:519–46.

Oldroyd G, Murray J, Poole P *et al*. The rules of engagement in the legume-rhizobial symbiosis. *Annu Rev Genet* 2011;**45**:119–44.

Pessi G, Ahrens C, Rehrauer H *et al*. Genome-wide transcript analysis of *Bradyrhizobium japonicum* bacteroids in soybean root nodules. *Mol Plant Microbe Interact* 2007; **20**:1353–63.

Peters J, Vangeloff A, Landick R. Bacterial transcription terminators: the RNA 3′-end chronicles. *J Mol Biol* 2011;**412**:793–813.

Pruitt K, Tatusova T, Maglott D. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 2005;**33**:D501–4.

Ribbe MW, Hu Y, Hodgson KO *et al*. Biosynthesis of nitrogenase metalloclusters. *Chem Rev* 2014;**114**:4063–80.

Siguier P, Gourbeyre E, Chandler M. Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiol Rev* 2014;**38**:865–91.

Siguier P, Filée J, Chandler M. Insertion sequences in prokaryotic genomes. *Curr Opin Microbiol* 2006a;**9**:526–31.

Siguier P, Perochon J, Lestrade L *et al*. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res* 2006;**34**:D32–6.

Siqueira AF, Ormeño-Orrillo E, Souza RC *et al*. Comparative genomics of *Bradyrhizobium japonicum* CPAC 15 and *Bradyrhizobium diazoefficiens* CPAC 7: elite model strains for understanding symbiotic performance with soybean. *BMC Genomics* 2014;**15**:420.

Sullivan J, Ronson C. Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. *Proc Natl Acad Sci USA* 1998;**95**:5145–9.

Tatusova T, Madden T. BLAST 2 sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol Lett* 1999;**174**:247–50.

Tsukui T, Eda S, Kaneko T *et al*. The type III secretion system of *Bradyrhizobium japonicum* USDA122 mediates symbiotic incompatibility with Rj2 soybean plants. *Appl Environ Microbiol* 2013;**79**:1048–51.

Uchiumi T, Ohwada T, Itakura M *et al*. Expression islands clustered on the symbiosis island of the *Mesorhizobium loti* genome. *J Bacteriol* 2004;**186**:2439–48.

Uniprot Consortium. Activities at the universal protein resource (UniProt). *Nucleic Acids Res* 2014;**42**:D191–8.

Van Sluys MA, Monteiro-Vitorello CB, Camargo LE *et al*. Comparative genomic analysis of plant-associated bacteria. *Annu Rev Phytopathol* 2002;**40**:169–89.

Varani A, Siguier P, Gourbeyre E. ISsaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. *Genome Biol* 2011;**12**:R30.

Wei M, Yokoyama T, Minamisawa K *et al*. Soybean seed extracts preferentially express genomic loci of *Bradyrhizobium japonicum* in the initial interaction with soybean, *Glycine max* (L.) Merr. *DNA Res* 2008;**15**:201–14.

Wiig JA, Hu Y, Lee CC *et al*. Radical SAM-dependent carbon insertion into the nitrogenase M-cluster. *Science* 2012;**337**:1672–5.