

Identification and functional analysis of recent IS transposition events in rhizobia

Ezequiel G. Mogro¹, Walter O. Draghi¹, Antonio Lagares¹ and Mauricio J. Lozano^{1*}

¹ Instituto de Biotecnología y Biología Molecular, Dep. Ciencias Biológicas - Facultad de Ciencias Exactas - Universidad Nacional de La Plata, CONICET CCT-LaPlata

Keywords: DLIS, Insertion sequence, transposition, symbiosis, rhizobia

*Corresponding author

Email address: mlozano@biol.unlp.edu.ar

ABSTRACT

Rhizobia are alpha- and beta- proteobacteria that, through the establishment of symbiotic interactions with leguminous plants, are able to fix atmospheric nitrogen as ammonium. The successful establishment of a symbiotic interaction is highly dependent on the availability of nitrogen sources in the soil, and on the specific rhizobia strain. Insertion sequences (ISs) are simple transposable genetic elements that can move to different locations within the host genome and are known to play an important evolutionary role, contributing to genome plasticity by acting as recombination hot-spots, and disrupting coding and regulatory sequences. Disruption of coding sequences may have occurred either in a common ancestor of the species or more recently. By means of ISCompare, we identified Differentially Located ISs (DLIS) in nearly related rhizobial strains of the genera *Bradyrhizobium*, *Mesorhizobium*, *Rhizobium* and *Sinorhizobium*. Our results revealed that recent IS transposition events don't seem to be playing a major role in adaptation. Nevertheless, DLIS could have a role enabling the activation and inactivation of certain genes that could dynamically affect the competition and survival of rhizobia in the rhizosphere.

INTRODUCTION

Rhizobia are alpha- and beta- proteobacteria that, through the establishment of symbiotic interactions with leguminous plants, are able to fix atmospheric nitrogen as ammonium. As a result of this interaction, rhizobia induce the development of a specialized root structure, the nodule, where they differentiate into bacteroids with the ability to fix N₂, providing the plant with a source of nitrogen in exchange for photosynthetic carbon mostly in the form of dicarboxylic acids (Gourion et al., 2015; Jones et al., 2007; Oldroyd et al., 2011; Poole et al., 2018; Shumilina et al., 2023). These bacteria are thus of particular interest due to their potential to improve crop yields and reduce the need for synthetic nitrogen fertilizers (Herridge et al., 2008).

The successful establishment of a symbiotic interaction between rhizobia and legume plants is highly dependent on the availability of nitrogen sources in the soil, and on the specific rhizobia strain, being some strains highly competitive for the colonization of plant roots but presenting low N₂-fixing efficiency (Checcucci et al., 2017). The colonization of the rhizosphere –the region of soil under the influence of plant roots– is the earliest step in the symbiotic process, and determines which rhizobia will end up in the nodules (Salas et al., 2017; Triplett and Sadowsky, 1992). The rhizosphere is a complex and dynamic environment where the interaction between plants, microorganisms, and soil particles takes place, thus playing a crucial role in plant growth and health (Walker et al., 2003). Nevertheless, the rhizosphere is not solely inhabited by rhizobia, but also by a diverse community of bacteria, fungi, and other microorganisms that compete for nutrients, but also for access to the host plant. The colonization of the rhizosphere is therefore a complex process that involves both plant-microbe interactions and competition with other rhizospheric organisms, and is influenced by various factors, including its growth rate, motility, and production of enzymes and other metabolites (Salas et al., 2017; Wheatley et al., 2020). Understanding the mechanisms and factors that influence rhizobia ability to colonize the rhizosphere can provide some insight into the dynamics of the rhizobia-legume symbiosis and potentially inform strategies for improving plant growth and productivity.

The genomes of rhizobia are very plastic, and are usually constituted by one chromosome, chromids –replicons of great size, difficult to distinguish from plasmids–, and several plasmids, some of them of great size –megaplasmids– and some smaller, generally considered as part of the accessory genome (Landeta et al., 2011; MacLean et al., 2007; Poole et al., 2018). Most of the genes required for the establishment of the symbiotic interaction with legume plants are located in such

megaplasmids, or in certain cases, in chromosomal integrative and conjugative elements (ICEs) (Poole et al., 2018), and there is evidence suggesting that they are actively mobilizing within rhizobial populations (Wheatley et al., 2020). Furthermore, rhizobial genomes contain a great number of insertion sequences (IS) (Kaneko, 2002, 2000; Nelson et al., 2018, this work).

Insertion sequences (ISs) are simple transposable genetic elements that can move to different locations within the host genome (Siguier et al., 2014). They are known to play an important evolutionary role, contributing to genome plasticity by acting as recombination hot-spots, and disrupting coding and regulatory sequences (Consuegra et al., 2021; Iida et al., 2015; Siguier et al., 2014; Vandecraen et al., 2017). In rhizobia, genome architecture has been shown to be under constant modification, with replicons continuously being cointegrated and excised (Guo et al., 2003). Moreover, the role of tandemly repeated ISs as drivers of genomic recombination events was demonstrated in artificial evolution experiments (Arashida et al., 2021). ISs can also modify the expression of genes, being a known case in rhizobia the mutation of *Sinorhizobium meliloti* 1021 *expR* –a LuxR family transcriptional regulator that controls the expression of the symbiotically active exopolysaccharide (EPS) EPS II– by the insertion of a ISRm2011-1 insertion sequence (Pellock et al., 2002). However, the great amount of partial –most likely inactive– transposases suggest that some of these insertions could have taken place in a more distant common ancestor. The role of recent transposition events in rhizobia has not been thoroughly explored thus far.

In this study, we used ISCompare (Mogro et al., 2021) software to identify ISs that have changed their location in nearly related rhizobial strains, and analysed the genes disrupted by these differentially located ISs (DLISs). Through this study, we have gained a better understanding of the impact of recent IS transposition events on rhizobial lifestyle.

MATERIALS AND METHODS

Identification of differentially located ISs with ISCompare

ISCompare was used to compare the genomes of several rhizobial strains belonging to *Bradyrhizobium diazoefficiens*, *Bradyrhizobium japonicum*, *Mesorhizobium ciceri*, *Mesorhizobium loti*, *Rhizobium etli*, *Rhizobium leguminosarum*, and *Sinorhizobium meliloti*. Genome sequences were obtained from the National Center for Biotechnology Information (NCBI) assembly database (Kitts et al., 2016). The accession numbers for the sequences used in this study are compiled in Table S1.A.

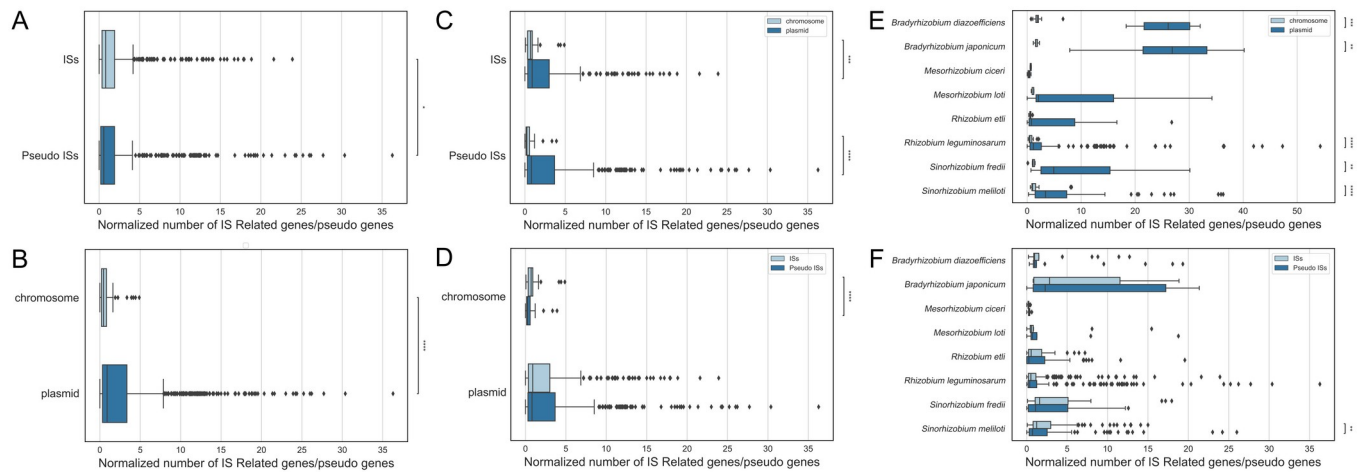


Figure 1. Distribution of transposase and IS related genes. A. Normalized IS and pseudo IS counts. B. Normalized number of IS elements by replicon. C-D. Normalized numbers of ISs and pseudo ISs by replicon. E. Normalized number of IS elements by species and replicon. F. Normalized number of IS and pseudo ISs by species. The number of transposases and IS related genes was estimated from the genbank files annotation using custom python scripts that looked for the terms 'transposase' and 'insertion seq' in the product descriptions of coding sequences and pseudo genes. Normalized counts were calculated as ISs / 100,000 bp. Significance was determined using the Mann-Whitney test and the normalized counts (*: $1.00e-02 < p \leq 5.00e-02$; **: $1.00e-03 < p \leq 1.00e-02$; ***: $1.00e-04 < p \leq 1.00e-03$; ****: $p \leq 1.00e-04$).

The sequences were downloaded in FASTA, CDS FASTA, and GBFF format. The sequences used as IS database for ISCompare were downloaded from <https://github.com/thanhleviet/Isfinder-sequences>.

ISCompare (Mogro et al., 2021) is a tool used to find differential located ISs between two closely related strains. Here we made pairwise comparisons using a selected strain as reference and all the complete closed genomes available as of 5/Jan/2023. To account for the most confident results only the results classified as DLIS by ISCompare, and with a complete IS match, were analysed in this study. The software was run with the following parameters: E-value = $1e-10$, minLength = 50, ISdiff = 50, scaffoldDiff = 20, minAlnLength = 50, surroundingLen = 500, surroundingLen2 = 500, shift = 0. The total number of DLIS located in accessory replicons –plasmids, megaplasmids– or in the main replicon –chromosome– was determined. DLIS counts were normalized by the species average size of the main and accessory replicons –the number of DLIS per 100,000 base pairs– to account for the difference in their sizes. The statistical analysis was done with python scripts using scipy (Virtanen et al., 2020). For the comparison of proportions Chi2 test was used summing the counts over all the strains within each species.

Distribution of IS in rhizobia

A custom python script was used to count the genes and pseudo-genes annotated as transposase or being part of an insertion sequence. In order to obtain this information, genbank files were processed with Biopython (Cock et al., 2009) and if the words 'transposase' or 'insertion seq' were found in a feature 'product' qualifier, a transposase was counted. In addition, the replicon information –transposon counts per replicon– was saved for further analysis.

DLIS functional analysis

To determine whether an IS was located within an intergenic region that could correspond to a promoter, the following analysis was done. First, the two closest CDS to a DLIS insertion site were found using the "closest" command from bedtools software (Quinlan and Hall, 2010). When the insertion site was located within 150 nucleotides upstream of a CDS, the DLIS was considered to be interrupting a probable promoter region. To identify DLIS interrupting a probable operon, the two closest CDS to the insertion site were determined with bedtools "closest" command, and the orientation and a distance between these CDS was analysed. A possible operon was considered if the two CDS were in the same orientation, and separated by a maximum of 100 nucleotides.

In the case of DLIS inserted within coding sequences, a functional classification was made using COG (Cluster of orthologous groups database, Galperin et al., 2021). COGs were assigned to each gene using the eggNOG-mapper web server with the eggNOG 5 database (Huerta-Cepas et al., 2019).

Figures were plotted using python matplotlib (Hunter, 2007) and seaborn (Waskom, 2021). Whenever it was required, final figures

were edited with Inkscape v 1.3 (The Inkscape team, <https://inkscape.org/>).

RESULTS

Analysis of IS and DLIS distribution in rhizobia

The genomes of all *B. diazoefficiens*, *B. japonicum*, *M. ciceri*, *M. loti*, *R. etli*, *R. leguminosarum*, and *S. meliloti* strains containing a complete assembly level were downloaded from NCBI Assembly database (Kitts et al., 2016, accessed 5/Jan/2023) and pairwise comparisons to identify DLIS were made with ISCompare (Table S1.B). For each species, a specific strain was used as reference for all pairwise comparisons. In order to estimate the proportion of active IS –DLIS with respect to the total number of IS– the total number of IS in a determined genome was estimated as the number of transposases plus the number of other genes annotated with functions related to insertion sequences (Figure 1, Table S1.C). Although the values are not exact –some IS may have been counted twice–, the estimated numbers of total ISs were in accordance with the previously reported (Kaneko, 2002, 2000; MacLean et al., 2007; Nelson et al., 2018). A recurrent observation is that chromosomal replicons tend to have lower number of total ISs whereas plasmids have a considerably higher amount (Figure 1.B, Table S1.E). This does not appear to be the case for DLIS (Figure 2.B., Table S2) which were detected with comparable frequency in plasmids (less but not significantly so) and chromosomal replicons. Of the rhizobia studied here, *Sinorhizobium* and *Bradyrhizobium* were the ones that showed the greatest number of DLIS (Figure 2.A) which agrees with the fact that they also present the highest number of ISs in their genomes (Figure 1.E-F, Table S1.C). Further, they also presented the highest proportion of IS that have recently changed their location in the genome, calculated as the number of DLIS / (total IS on ref strain + total IS on target strain) (Figure 3.A). Finally, a weak but significant correlation between the number of DLIS and the total number of IS was observed (Figure 3.B, $R^2 = 0.11$; p-value = 4.55 e-5).

Next, the distribution of DLIS insertion sites within genes, intergenic regions and possible operons was determined. Most DLIS presented insertion sites in intergenic regions (ca. 60%, Figure 4.A, Table S3.A) and the lowest proportion was observed for DLIS inserted within putative operons. *Mesorhizobium* and *Rhizobium etli* were the only ones that presented a higher proportion of DLIS inserted within coding sequences (Figure 4.B-C), although they also presented the lowest numbers of DLIS. Remarkably, most of the intergenic DLIS appear not to interrupt promoter regions (Figure 4.C).

Functional distribution of genes interrupted by DLIS

To analyse the cellular functions affected by DLIS insertions, we recovered the amino acid sequences of the complete version of the proteins from the genomes that didn't have the insertion sequence. All these genes were annotated with COGs using the eggNOG-mapper web server (Huerta-Cepas et al., 2019).

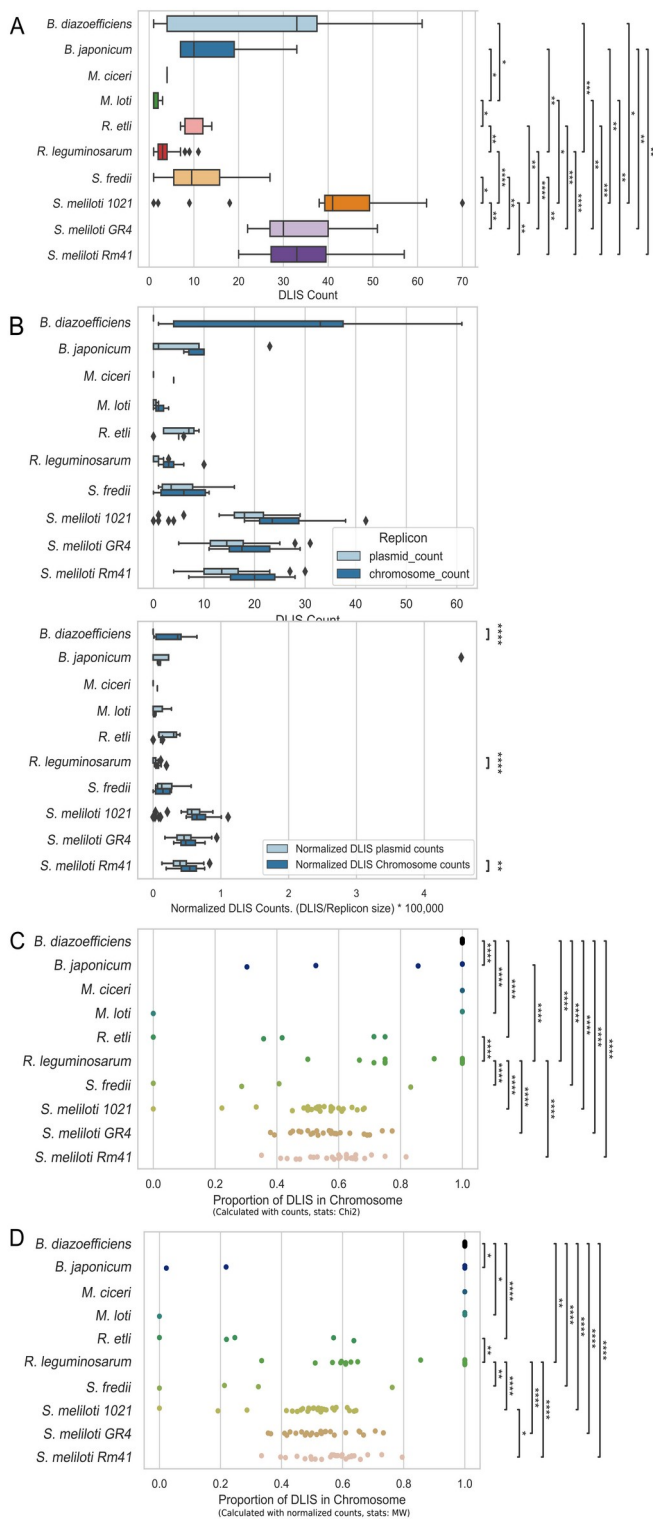


Figure 2. Distribution of DLIS. Chromosomes vs plasmids. A. Total DLIS counts. B. DLIS counts and normalized DLIS counts discriminated by rhizobia and replicon type. C. Proportions of DLIS in chromosomal replicons, calculated from DLIS counts. D. Proportions of DLIS in chromosomal replicons, calculated from DLIS normalized counts. DLIS were identified with ISCompare, and those with a full match for an insertion sequence and confidently identified as DLIS were analyzed. Significance was determined either using Chi2 test and the total DLIS counts in each species, or the Mann-Whitney test and the normalized DLIS counts (DLIS counts / 100,000 bp). *: $1.00e-02 < p \leq 5.00e-02$; **: $1.00e-03 < p \leq 1.00e-02$; ***: $1.00e-04 < p \leq 1.00e-03$; ****: $p \leq 1.00e-04$.

In the first analysis we used all the genes with an inserted DLIS (Figure 5). In such a case, each time a DLIS was inserted into a coding sequence of the genome used as reference, the disrupted gene could have been counted multiple times, as there could be many target genomes without the DLIS. In the second analysis, we only used the proteins with a DLIS insertion in the target genome, in that way, the complete sequence always corresponded to the reference genome which allowed us to avoid overcounting certain functions (Figure S1).

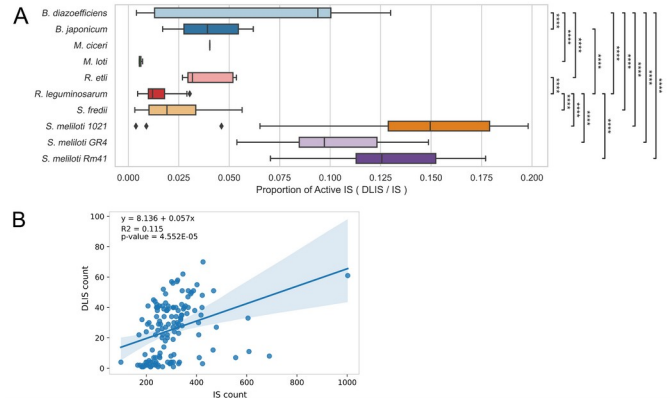


Figure 3. Proportion of recently active ISs. A. Proportion of active ISs. The proportion of recently active ISs was estimated as the relation of DLIS over the total IS count. Significance was determined using the Mann-Whitney test and the proportions of active ISs (DLIS / IS). *: $1.00e-02 < p \leq 5.00e-02$; **: $1.00e-03 < p \leq 1.00e-02$; ***: $1.00e-04 < p \leq 1.00e-03$; ****: $p \leq 1.00e-04$. B. Pearson correlation of IS counts and DLIS counts.

The overall results from the two analysis show that the top five COG categories –with the greatest number of DLIS– were S (Function unknown), K (Transcription), L (Replication, recombination and repair), M (Cell wall/membrane/envelope biogenesis) and Q (Secondary metabolites biosynthesis, transport and catabolism). L, S and K were also the principal categories in most rhizobia. Proteins with a classification in categories S and (-) corresponded majorly to hypothetical and conserved hypothetical proteins, or to proteins with unknown function. However, several proteins in the S category presented a more descriptive function, most remarkably some proteins with putative cellulose biosynthesis, chitinase, dihydroxyacetone kinase and nitrile hydratase activities. The L category contained DLIS that in most cases were inserted within genes encoding for transposases, integrases and reverse transcriptases, with only a few DLIS inserted in genes annotated as involved in DNA repair, nucleases, DNA polymerases, recombinases and methylases. For the K category, most of the proteins interrupted by DLIS corresponded to different families of transcriptional regulators, acetyltransferase domain containing proteins, and proteins involved in plasmid partition. Particularly, NodD1 and NifA presented DLIS insertions in *B. japonicum* USDA 6 and *S. meliloti* AK76 respectively. Of interest to us was that several proteins in the M (Cell wall/membrane/envelope biogenesis) COG category were interrupted by DLIS.

In particular, a capsular polysaccharide biosynthesis export transmembrane protein, and a choline-glycine betaine transporter were mutated in *S. meliloti* 1021 and *S. meliloti* GR4 respectively. Also *lpxK*, involved in the modification (phosphorylation) of lipid IVa, was interrupted in *S. meliloti* GR4 and 1021 genomes. In the E category, most of the insertions were within different enzymes annotated as arginine/lysine/ornithine decarboxylase, aspartate/tyrosine/aromatic aminotransferase, choline dehydrogenase, ethanolamine ammonia lyase, and phosphoribosyl anthranilate isomerase (Usg). Other insertions that could be relevant were observed in proteins belonging to the categories T (Signal Transduction) and P (Inorganic ion transport and metabolism). In particular DLIS insertions were observed in adenylate/guanylate cyclases, in two component systems (both in Histidine kinases and response regulators), and in several Na^+/H^+ , Mg^{2+}/Co^{2+} , Fe^{3+} and K^+ transport proteins. In the G category (Carbohydrate metabolism and transport) DLIS were found mainly in ABC transporters including a ribose/xylose/arabinose/galactoside and Branched-chain amino acid transport systems. Finally in the C category (Energy production and conversion), proteins annotated as aldehyde dehydrogenase, nitrite reductase, cytochrome bd terminal oxidase, and pyruvate ferredoxin/flavodoxin oxidoreductase activities presented DLIS insertions.

Certain genes presented DLIS insertions that were overcounted because they were mutated in the reference genome. These included genes annotated as major facilitator superfamily (MFS) transporter, hypothetical protein with adenylate kinase activity, polysaccharide biosynthesis protein (*lpsB2*), GNAT family N-acetyltransferase, LuxR family transcriptional regulator, Lrp/AsnC ligand binding domain-containing protein, LysR family transcriptional regulator, magnesium and cobalt transport protein CorA, methyltransferase domain-containing protein (annotated as NodS by eggNOG mapper), dihydroxyacetone kinase (*Dak2*), DedA family protein, Nitrile hydratase, RES domain-containing protein, tripartite tricarboxylate transporter, VIT family protein and the previously mentioned *lpxK*.

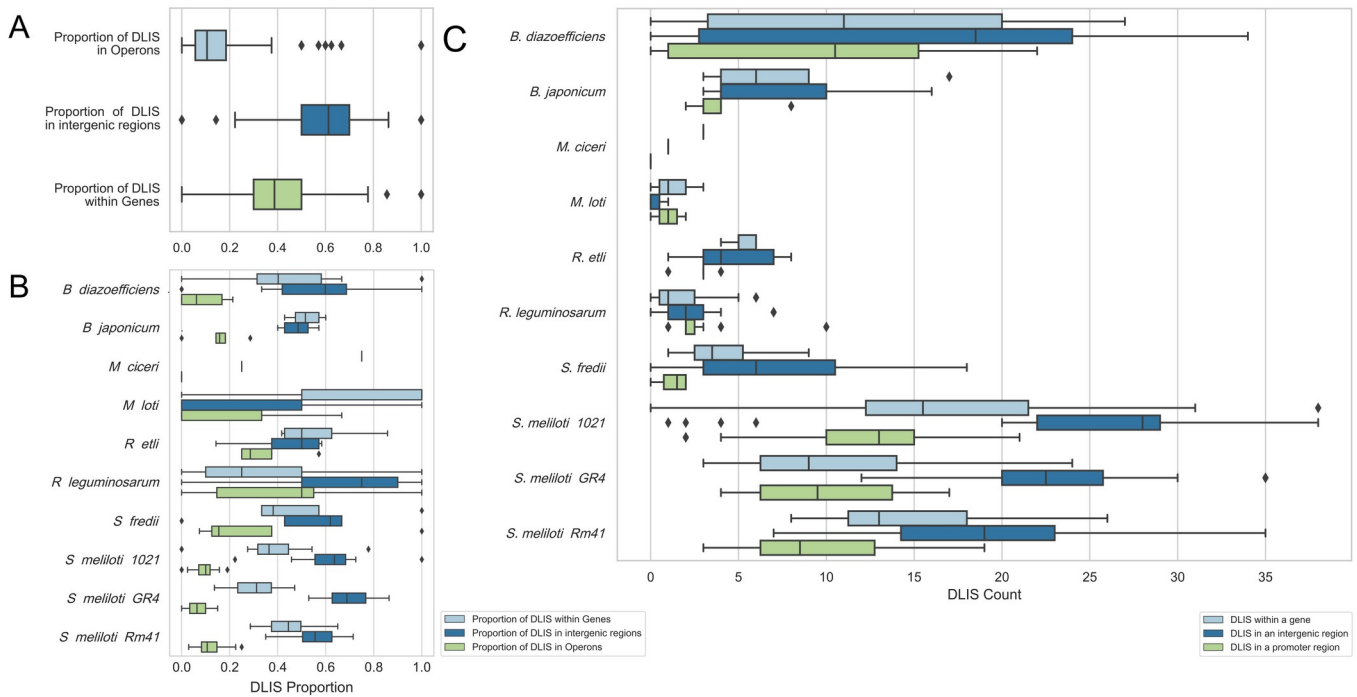


Figure 4. Distribution of DLIS in coding and intergenic regions. A. Proportion of DLIS inserted within genes, intergenic regions and putative operons. B. Proportion of DLIS inserted within genes, intergenic regions and putative operons discriminated by rhizobial species. C. Absolute count numbers of DLIS within genes, intergenic regions, and putative promoter regions. The count of the promoter regions is a subset of the counts for intergenic regions. When a DLIS was inserted within a short intergenic region between two divergent genes it was counted twice since it could be affecting the transcription of both genes. The location of the DLIS was determined using custom python scripts and the bedtools software as described in material and methods. Significance was determined using the Mann-Whitney test and the normalized counts.

DISCUSSION

Rhizobial plasmids are known to have a greater number of ISs than chromosomes and it has been shown that they can act as mediators of homologous recombination leading to genomic rearrangements (Guo et al., 2003). Surprisingly, in the case of DLIS, a greater number was observed in chromosomal replicons (Figure 2.C-D). DLIS are ISs inserted into highly conserved regions in a genome of a pair of closely related bacterial genomes and could therefore be broadly considered as active ISs. Our results appear to indicate that ISs tend to be more active in chromosomal replicons than in plasmids. This result was not expected, and could be a consequence of the high proportion of pseudo IS elements in plasmids (Figure 1.D), which, although perfectly capable of mediating homologous recombination, are inactive. To confidently identify recent transposition events, in this work we only considered DLIS with complete blast hits to ISs in the library. However, ISCompare also identifies DLISs involving partial IS blast hits which could be related to more ancestral insertions and were not considered.

Of the analysed rhizobia genera, *Sinorhizobium* and *Bradyrhizobium* presented the higher proportion of ISs, DLISs and active ISs. Furthermore, we found a weak but significant correlation ($r=0.115$, $p\text{-value}=4.552 \times 10^{-5}$, Figure 3.B) between the number of ISs and the number of DLISs. Thus, a more important contribution of ISs to environmental adaptation could be expected for genomes with greater numbers of ISs. However, most rhizobial DLISs were found to be inserted within intergenic regions, and far from the 5'-ends of coding sequences (Figure 4), suggesting that ISs might play a more important role in genome plasticity and adaptation by acting as recombination hot-spots than through transposition per se. This interpretation has to be taken with care, since intergenic regions could encode for small ORFs and RNAs not yet annotated. Furthermore, only a low proportion of the identified DLIS are expected to generate polar mutations by interrupting putative operons (Figure 4.A-B).

Regarding the interrupted genes, the majority corresponded to COG categories L (mostly transposases), S (unknown function) and - (not assigned), which reinforces that, in rhizobia, the adaptive mutations produced by IS transposition may be of less importance than those produced by IS-mediated homologous recombination. Nevertheless, we were able to identify some DLIS inserted within genes with a more precise annotation (Table 1). Some of those genes were reported to have a role in the symbiotic interaction with the plant host. Three of them corresponded to enzymes involved in the biosynthesis of the

lipopolysaccharide (LPS), *lpxK*, *lpsB2* and *kdsD*. LpxK catalyses the sixth step in the lipid A synthesis (Emptage et al., 2013); while LpsB2 is required for O-antigen biosynthesis and its mutants presented reduced motility, grew faster than the parental strain, and were more sensitive to maize benzoxazinones and polymyxin B (García-de los Santos and Brom, 1997; Ormeno-Orrillo et al., 2008). KdsD is required for the biosynthesis of 3-deoxy-d-manno-octulosonic acid (Kdo), a key sugar in the core region of LPS (Jenkins et al., 2023).

Furthermore, a DLIS inserted within *rpkl*, a gene encoding for a protein involved in capsular polysaccharide biosynthesis, was found in *S. meliloti* USDA1157. *rhiA*, the first gene of the rhizosphere-expressed genes operon (*rhiABC*), that was reported to influence nodulation (Cubo et al., 1992), was found to be interrupted by a DLIS in *R. leguminosarum* Vaf10 strain.

The *nifA* gene of *S. meliloti* –a regulatory nitrogen fixation gene required for the induction of several key *nif* and *fix* genes (Agron et al., 1993)– presented an inserted DLIS on strain AK76. NifA null mutants induce white nodules in the roots of the host plant, have reduced swarming ability, and present lower levels acyl-homoserine lactones and of extracellular proteins (Gong et al., 2007). In *B. japonicum* USDA 6, NodD1, the positive regulatory protein of the *nodYABC* operon (Stacey, 1995), was found to be interrupted by a DLIS. In that case, a shorter ORF is annotated in the genome, and it could be possible that its product is still functional.

Another protein that was found to be interrupted by a DLIS is NirB. NirB, was shown to be involved in the nitrate assimilatory pathway, and to participate indirectly in NO synthesis, possibly contributing with the denitrification pathway (Ruiz et al., 2019), and thus, to the generation of greenhouse gases. GshA (glutathione synthetase) involved in glutathione biosynthesis, was interrupted in *B. japonicum* CC829. Glutathione has been shown to be important for growth under defined conditions, and to play an important role in symbiosis (Harrison et al., 2005; Sobrevals et al., 2006; Taté et al., 2012). Finally, Carbonic anhydrase (CA) –an enzyme involved in the interconversion of carbon dioxide and bicarbonate– was found to be interrupted in *S. meliloti* AK57. Although CA was shown not to be essential for nodule development and nitrogen fixation, it was hypothesized that it could help in the protonation of extracellular NH₃ facilitating its diffusion and transport to the plant tissues (Kallonati et al., 2009).

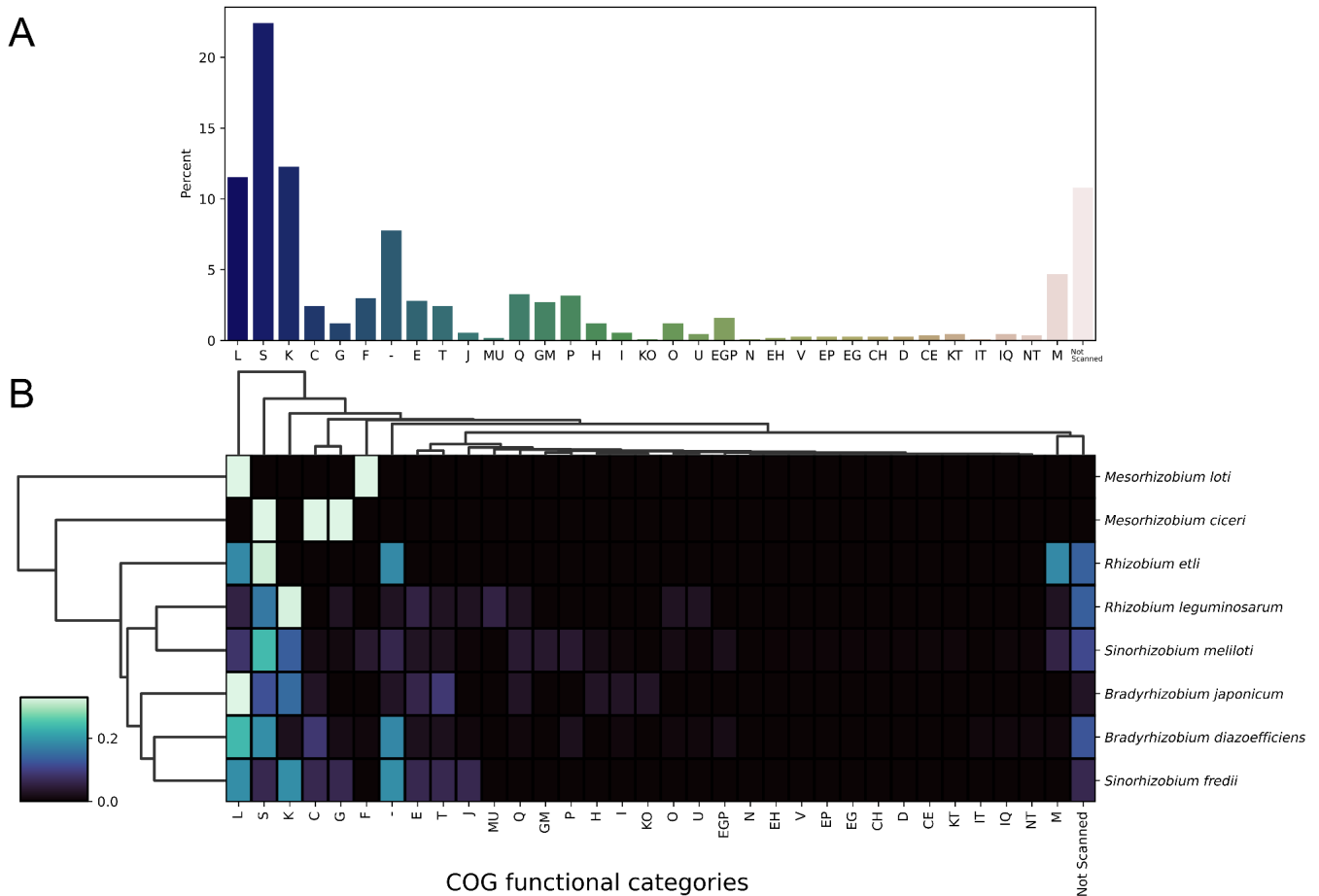


Figure 5. Distribution of COG functional categories. COGs functional categories for all the genes interrupted by a DLIS were assigned with eggNOG-mapper. A. Distribution of DLIS interrupted genes by COG functional category. B. The heatmap was generated with a custom python script using seaborn. COG categories are as follows: -: Not assigned; A: RNA processing and modification; B: Chromatin Structure and dynamics; C: Energy production and conversion; D: Cell cycle control and mitosis; E: Amino Acid metabolism and transport, F: Nucleotide metabolism and transport, G: Carbohydrate metabolism and transport; H: Coenzyme metabolism; I: Lipid metabolism; J: Translation; K: Transcription; L: Replication and repair; M: Cell wall/membrane/envelope biogenesis; N: Cell motility; O: Post-translational modification, protein turnover, chaperone functions; P: Inorganic ion transport and metabolism; Q: Secondary metabolites biosynthesis, transport and catabolism; R: General Functional Prediction only; S: Function Unknown; T: Signal Transduction; U: Intracellular trafficking and secretion; V: Defense mechanisms; W: Extracellular structures; X: Mobilome: prophages, transposons; Y: Nuclear structure; Z: Cytoskeleton.

Other genes encoding proteins not directly implicated in the interaction of rhizobia with their plant host, but that could have a role in rhizosphere colonization, were also found to be interrupted by DLISs, including *aspB* (aromatic aminotransferase), *acsA1* (Acyl-coenzyme A synthetases), *nthA* (nitrile hydratase subunit alpha), *cydB* (cytochrome d ubiquinol oxidase subunit II), *surF* (SURF1-like protein), *fhs* (formate-tetrahydrofolate ligase), *gabD* (Succinate-semialdehyde dehydrogenase), GlcF (glycolate oxidase complex, iron-sulphur subunit), *idnO* (Gluconate 5-dehydrogenase, GA5DH), *metB* (Cystathionine gamma-synthase), *pyk* (pyruvate kinase), and *eutB* (ethanolamine ammonia-lyase, large subunit). *AspB*, catalyses the transfer of an alpha amino group from aromatic amino acids, but also from aspartate, to different substrates. It was reported that it contributed, under high levels of exogenous tryptophan, to the biosynthesis of indole-acetic acid (IAA), and to nitrogen scavenging under nitrogen deprivation (Kittell et al., 1989). Also related to the IAA metabolism, *nthA* (coding for nitrile hydratase subunit alpha) was found interrupted by a DLIS in *S. meliloti* RM41 (Liu et al., 2017). An insertional mutation in *acsA1* was found in *S. meliloti* strain RCAM1115 which might impede its growth using acetate as carbon source (Aneja et al., 2002). *GabD* –Succinate-semialdehyde dehydrogenase– homologs are required for growth on γ -aminobutyrate (GABA) as the sole nitrogen source (Prell et al., 2009). GA5DH catalyses the transformation of 5-keto-d-gluconate to d-gluconate – that can enter the Entner-Doudoroff pathway to provide a source of carbon and NADP^{+} when energy is exhausted. However, under conditions of energy abundance, this enzyme catalyses the oxidation of d-gluconate to 5-keto-d-gluconate which can be used as a transient means of energy storage (Zhang et al., 2009).

In the transport systems category several genes presented DLIS insertions in specific strains which could impair the absorption of nutrients, and thus their ability to compete for rhizosphere colonization. Among them, *afuA* (ABC-type Fe^{3+} transport system, periplasmic component), *corA* (magnesium and cobalt transport protein),

MA20_24460 (sulfite exporter TauE/SaFe family protein), *tauZ* (putative sulfate exporter) *macA* (RND family efflux transporter MFP subunit), *modC* (Molybdenum import ATP-binding protein ModC), *rhcU* (type III secretion system export apparatus switch protein), *msmE* (Multiple sugar ABC transporter substrate-binding protein), *nhaA* (Na⁺/H⁺ antiporter), *ybaL* (monovalent cation proton antiporter 2) and *pstS* (phosphate ABC transporter substrate-binding protein PstS). Of these, *CorA* was shown to be important for growth on glucose at 21% O₂ (Wheatley et al., 2017). *MacA* is the membrane fusion protein of a ABC-Type efflux transporter, a type 1 secretion system that transports diverse molecules (antibiotics and peptides) across the inner and outer membranes (Alav et al., 2021; Eda et al., 2011). *ModC* is required for nitrogen fixation on limiting levels of molybdate and was reported to be the high-affinity molybdate transporter in *B. diazoefficiens* (Cheng et al., 2016; Tresierra-Ayala et al., 2011).

In the case of *NhaA*, it was shown to be induced under acid conditions, and it is likely that it functions by expelling protons toward the periplasm (Guerrero-Castro et al., 2018). Last, *pstS* is a part of the *pstSCAB* phosphate-specific transport operon that functions as high-affinity phosphate transporter (Guerrero-Castro et al., 2018). Moreover, an homologous of *Smc03123* (putative transcriptional regulator), a gene from *S. meliloti* 2011 which mutant was shown to have an altered fitness in the competence for rhizosphere colonization at 3 days post-inoculation (Salas et al., 2017), exhibited a DLIS in *S. meliloti* USDA1157.

Competition in the rhizosphere requires rhizobia to be able to use diverse compounds as carbon and nitrogen sources. Possessing a wide repertoire of metabolic pathways and transporters, enable rhizobia to thrive and compete for the nutrients exuded to the rhizosphere by the plant host. Dynamic inactivation of genes by DLIS could play an important role, granting a faster growth capability. However, it could also be detrimental, affecting posterior steps that occur during the interaction with the plant host, or under changing environmental conditions.

Table 1. List of genes interrupted by a DLIS and annotated with a gene name in the reference strain.

Gene Name	Product	Organism
<i>acaP</i>	Carbonic anhydrase	<i>S. meliloti</i> Ak57
<i>acsA1</i>	Acetyl-coenzyme A synthetase 1 – hypothetical protein	<i>S. meliloti</i> RCAM1115
<i>afuA</i>	ABC-type Fe ³⁺ transport system, periplasmic component	<i>S. meliloti</i> HM006
<i>aglR</i>	HTH-type transcriptional regulator AglR	<i>S. meliloti</i> Ak57
<i>aspB</i>	Aspartate/tyrosine/aromatic aminotransferase	<i>S. meliloti</i> 1021
<i>corA</i>	magnesium and cobalt transport protein CorA	<i>S. meliloti</i> GR4
<i>cydB</i>	cytochrome d ubiquinol oxidase subunit II	<i>B. diazoefficiens</i> NK6
<i>degP4</i>	Trypsin-like serine protease, periplasmic, C-terminal PDZ domain	<i>S. meliloti</i> USDA1106
<i>deoR1</i>	transcriptional regulator protein, Putative sugar-binding domain	<i>S. meliloti</i> USDA1157
<i>dinB</i>	DNA polymerase IV 2, error-prone DNA polymerase	<i>S. meliloti</i> HM006
<i>dnaB</i>	putative replicative DNA helicase	<i>S. meliloti</i> GR4
<i>eutB</i>	Ethanolamine ammonia lyase large subunit	<i>S. meliloti</i> RMO17
<i>fhs</i>	formate–tetrahydrofolate ligase	<i>M. loti</i> 582
<i>flaR</i>	Adenylate kinase; ATPase AAA, CRISPR/Cas system-associated protein Cas3	<i>S. meliloti</i> RU11_001, Rm41
<i>gabD</i>	Succinate-semialdehyde dehydrogenase [NADP(+)] GabD	<i>S. meliloti</i> AK83
<i>glcF</i>	putative glycolate oxidase iron-sulfur subunit	<i>S. meliloti</i> GR4
<i>groS</i>	co-chaperone GroES	<i>B. diazoefficiens</i> F07S3
<i>gshA</i>	glutathione synthetase	<i>B. japonicum</i> CC829
<i>idnO</i>	Gluconate 5-dehydrogenase	<i>S. meliloti</i> M162
<i>kdgK</i>	2-dehydro-3-deoxygluconokinase, pfkB family carbohydrate kinase	<i>M. ciceri</i> R30
<i>kdsD</i>	arabinose-5-phosphate isomerase, SIS family. GutQ KpsF subfamily	<i>S. meliloti</i> KH35c
<i>lpsB2</i>	polysaccharide biosynthesis protein CapD	<i>S. meliloti</i> 1021
<i>lpxK</i>	tetraacyldisaccharide 4'-kinase	<i>S. meliloti</i> 1021, GR4
<i>luxR</i>	transcriptional regulator, autoinducer-binding domain-containing protein	<i>S. meliloti</i> RU11_001
MA20_06130	indolepyruvate ferredoxin oxidoreductase family protein	<i>B. diazoefficiens</i> USDA_110
MA20_08550	inner membrane transporter Tim44-related protein	<i>R. etli</i> CFN_42
MA20_09650	glutathionylspermidine synthase family protein	<i>B. diazoefficiens</i> F07S3
MA20_22950	acetyl-CoA carboxylase biotin carboxylase subunit	<i>B. diazoefficiens</i> NK6
MA20_24460	sulfite exporter TauE/SafE family protein	<i>R. leguminosarum</i> 248
MA20_31720	transglutaminase-like cysteine peptidase	<i>B. diazoefficiens</i> 172S4
MA20_35415	SDR family oxidoreductase	<i>B. diazoefficiens</i> NK6
MA20_36515	COG0790 FOG TPR repeat, SEL1 subfamily, Hypothetical protein	<i>B. diazoefficiens</i> 172S4
MA20_40030	Zn-dependent protease	<i>B. japonicum</i> USDA_6
<i>macA</i>	ABC-type family efflux transporter MFP subunit	<i>S. meliloti</i> Ak57
<i>metB</i>	Cystathionine gamma-synthase	<i>S. meliloti</i> KH46
<i>modC</i>	Molybdenum import ATP-binding protein ModC	<i>S. meliloti</i> AK83
<i>msmE</i>	Multiple sugar ABC transporter substrate-binding protein	<i>S. fredii</i> CCBau_45436
<i>nhaA</i>	Na ⁺ /H ⁺ antiporter	<i>S. meliloti</i> KH35c
<i>nifA</i>	Nif-specific regulatory protein	<i>S. meliloti</i> AK76
<i>nirB</i>	Nitrite reductase [NAD(P)H]	<i>S. meliloti</i> USDA1106
<i>nodD1</i>	LysR family transcriptional regulator	<i>B. japonicum</i> USDA_6
<i>norG</i>	Transcriptional regulator with HTH domain and aminotransferase MocR family domain	<i>S. meliloti</i> S35m
<i>nthA</i>	nitrile hydratase subunit alpha	<i>S. meliloti</i> Rm41
<i>parB</i>	ParB/RepB/Spo0J family partition protein	<i>R. leguminosarum</i> 31B
<i>pepT</i>	peptidase T	<i>R. leguminosarum</i> Vaf-108
<i>pstS</i>	phosphate ABC transporter substrate-binding protein PstS	<i>B. diazoefficiens</i> NK6
<i>psuG</i>	putative enzyme involved in pigment biosynthesis, pseudouridine degradation pathway	<i>S. meliloti</i> RCAM1115
<i>pyk</i>	pyruvate kinase	<i>B. diazoefficiens</i> NK6
QU41_18010	transcriptional regulator, Lrp/AsnC ligand binding domain	<i>S. meliloti</i> Rm41
<i>rbn</i>	ribonuclease BN (exoribonuclease), putative membrane protein	<i>S. meliloti</i> M162
<i>rhcU</i>	EscU/YscU/HrcU familyT3SS apparatus switch protein	<i>B. diazoefficiens</i> USDA_110
<i>rhiA</i>	RhiA, rhizosphere-expressed protein	<i>R. leguminosarum</i> Vaf10
<i>rkpl</i>	Capsular polysaccharide biosynthesis protein rkpl, Sulfatase	<i>S. meliloti</i> USDA1157
<i>sbcD</i>	exonuclease SbcCD subunit D	<i>B. diazoefficiens</i> USDA_110
<i>sqr</i>	NAD(P)/FAD-dependent oxidoreductase	<i>B. diazoefficiens</i> H12S4
<i>surF</i>	surfeit locus 1 family protein, SURF1-like	<i>S. meliloti</i> USDA1021
<i>tauZ</i>	UPF0324 membrane protein, TIGR00698 family, Putative sulfate exporter	<i>S. meliloti</i> AK21, GR4
<i>terC</i>	transmembrane protein	<i>S. meliloti</i> KH46
<i>tyrS</i>	tyrosine–tRNA ligase	<i>R. leguminosarum</i> RCAM2802
<i>usg</i>	Usg protein, hypothetical protein	<i>S. meliloti</i> KH35c, AK21
<i>virB6</i>	VirB6 type IV secretion protein	<i>S. meliloti</i> USDA1157
<i>ybaL</i>	Inner membrane protein ybaL, monovalent cation proton antiporter 2	<i>S. meliloti</i> AK83
<i>yceA</i>	Putative rhodanese-related sulfurtransferase	<i>S. meliloti</i> RU11_001
<i>ykuE</i>	Calcineurin-like phosphoesterase, hypothetical protein	<i>S. meliloti</i> KH46
<i>yqaA</i>	hypothetical membrane protein	<i>S. meliloti</i> 1021, Rm41
<i>ytfG</i>	NmrA-like family, hypothetical protein	<i>S. meliloti</i> SM11
<i>zapE</i>	cell division protein ZapE, AFG1-like ATPase	<i>B. diazoefficiens</i> XF7

CONCLUSION

In this work we analysed recent transposition events in rhizobia by identifying the DLISs by means of ISCompare. Our results revealed that recent IS transposition events don't seem to be playing a major role in adaptation. This hypothesis is supported by the vast majority of

DLIS being inserted within intergenic regions, and by the fact that most DLIS inserted within genes correspond to genes coding for hypothetical proteins. A more important role of ISs could be to mediate homologous recombination, allowing the dynamic interchange of information between chromosomal and plasmid replicons, and between different rhizobial species. Nevertheless, DLIS could have a role en-

