GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Phylogenetic analysis reveals gene conversions in multigene families of rhizobia

J.E. Hernandez-Salmeron and G. Santoyo

Laboratorio de Recombinacion y Diversidad Genomica,
Instituto de Investigaciones Quimico Biologicas,
Universidad Michoacana de San Nicolas de Hidalgo, Morelia,
Michoacan, Mexico

Corresponding author: G. Santoyo
E-mail: gsantoyo@umich.mx / gustavo_santoyo@yahoo.com

Genet. Mol. Res. 10 (3): 1383-1392 (2011)
Received October 21, 2010
Accepted January 16, 2011
Published July 12, 2011
DOI 10.4238/vol10-3gmr1118

ABSTRACT. Gene families are an important and intrinsic trait of rhizobial species. These gene copies can participate in non-reciprocal recombination events, also called gene conversions. Gene conversion has diverse roles, but it is usually implicated in the evolution of multigene families. Here, we searched for gene conversions in multigene families of six representative rhizobial genomes. We identified 11 gene families with different numbers of copies, genome location and function in CFN42 and CIAT652 strains of *Rhizobium etli*, *Rhizobium* sp NGR234, *Mesorhizobium loti* MAFF303099, *Sinorhizobium meliloti* 1021, and *Bradyrhizobium japonicum* USDA110. Gene conversions were detected by phylogenetic inference in the *nifD* and *nifK* gene families in *R. etli*. Sequence analysis confirmed multiple gene conversions in these two gene families. We suggest that gene conversion events have an important role in homogenizing multigene families in rhizobia.

Key words: Rhizobia; Gene conversion; Concerted evolution; Multigene families

INTRODUCTION

Bacteria usually contain multigene families as an intrinsic trait of their genomes (Achaz et al., 2002; Treangen et al., 2009). Some of them exhibit extraordinary similarity between all its members, which suggests a recent origin after duplication, or a homogenizing mechanism called gene conversion (Santoyo and Romero, 2005). Gene conversion has been widely documented in diverse multigene families, including the rRNA genes in *Escherichia coli* (Liao, 2000), *tuf* in *Salmonella typhimurium* (Abdulkarim and Hughes, 1996) and *Vibrio cholerae* (Lathe III and Bork, 2001), *pilE* in *Neisseria gonorrhoeae* and *N. meningitidis* (Haas and Meyer, 1986; Kline et al., 2003), and *hop* in *Chlamydia pneumoniae* and *Helicobacter pylori* (Jordan et al., 2001), among others (Santoyo and Romero, 2005). Gene conversion also plays essential roles in antigenic variation, an important mechanism in bacterial pathogens to evade the host immune system, as well as other positive and detrimental roles (Deitsch et al., 1997).

In rhizobial genomes, gene conversion has been studied only in the *nifH* family of *Rhizobium etli*, which codes for the enzyme nitrogenase (Rodriguez and Romero, 1998; Santoyo et al., 2005). However, it is known that *R. etli* and other rhizobial genera contain a huge amount of reiterated DNA sequences, including multigene families, which can be targets for gene conversion events (Gonzalez et al., 2006; Orozco-Mosqueda et al., 2009). Genome analysis of *R. etli* shows the presence of 133 families of identical repeats, while *Sinorhizobium meliloti* displays 24 long sequence repeats (Gonzalez et al., 2006; Orozco-Mosqueda et al., 2009). Other rhizobia such as *Mesorhizobium loti* and *Bradyrhizobium japonicum* have been completely sequenced, but the number of gene families is not mentioned (Kaneko et al., 2000, 2002). Although some rhizobia have a huge amount of identical repeats, not all of them belong to gene coding families.

Gene conversion is experimentally difficult to study in bacteria, due to the problem of recovering all products of a gene conversion event (Liao, 2000). Therefore, phylogenetic inference has been useful to study this mechanism in different gene families in bacteria. For example, multiple independent gene conversion events were detected in outer membrane proteins of *H. pylori* and *C. pneumoniae* (Jordan et al., 2001). In another study, gene conversions were also detected by phylogenetic evidence in the *tuf* genes of different bacteria (Lathe III and Bork, 2001).

In this study, gene conversion was surveyed in 11 multigene families of *R. etli*, strains CFN42 and CIAT652, by phylogenetic analysis. Gene conversions were also searched in orthologs of the multigene families in other complete rhizobial genomes, such as *Rhizobium* sp NGR234 (Schmeisser et al., 2009), *M. loti* MAFF303099 (Kaneko et al., 2000), *S. meliloti* 1021 (Galibert et al., 2001), and *B. japonicum* USDA110 (Kaneko et al., 2002).

MATERIAL AND METHODS

Identification of multigene families

Members of multigene families were identified by literature search in PUBMED (http://www.ncbi.nlm.nih.gov/pubmed) of the National Center for Biotechnology Information (NCBI). Eleven multigene families of the *R. etli* CFN42 and CIAT652 strains were chosen and analyzed in this study because of their biological significance. It is worth noting that other long DNA repeats are present in the *R. etli* genome; however, our study focused on gene coding families. The nucleotide sequence of each copy was obtained from GenBank (NCBI) with the following GenID

numbers: *R. etli* CFN42, *nifH* (1005068, 1005033, 1004995), *nifD* (1005069, 1004996), *nifK* (1005079, 1004997), *nodD* (6402815, 6402862, 6402864), *ccm* (3896046, 3896045, 3896047), *adhC* (3895921, 3892269), *tuf* (3892582, 3892567), *fix* (1005327, 1005055, 1005054), *groEL* (3894519, 3891062), *dnaJ* (3890921, 3890921, 3894442), and *purU* (6400149, 6398142); *R. etli* CIAT652, *nifH* (6402850, 6402804, 6402760), *nifD* (6402759, 6402805, 6402849), *nifK* (6402758, 6402848), *nodD* (6402815, 6402862, 6402864), *fix* (6402772, 6402773, 6402774), and *groEL* (6403417, 6399413); *Rhizobium* sp NGR234, *nifH* (962478, 962500), *nifD* (962502, 962482), *nifK* (962485, 962471), *fix* (962188, 962189), *groES* (7788251, 7791635), *groEL* (7792170, 7788321, 7788250, 7791634), and *dnaJ* (7792773, 7791032, 7791742); *R. leguminosarum*, *fix* (4403906, 4403966), *groEL* (4402951, 4398962, 4401984, 4398208), *etfA* (4401546, 4398246), *etfB* (4401547, 4398247), and *hemA* (4398963, 4402319); *M. loti*, *nifA* (1229107, 1229090, 1229044), *repA* (1224477, 1231331), *repB* (1224476, 1231330), *repC* (1224475, 1231329), *groES* (1231535, 1230998, 1226411, 1226287, 1229067), *groEL* (1231471, 1230997, 1226412, 1226286, 1229066), and *dnaJ* (1228214, 1230424); *B. japonicum*, *groES* (1053428, 1055491, 1048455, 1054913, 1051464), *groEL* (1053427, 1052432, 1055489, 1048307, 1051467, 1051462, 1047245), *rpoN* (1054956, 1054195), *dnaJ* (1052529, 1052410, 1049300, 1048922), and *fix* (1055510, 1054820, 1054853), and *S. meliloti nod* (1235512, 1235224, 1235494), *fix* (1235483, 1235482), *fixO3* (1235141, 1235447, 1235700), *groES* (1235290, 1235216, 1232434), and *groEL* (1235289, 1235215, 1237337, 1232840, 1232433).

Orthologs of 11 multigene families were also searched in other rhizobia by similarity blast genome research (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Genomes analyzed included *Rhizobium* sp NGR234 (NC_012587.1), *R. leguminosarum* (NC_008380.1), *B. japonicum* (NC_004463.1), *S. meliloti* (NC_003047.1), and *M. loti* (NC_002678.2).

Phylogenetic analysis

A multiple sequence alignment was generated with ClustalX, and the phylogenetic analysis of the multigene families was carried out with the MEGA 4.0 program (Tamura et al., 2007). To obtain a confidence value for the aligned sequence dataset, a bootstrap analysis of 1000 replications was done. Phylogenetic trees were constructed by the neighbor-joining method based on Kimura's two-parameter distance (Kimura, 1980). Other methods were also employed, but irrespective of that, the overall tree topologies were similar in all cases. Alignment analysis of the gene sequences was done with the help of the ClustalW program (http://www.ebi.ac.uk/Tools/clustalw2/).

RESULTS

Identification of multigene families in rhizobia

Rhizobial genomes are an interesting model to study recombination and genome dynamics mechanisms, due to the wide presence of multigene families. In some cases, these multiple gene copies are located on different replicons, either chromosome or plasmids. Recombination between these inter-replicon copies may lead to diverse genome rearrangements (Orozco-Mosqueda et al., 2009).

Here, we identified and analyzed 11 multigene families in two *R. etli* strains (CFN42

and CIAT652). Some of these gene families were also found in the genome of other rhizobia. For example, the chaperonin *groEL* gene family was found to be reiterated in all the rhizobial genomes analyzed in this study. The number of gene copies varied: 4 copies are present in *R. etli*, *Rhizobium* sp NGR234 and *R. leguminosarum*, 5 in *S. meliloti* and *M. loti*, and 7 in *B. japonicum*. Other gene families, including *nifH*, are present in two or three copies in the *Rhizobium* genera, but not in others. Table 1 shows the list and copy number of the gene families analyzed here for gene conversions. It is interesting that diverse gene copies are distributed in different replicons, which can be targets for recombination, thereby generating genome rearrangements such as deletions, inversions and amplifications. This is an interesting subject for further studies.

Table 1. Multigene families in the order Rhizobiales.

Genome	Multigene family	Function	Number of copies	Genomic localization
Rhizobium etli CFN42	*nifH*	Nitrogenase reductase	3	Plasmid
	nifD	Nitrogenase subunit α	3	Plasmid
	nifK	Nitrogenase subunit β	2	Plasmid
	nodD	Transcriptional regulator	3	Plasmid
	ccm	Cytochrome c	3	Chromosome/plasmid (3)
	adhC	Formaldehyde dehydrogenase	2	Chromosome/plasmid
	tuf	Elongation factor	2	Chromosome
	fix	Electron transfer	3	Plasmid
	groEL	Chaperonin	4	Chromosome (3)/plasmid
	dnaJ	Chaperone	3	Chromosome
	purU	Purine biosynthesis	2	Chromosome
Rhizobium sp NGR234	*nifH*	Nitrogenase reductase	2	Plasmid
	nifD	Nitrogenase subunit α	2	Plasmid
	nifK	Nitrogenase subunit β	2	Plasmid
	fixA	Electron transport	3	Plasmid
	groES	Chaperonin	2	Chromosome/plasmid
	groEL	Chaperonin	4	Chromosome (2)/plasmid (2)
	dnaJ	Chaperone	3	Chromosome
Rhizobium leguminosarum	*fix*	Electron transport	2	Plasmid
	groEL	Chaperonin	4	Chromosome (2)/plasmid (2)
	etfA	Electron acceptor	2	Chromosome/plasmid
	etfB	Electron acceptor	2	Chromosome/plasmid
	hemA	Coenzyme	2	Chromosome/plasmid
Mesorhizobium loti	*nifA*	Regulatory protein	3	Chromosome
	repA	Replication	2	Plasmid
	repB	Replication	2	Plasmid
	repC	Replication	2	Plasmid
	groES	Co-chaperonin	5	Chromosome (4)/plasmid
	groEL	Chaperonin	5	Chromosome (4)/plasmid
	dnaJ	Chaperone	2	Chromosome
Bradyrhizobium japonicum	*groES*	Co-chaperonin	5	Chromosome
	groEL	Chaperonin	7	Chromosome
	rpoN	Sigma factor	2	Chromosome
	dnaJ	Chaperone	4	Chromosome
	fix	Electron transport	3	Chromosome
Sinorhizobium meliloti	*nodD*	Regulation of nodulation	3	Plasmid
	fix	Electron transport	2	Plasmid
	fixO3	Cytochrome c oxidase	3	Plasmid
	groES	Chaperonin	3	Plasmid
	groEL	Chaperonin	5	Chromosome (2)/plasmid (3)

Phylogenetic analysis of multigene families

The hypothesis states that orthologs present in two strains or very closely related species are more evolutionarily related than paralogous genes. This is because orthologs shared a

more recent common ancestor at the time of strain divergence, while paralogous genes shared an ancestor at the time of duplication, previous to strain divergence (Jordan et al., 2001). When orthologous genes were not found as multigene families in some rhizobial genera, we were unable to prove the above hypothesis. However, it was really useful to have the complete genome sequence available for the two strains of *R. etli* (CFN42 and CIAT652).

Figure 1 shows the phylogeny of the *groEL* multigene family. This confirms the prediction that orthologous genes are more closely related than paralogs within the same genome, and therefore, evidence for gene conversions is lacking. In some cases, the *groEL* copies were grouped in the same clade for *M. loti* and *B. japonicum*, but these species are too divergent, and these orthologs were probably duplicated after species divergence. In addition, sequence analysis did not reveal gene conversions within this *groEL* family in any rhizobial analyzed here.

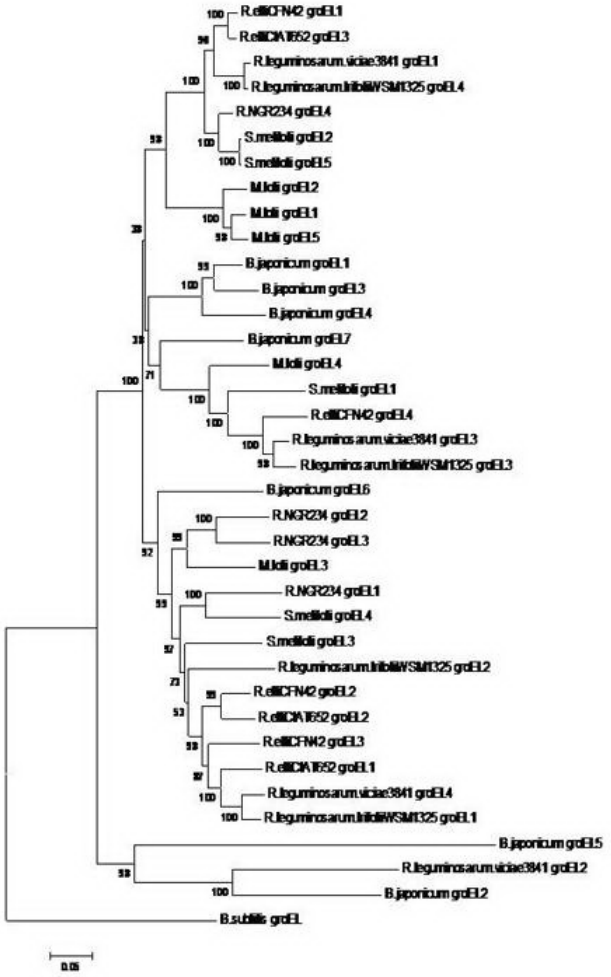

Figure 1. Phylogenetic tree of the *groEL* multigene family in diverse rhizobial genomes. A multiple sequence alignment was generated with ClustalX, and construction of the phylogeny was carried out by the neighbor-joining method of the MEGA 4.0 program. A bootstrap analysis of 1000 replications was done.

Interestingly, the phylogenies of the *nifD* and *nifK* multigene families in *R. etli* strains clearly showed evidence of apparent gene conversions, because paralogs are more related than orthologs (Figures 2 and 3). The phylogenies of the rest of the multigene families in all the rhizobial genomes studied here did not show evidence of gene conversions (data not shown).

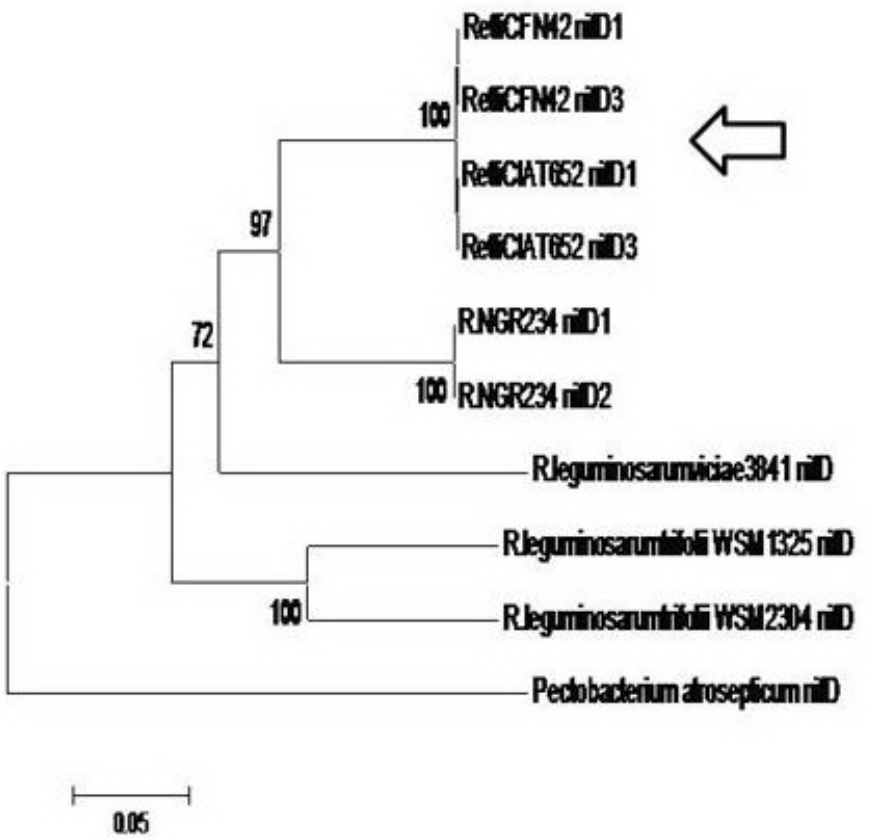

Figure 2. *nifD* multigene family phylogeny. Arrow indicates the clade of the probable gene conversions. See text for details.

Sequence analysis of gene conversion events

An alignment analysis of the *nifD* sequences reveals that the two paralogous gene copies in the *R. etli* CFN42 genome are completely identical, as well as the gene copies present within *R. etli* CIAT652. Nonetheless, when we compared the *nifD* sequences between strains, two sequence polymorphisms were detected (Figure 4). These nucleotide polymorphisms are indicative of probable gene conversion events. Sequence alignment of the *nifK* gene family in the two *R. etli* strains also reveals a more complex pattern of gene conversions, since we identified 13 polymorphisms between the orthologous copies (Figure 5).

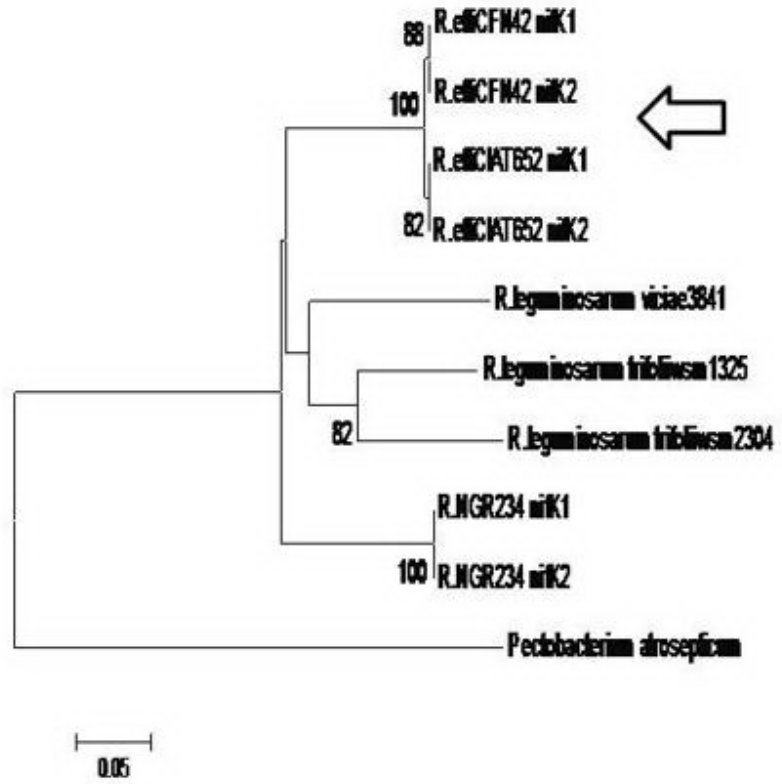

Figure 3. *nifK* multigene family phylogeny. Arrow indicates the clade of the probable gene conversions. See text for details.

Figure 4. Sequence polymorphisms show evidence of gene conversion in the *nifD* multigene family (indicated by rectangles).

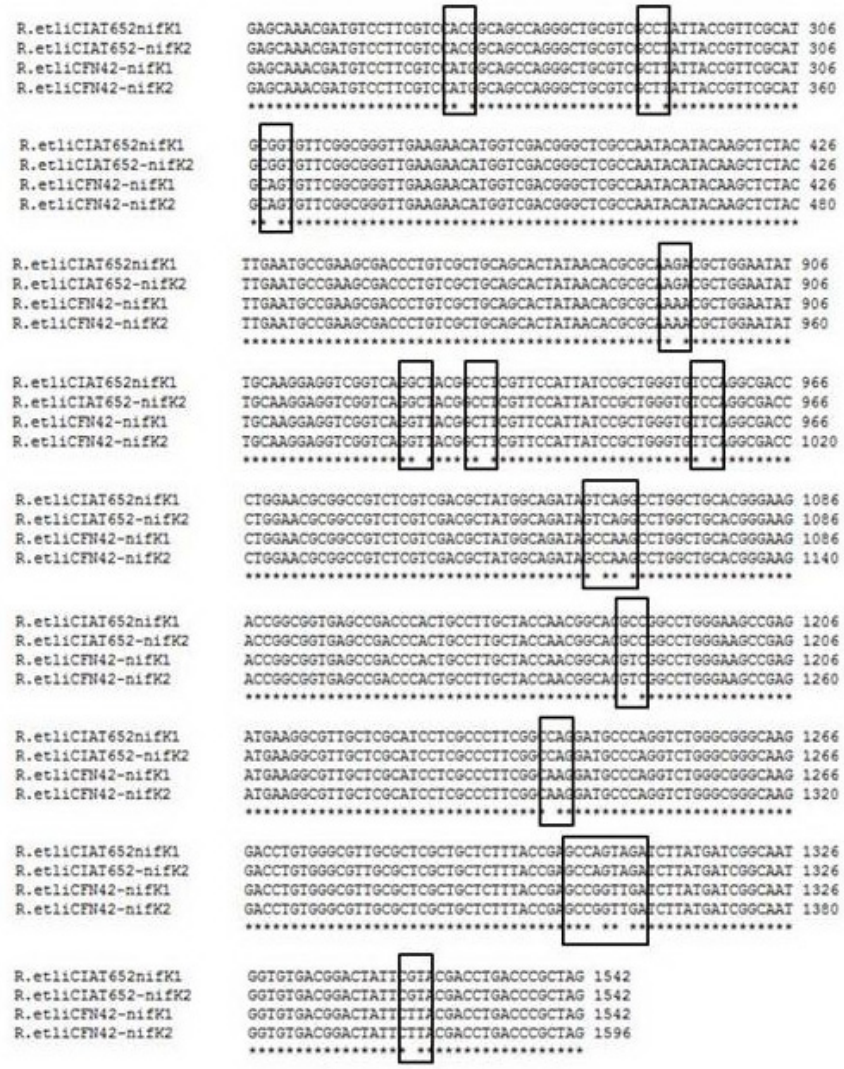

Figure 5. Multiple sequence polymorphisms show evidence of gene conversion in the *nifK* multigene family (indicated by rectangles).

DISCUSSION

An interesting outcome of a recombination event between identical or highly similar gene copies is the non-reciprocal transfer of genetic information, resulting in a homogenizing mechanism known as gene conversion. Gene conversion has been studied in diverse bacteria, where it plays an important role in the concerted evolution of multigene families (Santoyo and Romero, 2005).

This study provided phylogenetic evidence of gene conversions in two multigene

families of rhizobia, which was also confirmed by sequence comparative analysis.

It has been proposed that transformation could be a homogenization mechanism in Hop genes in *H. pylori* (Alm et al., 1999). However, there is no evidence that *R. etli* is transformable. In fact, laboratory experiments have shown an extremely low efficiency in transforming it by diverse protocols and methods (Romero D, personal communication). Therefore, gene conversion appears to be the most probable homogenizing mechanism. Also, a study by Santoyo et al. (2005) showed experimentally that gene conversion can homogenize the *nifH* multigene family in *R. etli*. They also determined that the length of gene conversion events between the *nifH* copies is approximately 100 to 800 bp. For the *nifD* family, the two polymorphisms have a distance of 170 bp. However, for the *nifK* family, there are multiple gene conversions, with the longest distance between the first one and last one of 1254 bp. In this way, apparent gene conversions between the *nifD* and *nifK* paralogous gene copies are within this probable size range.

It is interesting to note that in bacterial pathogens, gene conversion could provide a selective advantage by homogenizing housekeeping genes or those implicated in survival in adverse environments (Deitsch et al., 1997; Liao, 2000). However, the biological significance of gene conversions in *nif* multigene families (other than evolving in a concerted way) is still unknown. These *nif* genes play no essential role in viability or survival, since *Rhizobium* species can live saprophytically. Nonetheless, *nif* genes are important during nodulation processes because they are expressed to fix nitrogen and promote plant health. The complete genome sequences of the other six *R. etli* strains from different geographical origins are currently being determined (González et al., 2010). Genomic comparison between the multigene families or long DNA repeats would reveal whether gene conversion is involved in not only homogenizing non-housekeeping functions but also other functions.

ACKNOWLEDGMENTS

Research supported by Coordinacion de la Investigacion Cientifica-UMSNH (Programa de Investigacion 2009/2010). J.E. Hernandez-Salmeron received a Master's Scholarship from CONACYT-Mexico (#232463).

REFERENCES

Abdulkarim F and Hughes D (1996). Homologous recombination between the *tuf* genes of *Salmonella typhimurium*. *J. Mol. Biol.* 260: 506-522.

Achaz G, Rocha EP, Netter P and Coissac E (2002). Origin and fate of repeats in bacteria. *Nucleic Acids Res.* 30: 2987-2994.

Alm RA, Ling LS, Moir DT, King BL, et al. (1999). Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* 397: 176-180.

Deitsch KW, Moxon ER and Wellems TE (1997). Shared themes of antigenic variation and virulence in bacterial, protozoal, and fungal infections. *Microbiol. Mol. Biol. Rev.* 61: 281-293.

Galibert F, Finan TM, Long SR, Puhler A, et al. (2001). The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science* 293: 668-672.

Gonzalez V, Santamaria RI, Bustos P, Hernandez-Gonzalez I, et al. (2006). The partitioned *Rhizobium etli* genome: genetic and metabolic redundancy in seven interacting replicons. *Proc. Natl. Acad. Sci. U. S. A* 103: 3834-3839.

Gonzalez V, Acosta JL, Santamaria RI, Bustos P, et al. (2010). Conserved symbiotic plasmid DNA sequences in the multireplicon pangenomic structure of *Rhizobium etli*. *Appl. Environ. Microbiol.* 76: 1604-1614.

Haas R and Meyer TF (1986). The repertoire of silent pilus genes in *Neisseria gonorrhoeae*: evidence for gene conversion.

Cell 44: 107-115.
Jordan IK, Makarova KS, Wolf YI and Koonin EV (2001). Gene conversions in genes encoding outer-membrane proteins in *H. pylori* and *C. pneumoniae*. *Trends Genet.* 17: 7-10.
Kaneko T, Nakamura Y, Sato S, Asamizu E, et al. (2000). Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. *DNA Res.* 7: 331-338.
Kaneko T, Nakamura Y, Sato S, Minamisawa K, et al. (2002). Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA Res.* 9: 189-197.
Kimura M (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16: 111-120.
Kline KA, Sechman EV, Skaar EP and Seifert HS (2003). Recombination, repair and replication in the pathogenic *Neisseriae*: the 3 R's of molecular genetics of two human-specific bacterial pathogens. *Mol. Microbiol.* 50: 3-13.
Lathe WC III and Bork P (2001). Evolution of *tuf* genes: ancient duplication, differential loss and gene conversion. *FEBS Lett.* 502: 113-116.
Liao D (2000). Gene conversion drives within genic sequences: concerted evolution of ribosomal RNA genes in bacteria and archaea. *J. Mol. Evol.* 51: 305-317.
Orozco-Mosqueda MC, Altamirano-Hernandez J, Farias-Rodriguez R, Valencia-Cantero E, et al. (2009). Homologous recombination and dynamics of rhizobial genomes. *Res. Microbiol.* 160: 733-741.
Rodriguez C and Romero D (1998). Multiple recombination events maintain sequence identity among members of the nitrogenase multigene family in *Rhizobium etli*. *Genetics* 149: 785-794.
Santoyo G and Romero D (2005). Gene conversion and concerted evolution in bacterial genomes. *FEMS Microbiol. Rev.* 29: 169-183.
Santoyo G, Martinez-Salazar JM, Rodriguez C and Romero D (2005). Gene conversion tracts associated with crossovers in *Rhizobium etli*. *J. Bacteriol.* 187: 4116-4126.
Schmeisser C, Liesegang H, Krysciak D, Bakkou N, et al. (2009). *Rhizobium* sp. strain NGR234 possesses a remarkable number of secretion systems. *Appl. Environ. Microbiol.* 75: 4035-4045.
Tamura K, Dudley J, Nei M and Kumar S (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24: 1596-1599.
Treangen TJ, Abraham AL, Touchon M and Rocha EP (2009). Genesis, effects and fates of repeats in prokaryotic genomes. *FEMS Microbiol. Rev.* 33: 539-571.