Journal of Bacteriology, Dec. 1997, p. 7488–7496 Vol. 179, No. 23
0021-9193/97/$04.00+0

Structure and Evolution of NGRRS-1, a Complex, Repeated Element in the Genome of *Rhizobium* sp. Strain NGR234

X. PERRET,[1] V. VIPREY,[1] C. FREIBERG,[2] and W. J. BROUGHTON[1]*

Laboratoire de Biologie Moleculaire de Plantes Supérieures, University of Geneva, 1292 Chambésy, Geneva, Switzerland,[1] and Abteilung Genomanalysis, Institut für Molekulare Biotechnologie, 07745 Jena, Germany[2]

Received 11 June 1997/Accepted 16 September 1997

Much of the remarkable ability of *Rhizobium* sp. strain NGR234 to nodulate at least 110 genera of legumes, as well as the nonlegume *Parasponia andersonii*, stems from the more than 80 different Nod factors it secretes. Except for *nodE*, *nodG*, and *nodPQ*, which are on the chromosome, most Nod factor biosynthesis genes are dispersed over the 536,165-bp symbiotic plasmid, pNGR234*a*. Mosaic sequences and insertion sequences (ISs) comprise 18% of pNGR234*a*. Many of them are clustered, and these IS islands divide the replicon into large blocks of functionally related genes. At 6 kb, NGRRS-1 is a striking example: there is one copy on pNGR234*a* and three others on the chromosome. DNA sequence comparisons of two NGRRS-1 elements identified three types of IS, NGRIS-2, NGRIS-4, and NGRIS-10. Here we show that all four copies of NGRRS-1 probably originated from transposition of NGRIS-4 into a more ancient IS-like sequence, NGRIS-10. Remarkably, all nine copies of NGRIS-4 have transposed into other ISs. It is unclear whether the accumulation of potentially mutagenic sequences in large clusters is due to the nature of the IS involved or to some selection process. Nevertheless, a direct consequence of the preferential targeting of transposons into such IS islands is to minimize the likelihood of disrupting vital functions.

It has long been recognized that repetitive DNA is an important feature of eukaryotic genomes (8). Bacterial genomes, in contrast, are often regarded as being less complex, containing mostly single-copy genetic elements. Although this might be true for small genomes, such as that of *Methanococcus jannaschii* (11), many bacterial loci are in fact duplicated. Rhizobial genomes are also composed of several replicons. Large plasmids make up half of the genome in some species (47) and harbor many reiterated DNA sequences (32). Repeated sequences vary in size from 100 bases for *Rhizobium*-specific intergenic mosaic elements (36) to several kilobases in the duplicated *nifHDK* genes of *Rhizobium* sp. strain NGR234 which code for the enzyme nitrogenase (3). Transposable elements (transposons and insertion-like sequences) are also broadly distributed in bacteria, where they have many effects, including gene rearrangements, insertions, and deletions (18). In addition to causing deletions and rearrangements of genes, transposition also causes insertional mutations which may disrupt genes or modify their expression.

A number of insertion sequence (IS) elements have been identified in rhizobia; some of them are linked to spontaneous symbiotic mutants (15, 35). In *R. meliloti*, IS*Rm1* is present in 1 to 11 copies (53) and, like IS*Rm2*, is preferentially found in association with symbiotic genes (15, 44). Other elements such as IS*Rm3*, IS*Rm4*, and IS*Rm5* were also characterized (30, 48, 54), suggesting that most *R. meliloti* strains carry over 50 IS copies per genome (30). In *R. leguminosarum* bv. viciae strains, IS*RI2* is present in low copy numbers and seems preferentially linked to plasmids (33). Among the family of repeated elements RSα, RSβ, RSγ, RSδ, and RSε, which are closely linked to the nitrogen fixation genes of *Bradyrhizobium japonicum* USDA110, only RSα (present in 12 copies) has properties similar to those of an IS element (25). Similarly, at least one copy of the hyperreiterated DNA region HRS1 of *B. japonicum* USDA424 is adjacent to the *fixRnifA* locus of this strain (24).

Various reiterated sequences have been found in NGR234 (3, 31, 36, 37, 52). Among them, NGRRS-1 was identified during the construction of the ordered cosmid library of pNGR234*a* (37). Another three copies of NGRRS-1, extending over more than 5 kb, were found on the chromosome (38). In this report, we show that all copies of NGRRS-1 probably originated from transposition of one insertion element (NGRIS-4) into a more ancient, IS-like sequence, NGRIS-10. Comparison of the complete DNA sequence of two NGRRS-1 elements identified another type of IS, NGRIS-2. In fact, a large fraction of the NGR234 genome is comprised of IS-like elements, most of which are clustered in several islands. This irregular distribution and the dynamics of dissemination, provide new insights into the stability, plasticity, and evolution of NGR234 genome.

MATERIALS AND METHODS

Microbiological techniques. Bacterial strains and plasmids used in this study are listed in Table 1. *Escherichia coli* was grown on Luria-Bertani medium or Terrific broth (45). Strains of *Bradyrhizobium*, *Rhizobium*, and *Agrobacterium* were grown in tryptone-yeast medium (5) or *Rhizobium* minimal medium (10). pBluescript KS+ recombinants were raised in *E. coli* DH5α, while Lorist2 cosmid clones were grown in *E. coli* 1046. Antibiotics were used at the following concentrations: chloramphenicol, 25 μg/ml; rifampin, kanamycin, and spectinomycin, 50 μg/ml; and streptomycin and ampicillin, 100 μg/ml.

DNA isolation and sequencing. DNA of cosmids and pBluescript KS+ clones was prepared by standard alkaline procedures followed by purification on CsCl gradients (45). The complete DNA sequence of cosmid pXB807 was established by using standard protocols for sequencing GC-rich cosmid clones (16). DNA sequences of the NGRRS-1*b*, -1*c*, and -1*d* elements were produced by manual, dideoxy methods (46) using double-stranded templates and Sequenase II (United States Biochemical Corp., Cleveland, Ohio). Selected subclones (Table 1; Fig. 1), *Exo*III nuclease deletions and primer walking (45) were used to determine the DNA sequence of the 6.8-kb *Eco*RI-*Pst*I fragment encompassing the NGRRS-1*b* locus of pXB826.

DNA labeling and hybridization procedures. ^{32}P labeling of the *Sau*3AI fragments of NGR234 remaining after subtraction against genomic DNA of USDA257 was performed by three cycles of PCR amplification (6). Inserts from selected pBluescript KS+ clones were radioactively labeled by PCR amplifica-

* Corresponding author. Mailing address: Laboratoire de Biologie Moléculaire des Plantes Supérieures, University of Geneva, 1 chemin de l'Impératrice, 1292 Chambésy, Geneva, Switzerland.

TABLE 1. Bacterial strains and plasmids used in this study

Strain, plasmid, or vectors	Relevant characteristics	Reference or source
Strains		
E. coli DH5α	*recA1* *ϕ80 lacZ*ΔM15	20
E. coli 1046	*recA1*	12
Rhizobium sp. strain NGR234	Broad host range, isolated from *Lablab purpureus*, Rif^r	49
Rhizobium sp. strain ANU265	Sym plasmid-cured derivative of NGR234, Spc^r	34
R. fredii USDA257	Broad host range, isolated from *Glycine soja*, Km^r	22
R. meliloti 2011	Wild-type isolate from *Medicago sativa*	43
R. loti NZP4010	Cryptic plasmid-cured derivative of NZP2037, Rif^r Sm^r	13
Rhizobium sp. strain WBM16	Isolated from *Leucaena hut*	9
B. elkanii USDA76	Isolated from *Glycine max*, Cm^r	29
B. japonicum CB756	Isolated from *Macrotyloma africanum*	21
A. rhizogenes R1600	Ap^r Km^r	
A. rhizogenes A4RSII	pRiA4, Rif^r	23
Cosmids		
Lorist2	5.6-kb cosmid vector, Km^r	19
pXB807	Lorist2 clone of pNGR234*a*	37
pXB826, pXB953, pXB1539	Lorist2 clones of the NGR234 chromosome	38
Bluescript clones		
pBluescript KS+	High-copy-number, ColE1 based phagemid, Ap^r	Stratagene, La Jolla, Calif.
pXB807H-0.8	0.8-kb *Hin*dIII fragment of pXB807, covers the right border of NGRRS-1*a*	This work
pXB807P-0.6	0.6-kb *Pst*I fragment of pXB807, covers the left border of NGRRS-1*a*	This work
pXB826P-1	1-kb *Pst*I fragment of pXB826, covers the right border of NGRRS-1*b*	This work
pXB826PH-1	1-kb *Pst*I-*Hin*dIII fragment of pXB826	This work
pXB826H-0.5	0.5-kb *Hin*dIII fragment of pXB826	This work
pXB826X-2	2-kb *Xho*I fragment of pXB826	This work
pXB826P-0.8	0.8-kb *Pst*I fragment of pXB826	This work
pXB826P-3	3-kb *Pst*I fragment of pXB826	This work
pXB826XP-0.9	0.9-kb *Xho*I-*Pst*I fragment of pXB826	This work
pXB826CP-0.6	0.6-kb *Cla*I-*Pst*I fragment of pXB826	This work
pXB826C-1.8	1.8-kb *Cla*I fragment of pXB826, internal to NGRIS-4*c*	This work
pXB826PX-2	2-kb *Pst*I-*Xho*I fragment of pXB826	This work
pXB826C-0.6	0.6-kb *Cla*I fragment of pXB826P-3	This work
pXB826EP-0.6	0.6-kb *Eco*RI-*Pst*I fragment of pXB826, covers the left border of NGRRS-1*b*	This work
pXB953H-4	4-kb *Hin*dIII fragment of pXB953, covers the right border of NGRRS-1*c*	This work
pXB953H-0.5	0.5-kb *Hin*dIII fragment of pXB953	This work
pXB953XP-0.5	0.5-kb *Xho*I-*Pst*I fragment of pXB953, covers the left border of NGRRS-1*c*	This work
pXB953X-2.5	2.5-kb *Xho*I fragment of pXB953	This work
pXB1539H-1.4	1.4-kb *Hin*dIII fragment of pXB1539, covers the right border of NGRRS-1*d*	This work
pXB1539H-0.5	0.5-kb *Hin*dIII fragment of pXB1539	This work
pXB1539P-0.8	0.8-kb *Pst*I fragment of pXB1539, covers the left border of NGRRS-1*d*	This work

tion using T3-T7 primers flanking the entire insert. Southern blots of DNA samples restricted by endonucleases and multiple samples of nondigested DNA (dot blots) were probed by using standard hybridization conditions (45).

RESULTS

Characterization of the four NGRRS-1 loci. The canonical ordered cosmid library of NGR234 was constructed by comparing the *Hin*dIII fingerprints of 1,014 cosmids, 227 of which hybridized to pNGR234*a* (37, 38). Interestingly, a significant portion of these clones did not map to pNGR234*a*. Rather, they formed independent contigs, suggesting that a number of pSym sequences are reiterated. Among these, NGRRS-1 is repeated four times. At ≅6 kb, NGRRS-1, which has conserved *Hin*dIII, *Pst*I, and *Xho*I restriction sites (Fig. 1A), is the largest repeat known in NGR234. To facilitate its analysis, four cosmids, pXB807, pXB826, pXB953, and pXB1539, were selected to represent, respectively, NGRRS-1*a* of pSym as well as the three, presumably chromosomally borne copies NGRRS-1*b*, NGRRS-1*c*, and NGRRS-1*d*. Cloning and sequencing of several restriction fragments covering the ends of the four repeats (Fig. 1A), helped identify their border regions. Although all four of the 3′-end borders are almost identical (97% at the nucleotide level) and start to diverge at the same nucleotide (Fig. 2), the 5′ ends of NGRRS-1*a* and NGRRS-1*c* are truncated compared with those of NGRRS-1*b* and NGRRS-1*d* (Fig. 2). To fully characterize one of the repeats, a series of overlapping subclones of pXB826 was used to establish the DNA sequence of the 6.8-kb *Eco*RI-*Pst*I region which carries NGRRS-1*b* (Table 1; Fig. 1B).

Comparison of NGRRS-1*a* and -1*b*. Concomitantly, the sequence of NGRRS-1*a* was obtained during the automated shotgun sequencing of pNGR234*a* (17). Multiple alignments of the NGRRS-1*a* and -1*b* sequences, together with those of the border regions of the repeats carried by pXB953 and pXB1539,

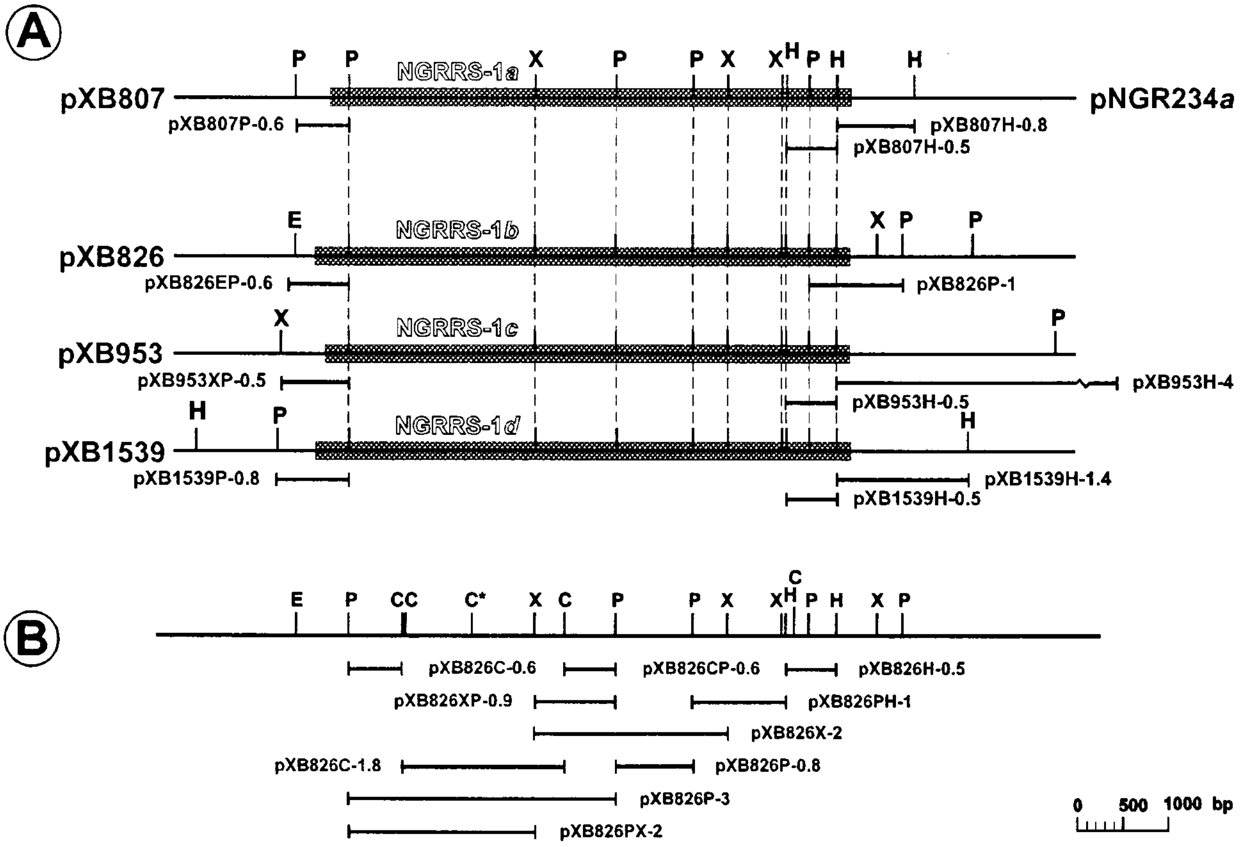
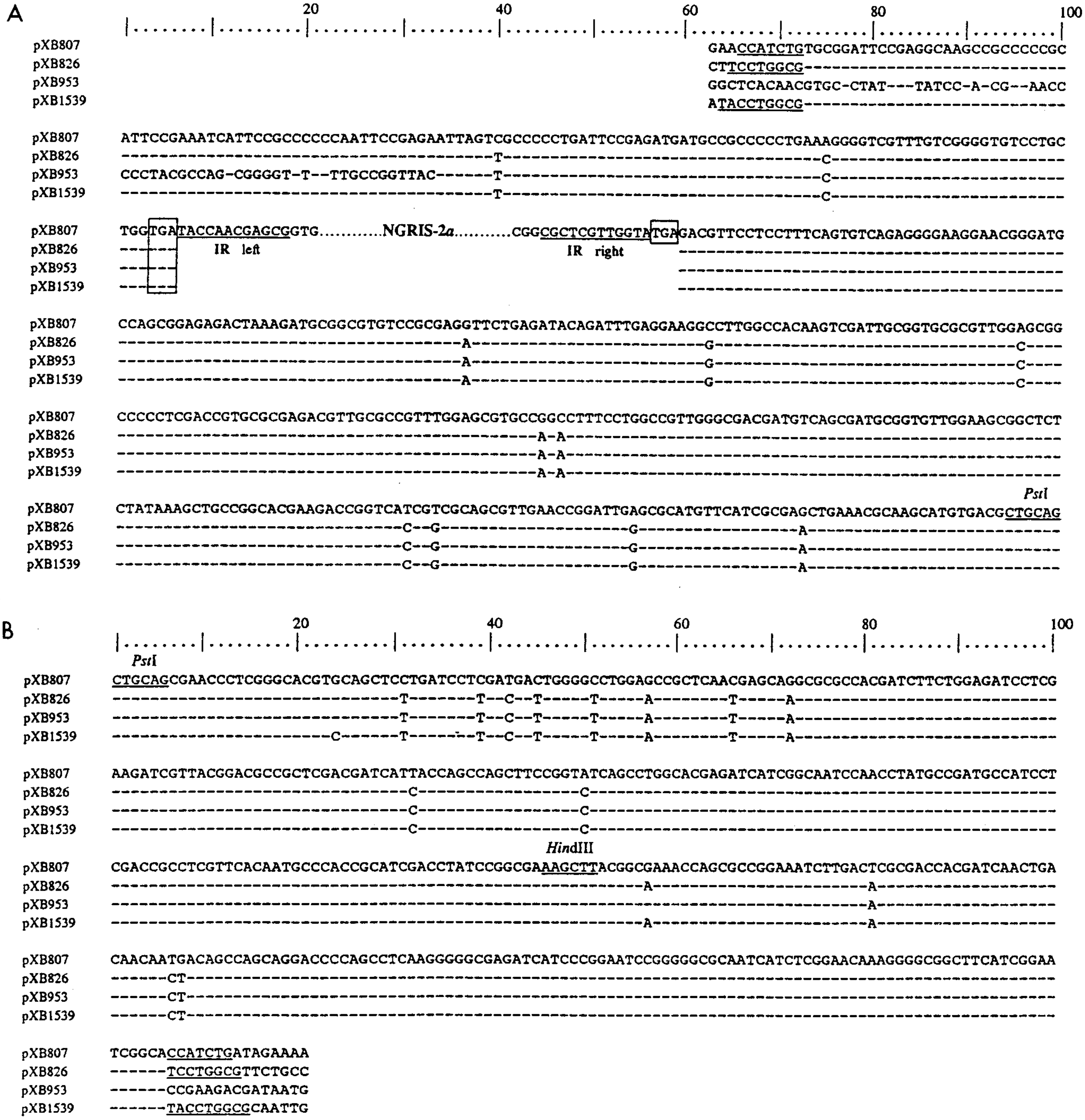


FIG. 1. (A) Restriction map of the four copies of NGRRS-1 carried by cosmids pXB807 (from pNGR234*a*), pXB826, pXB953, and pXB1539. Subclones used to sequence the borders of NGRRS-1 (Fig. 2) are displayed below the partial restriction map of each cosmid. Each repeat is represented by a shaded box, with dashed lines linking the conserved restriction sites. (B) Subclones used to determine the DNA sequence of NGRRS-1*b* are placed below the partial restriction map of cosmid pXB826. Restriction sites: C, *Cla*I; C*, *Cla*I site methylated in *E. coli* DH5α; E, *Eco*RI; H, *Hin*dIII; P, *Pst*I; X, *Xho*I.

showed that NGRRS-1*b* covers 5,919 nucleotides. There are two major differences between NGRRS-1*a* and NGRRS-1*b*. NGRRS-1*a* contains a 2,561-bp-long insertion into the left border region, and 116 irregularly distributed point mutations differ along the two repeated elements (Fig. 3).

NGRIS-2 has the characteristics of an insertion element. The 2,561-bp insertion into the 5′-end border region of NGRRS-1*a* (Fig. 2 and 3) presents many of the properties of IS elements. It is delimited by two perfectly conserved 12-bp inverted repeats directly flanked by short 3-bp direct repeats of target duplication (Table 2). Genemark analysis (7) of NGRIS-2 predicted the presence of three putative genes transcribed in the same orientation and encoding proteins of 21 (Orf1-IS*2*), 43 (Orf2-IS*2*), and 15 (Orf3-IS*2*) kDa, respectively. Comparison of these hypothetical products with published sequences by using the BLAST program (2) showed significant homologies to putative transposases of various microorganisms (Table 3). Interestingly, a 711-bp fragment of RFRS9, a member of the *Rhizobium fredii* family of repetitive sequences (28), showed extensive homology to two distinct segments of NGRIS-2 separated by almost 1,400 nucleotides (Fig. 3). Thus, DNA sequences in NGRIS-2 from positions 1 to 425 and positions 1,824 to 2,104 present 91% identity with two contiguous blocks of sequences in RFRS9 (positions 710 to 287 and positions 286 to 1, respectively) (Fig. 3). Moreover, the amino-terminal part of the hypothetical protein encoded by the single open reading frame (ORF) described for RFRS9 presents 91% identity and 97% similarity with the putative *orf3*-IS*2* product when conserved and less conserved amino acids are included in the analysis (Table 3). Although functional transposition was not demonstrated, homology searches within the 536-kb sequence of pNGR234*a* (17) revealed the presence of a second and completely identical copy of NGRIS-2. This element, referred to as NGRIS-2*b* and carried by cosmid pXB43, is 17 kb upstream of NGRIS-2*a* (in the sense of transcription of the three putative ORFs).

Irregular distribution of polymorphisms along NGRRS-1*a* and -1*b*. Among the 116 nucleotide differences identified by aligning the NGRRS-1*a* and -1*b* sequences, 115 are clustered into two groups of 25 and 90 polymorphic sites. The first 25 are within the 940 bp which include the 5′-border region of NGRRS-1, while the remaining 90 differences are in the last 1,614 bp of the repeated element (Fig. 3). Surprisingly, only a single polymorphic position was found in the 3,324-bp central portion of NGRRS-1, which was also shown to be perfectly duplicated in another part of the symbiotic plasmid. Together these data suggest the presence within NGRRS-1 of another autonomously mobile element, NGRIS-4. Although there are no inverted repeats, NGRIS-4 is flanked by four-nucleotide direct repeats of target duplication (Table 2). Unlike those of the two NGRIS-2 copies, which are similar but not identical, the target duplication sites of the three NGRIS-4 elements are identical (5′_AAGG_3′). Genemark analysis predicted four putative genes transcribed in the same direction as in NGRIS-2*a* and encoding products of 6,105 (Orf1-IS*4*), 16,836 (Orf2-IS*4*), 9,703 (Orf3-IS*4*) and 78,742 (Orf4-IS*4*) Da (Fig. 3;



```
A                             20                  40                  60                  80                 100
          |.........|.........|.........|.........|.........|.........|.........|.........|.........|.........|
pXB807                                                                  GAACCATCTGTGCGGATTCCGAGGCAAGCCGCCCCCGC
pXB826                                                                  CTTCCTGGCG----------------------------
pXB953                                                                  GGCTCACAACGTGC-CTAT---TATCC-A-CG--AACC
pXB1539                                                                 ATACCTGGCG----------------------------

pXB807    ATTCCGAAATCATTCCGCCCCCCAATTCCGAGAATTAGTCGCCCCCTGATTCCGAGATGATGCCGCCCCCTGAAAGGGGTCGTTTGTCGGGGTGTCCTGC
pXB826    ---------------------------------------T----------------------------------C-------------------------
pXB953    CCCTACGCCAG-CGGGGT-T--TTGCCGGTTAC------T----------------------------------C-------------------------
pXB1539   ---------------------------------------T----------------------------------C-------------------------

pXB807    TGGTGATACCAACGAGCGGTG...........NGRIS-2a..........CGGCGCTCGTTGGTATGAGACGTTCCTCCTTTCAGTGTCAGAGGGGAAGGAACGGGATG
pXB826    ------      IR left                                  IR right     -----------------------------------------
pXB953    ------                                                            -----------------------------------------
pXB1539   ------                                                            -----------------------------------------

pXB807    CCAGCGGAGAGACTAAAGATGCGGCGTGTCCGCGAGGTTCTGAGATACAGATTTGAGGAAGGCCTTGGCCACAAGTCGATTGCGGTGCGCGTTGGAGCGG
pXB826    ------------------------------------A-------------------------G--------------------------------C----
pXB953    ------------------------------------A-------------------------G--------------------------------C----
pXB1539   ------------------------------------A-------------------------G--------------------------------C----

pXB807    CCCCCTCGACCGTGCGCGAGACGTTGCGCCGTTTGGAGCGTGCCGGCCTTTCCTGGCCGTTGGGCGACGATGTCAGCGATGCGGTGTTGGAAGCGGCTCT
pXB826    --------------------------------------------A-A-----------------------------------------------------
pXB953    --------------------------------------------A-A-----------------------------------------------------
pXB1539   --------------------------------------------A-A-----------------------------------------------------
                                                                                                        PstI
pXB807    CTATAAAGCTGCCGGCACGAAGACCGGTCATCGTCGCAGCGTTGAACCGGATTGAGCGCATGTTCATCGCGAGCTGAAACGCAAGCATGTGACGCTGCAG
pXB826    ------------------------------C--G--------------------G-----------------A---------------------------
pXB953    ------------------------------C--G--------------------G-----------------A---------------------------
pXB1539   ------------------------------C--G--------------------G-----------------A---------------------------

B                             20                  40                  60                  80                 100
          |.........|.........|.........|.........|.........|.........|.........|.........|.........|.........|
           PstI
pXB807    CTGCAGCGAACCCTCGGGCACGTGCAGCTCCTGATCCTCGATGACTGGGGCCTGGAGCCGCTCAACGAGCAGGCGCGCCACGATCTTCTGGAGATCCTCG
pXB826    ------------------------------T-------T--C--T-----T-----A--------T-----A----------------------------
pXB953    ------------------------------T-------T--C--T-----T-----A--------T-----A----------------------------
pXB1539   -----------------------C------T-------T--C--T-----T-----A--------T-----A----------------------------

pXB807    AAGATCGTTACGGACGCCGCTCGACGATCATTACCAGCCAGCTTCCGGTATCAGCCTGGCACGAGATCATCGGCAATCCAACCTATGCCGATGCCATCCT
pXB826    -------------------------------C-----------------C--------------------------------------------------
pXB953    -------------------------------C-----------------C--------------------------------------------------
pXB1539   -------------------------------C-----------------C--------------------------------------------------
                                                        HindIII
pXB807    CGACCGCCTCGTTCACAATGCCCACCGCATCGACCTATCCGGCGAAAGCTTACGGCGAAACCAGCGCCGGAAATCTTGACTCGCGACCACGATCAACTGA
pXB826    --------------------------------------------------------A-----------------------A-------------------
pXB953    --------------------------------------------------------------------------------A-------------------
pXB1539   --------------------------------------------------------A-----------------------A-------------------

pXB807    CAACAATGACAGCCAGCAGGACCCCAGCCTCAAGGGGGCGAGATCATCCCGGAATCCGGGGGCGCAATCATCTCGGAACAAAGGGGCGGCTTCATCGGAA
pXB826    ------CT--------------------------------------------------------------------------------------------
pXB953    ------CT--------------------------------------------------------------------------------------------
pXB1539   ------CT--------------------------------------------------------------------------------------------

pXB807    TCGGCACCATCTGATAGAAAA
pXB826    ------TCCTGGCGTTCTGCC
pXB953    ------CCGAAGACGATAATG
pXB1539   ------TACCTGGCGCAATTG
```

FIG. 2. Multiple alignments of the 5′ (A) and 3′ (B) ends of DNA sequences bordering the four NGRRS-1*a*, -1*b*, -1*c*, and -1*d* copies carried by cosmids pXB807 (part of pNGR234*a*), pXB826, pXB953, and pXB1539, respectively. Identical nucleotide bases are replaced by hyphens. Landmark *Pst*I and *Hin*dIII restriction sites are labeled and underlined. The position of insertion of the NGRIS-2*a* element into the sequence of NGRRS-1*a* is shown: the 12-bp perfect inverted repeats bordering NGRIS-2*a* are underlined and marked IR left and IR right. The trinucleotide TGA direct duplication of target DNA is boxed. Three different target insertion sites of NGRRS-1 into the genomic regions covered by cosmids pXB807, pXB826, and pXB1539 are underlined.

Table 3). Although the Orf2-IS*4* and Orf4-IS*4* hypothetical proteins show weak homologies to protein 1 of IS*895* of the cyanobacterium *Anabaena* sp. strain PCC 7120 (1) and to site-specific recombinases such as TnpX and Xisf (Table 3), respectively, no clear homolog was found in the databases.

NGRIS-10, the remains of another IS-like element disrupted by NGRIS-4. Clustering of polymorphic nucleotides at both ends of NGRRS-1*a* and -1*b*, together with the IS-like features of NGRIS-4, suggested that all four copies of NGRRS-1 resulted from transposition of NGRIS-4 into another multicopy element. To reconstruct this repeat, the NGRIS-4*c* DNA sequence (3,316 bp) together with its 4-bp target duplication site was removed in silico from NGRRS-1*b*. This reconstructed 2,599-bp element was called NGRIS-10. It carries two putative genes encoding hypothetical proteins of ca. 59 (*orf1*-IS*10*) and 28 (*orf2*-IS*10*) kDa and is bordered by

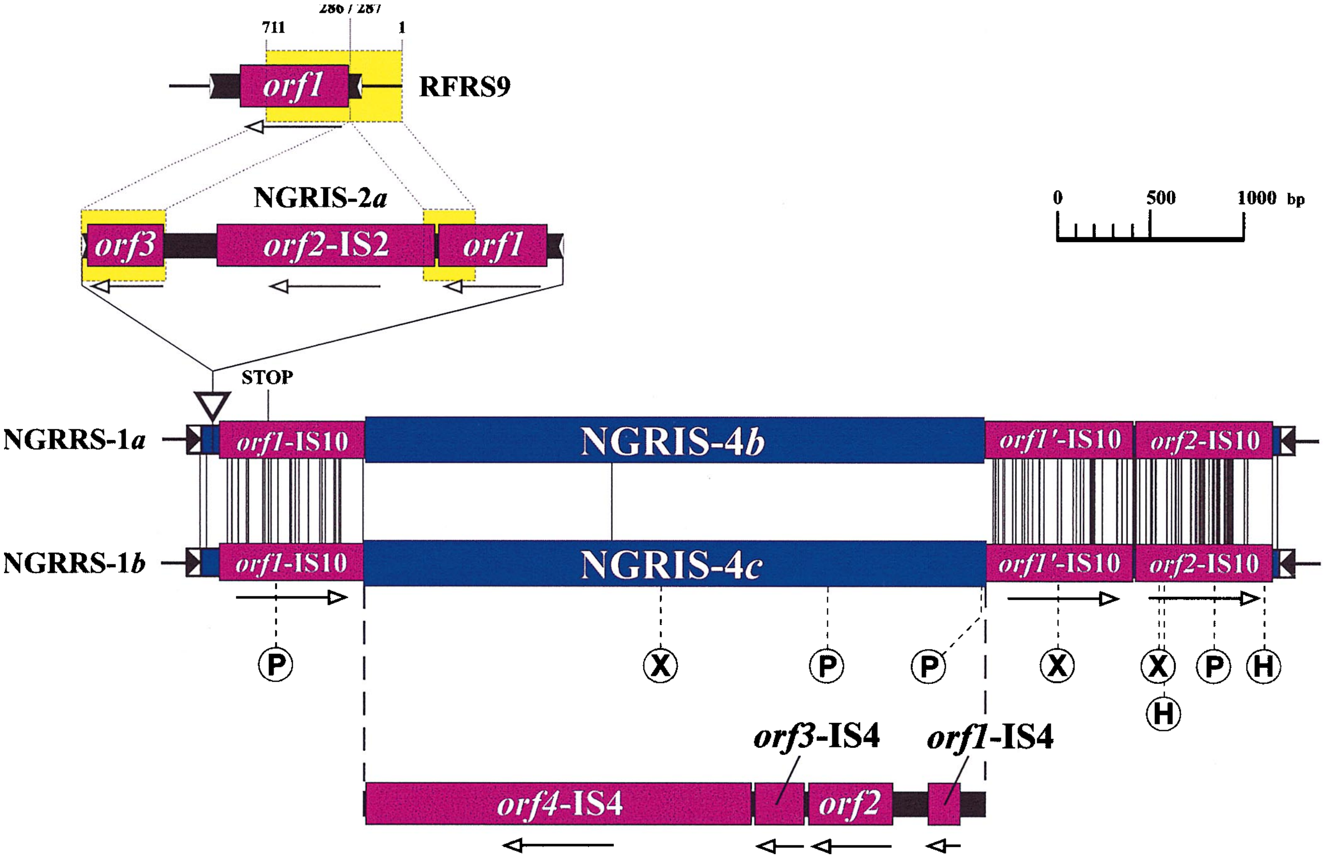


FIG. 3. Scheme presenting the various IS-like elements which compose the NGRRS-1*a* and NGRRS-1*b* repeats. Respective positions, lengths, and senses of transcription (marked by arrows) of the putative coding ORFs identified within NGRIS-2, NGRIS-4, and the reconstructed NGRIS-10 are marked by magenta boxes. Vertical lines between the NGRRS-1*a* and -1*b* repeats correspond to the positions of polymorphic nucleotides in the two sequences. The premature stop codon in *orf1*-IS*10*, which results from a single nucleotide substitution in the NGRRS-1*a* sequence (TGG into TGA), is also reported (stop). Small inverted triangles represent terminal inverted repeats identified in NGRIS-2 and NGRIS-10. The site of insertion of NGRIS-2 into NGRRS-1*a* is marked with a large triangle. Conserved restriction sites presented in Fig. 1A are reported with dashed lines and circled H (*Hin*dIII), P (*Pst*I), and X (*Xho*I). Respective positions of the two blocks of sequence homology between NGRIS-2 and RFRS9 are shown as yellow boxes linked with thin dashed lines.

75- and 74-bp imperfect inverted repeats. Interestingly, the target duplication sites of NGRIS-10 identified in pXB807, pXB826, and pXB1539 are different in size and sequence (Table 2). Both hypothetical products encoded by *orf1*-IS*10* and *orf2*-IS*10* show strong homology to various putative transposases such as that of IS*1162* of *Pseudomonas fluorescens* (Table 3). Although no whole NGRIS-10 element was found, hybridizations and BLAST searches for similar proteins encoded by pNGR234*a* identified several homologs. Among these, *y4UI* and *y4UH* (carried by cosmid pXB296 [16]) en-

TABLE 2. Main characteristics of the IS-like elements identified in NGRRS-1 and in pNGR234*a*

Element	Size (bp)	Target duplication	Inverted repeats	Sym plasmid linked	Copy of NGRRS-1:	Cosmid
NGRIS-2*a*	2,558	5′_TGA_3′	5′_TACCAACGAGCG_3′[a]	Yes	NGRRS-1*a*	pXB807
NGRIS-2*b*	2,558	5′_TTA_3′	5′_TACCAACGAGCG_3′[a]	Yes	None	pXB43
NGRIS-4*a*	3,316	5′_AAGG_3′	None found	Yes	None	pXB1459
NGRIS-4*b*	3,316	5′_AAGG_3′	None found	Yes	NGRRS-1*a*	pXB807
NGRIS-4*c*	3,316	5′_AAGG_3′	None found	No	NGRRS-1*b*	pXB826
NGRIS-4*d*	—[b]	—	—	No	NGRRS-1*c*	pXB953
NGRIS-4*e*	—	—	—	No	NGRRS-1*d*	pXB1539
NGRIS-10*a*	2,599	5′_CCATCTG_3′	75 bp/74 bp[d]	Yes	NGRRS-1*a*	pXB807
NGRIS-10*b*	2,599	5′_TCCTGGCG_3′	75 bp/74 bp[d]	No	NGRRS-1*b*	pXB826
NGRIS-10*c*	*[c]	*	*	No	NGRRS-1*c*	pXB953
NGRIS-10*d*	*	5′_TACCTGGCG_3′	75 bp/74 bp[d]	No	NGRRS-1*d*	pXB1539

[a] Perfect 12-bp inverted repeat.

[b] —, no complete DNA sequence available, but RFLP analysis suggests that all NGRIS-4 elements listed above have similar sizes and structures and probably the same insertion sites within NGRRS-1.

[c] *, only left and right border regions of NGRIS-10*c* and -10*d* have been sequenced. The left border of NGRIS-10*c* is truncated by 71 bp.

[d] Imperfect inverted repeats of 75 and 74 bases of the left border and right border, respectively.

TABLE 3. Best homologies to the hypothetical proteins encoded by NGRIS-2, NGRIS-4, and NGRIS-10[a]

Protein	Size (aa)	Homolog	Accession no.	Size (aa)	Identity (%)	Similarity (%)
Orf1-IS2	192	*orfA* product of IS*1238* of *Acetobacter xylinum*[b]	U22323	197	46	80
Orf2-IS2	387	TnpA of IS*4321*L of *Enterobacter aerogenes*[c]	U60777	334	40	76
		Putative transposase of IS*1328* of *Yersinia enterocolitica*[c]	Z48244	334	40	74
Orf3-IS2	135	*orfB* product of IS*1238* of *A. xylinum*	U22323	189	67	87
		Single *orf* product of RFRS9 of *R. fredii* USDA257	U18764	222	91	97
Orf1-IS4	56	No homolog found in databases				
Orf2-IS4	149	Partial homology with protein 1 of IS*895* of *Anabaena*	PID: G142027	189		
Orf3-IS4	88	No homolog found in databases				
Orf4-IS4	694	Partial homology with TnpX recombinase of Tn*4451*	PID: n.a.	500		
		Partial homology with Xisf recombinase of *Anabaena*	PID: G14904	613		
Orf1-IS10	516	*y4UI* hypothetical product of pNGR234*a*[c]	PID: G1486430	514	59	87
		Putative transposase of IS408 of *Pseudomonas cepacia*[c]	L09108	518	46	82
		orf1 putative product of NGRIS-3[d]		516	47	75
Orf2-IS10	245	*y4UH* hypothetical product of pNGR234*a*	PID: G1486429	248	60	87
		orf2 putative product of NGRIS-3[b]		258	55	85
		orf2 of IS1162 of *Pseudomonas fluorescens*[b]	PID: E108313	231	42	72

[a] Identity and homology levels shown were calculated on the basis of the entire protein.
[b] For best alignments, produced with the introduction of one gap of a single amino acid.
[c] For best alignments, produced with the introduction of two gaps of a single amino acid.
[d] For best alignments, produced with the introduction of a maximum of five gaps of one or two amino acids.

coded products which are 59% identical and 87% homologous to Orf1-IS*10* and Orf2-IS*10*, respectively. Both putative genes have the same organization as, as well as 60% homology at the nucleotide level to, those of NGRIS-10.

Distribution of NGRIS-2, NGRIS-4, and NGRIS-10 in NGR234. Traditionally, the copy number of repeated elements is estimated by probing Southern blots of restricted genomic DNA. Further restriction fragment length polymorphism (RFLP) and/or sequencing data are generally required, however, to define the extent of homology between the regions. As an alternative, we used a two-step hybridization procedure to identify loci homologous to each of the elements composing NGRRS-1. First, dot blots of DNA prepared from the 309 cosmids that cover more than 97% of the NGR234 genome were hybridized with ^{32}P-labeled probes internal to NGRIS-2, NGRIS-4, or NGRIS-10. For nonoverlapping clones, an initial estimate of the probable number of copies of the labeled sequence was obtained. RFLP data were obtained by probing Southern blots prepared with *Eco*RI-, *Hin*dIII-, *Pst*I-, and *Xho*I-restricted DNAs of the 14 positive cosmids. Hybridization results (summarized in Table 4) confirmed that no more than four copies of the NGRRS-1 element exist in NGR234. Although two independent repeats, NGRRS-3*a* and -3*b* (Table 4), have structures similar to that of NGRRS-1, their RFLP patterns differ significantly. In fact, each of the two NGRRS-3 copies is composed of one NGRIS-4 sequence inserted into a homolog of NGRIS-10 (named NGRIS-11). Thus, nine complete copies of NGRIS-4 were identified. Although sequence data exist for only three of them, identical RFLP patterns suggest that all copies are highly homologous. NGRIS-4*c* differs from NGRIS-4*a* and -4*b* by a single nucleotide. Interestingly, only four copies of NGRIS-10 were identified, and all of them are disrupted by NGRIS-4. This systematic disruption of all NGRIS-10 and IS*11* copies by six of the nine NGRIS-4 elements clearly demonstrates that both elements are preferential targets for insertion by NGRIS-4. Apart from the two identical copies of NGRIS-2 located on pNGR234*a*, no other element of this kind was identified elsewhere in the genome. Rather, four different homologous regions with restriction patterns different from those of NGRIS-2 were found on the chromosome. The close linkage of one of these loci to NGRRS-3*a* suggests that many IS elements have tendencies to assemble into large and complex clusters.

Subtractive DNA hybridization against total genomic DNA of *R. fredii* USDA257 was used to isolate a pool of *Sau*3AI fragments specific to NGR234 (39). Probing of Southern blots prepared with *Eco*RI-, *Pst*I-, and *Xho*I-restricted DNAs of cosmids pXB807, pXB826, pXB953, and pXB1539, using ^{32}P-

TABLE 4. Homologous repeats that are neither in pNGR234*a* nor in the four copies of NGRRS-1[a]

Element	Probe(s)	Hybridizing cosmids	New repeats identified
NGRIS-2	2.4-kb PCR product	pXB225, pXB632	
		pXB74, pXB290	NGRRS-2*a*
		pXBS12, pXA1-H3	NGRRS-2*b*
		pXB70, pXB1302	
NGRIS-4	pXB826P-0.8	pXB225, pXB632	NGRIS-4*g* (NGRRS-3*a*)
	pXB826XP-0.9	pXB198, pXB636	NGRIS-4*f* (NGRRS-3*b*)
		pXB273, pXA4	NGRIS-4*h*
		pXB445, pXB1526	NGRIS-4*i*
NGRIS-10	pXB826H-0.5	pXB225, pXB632	NGRIS-11*a* (NGRRS-3*a*)
	pXB807X-0.5	pXB198, pXB636	NGRIS-11*b* (NGRRS-3*b*)

[a] Overlapping cosmids are listed on the same line. NGRIS-2*a* was PCR amplified by using primers NGRIS2A (5′-GCGCCGTTTCTGACTCTCATGGG-3′) and NGRIS2B (5′-GAGCGGTGATCATGACCGATGCG-3′), as well as pXB807 DNA.

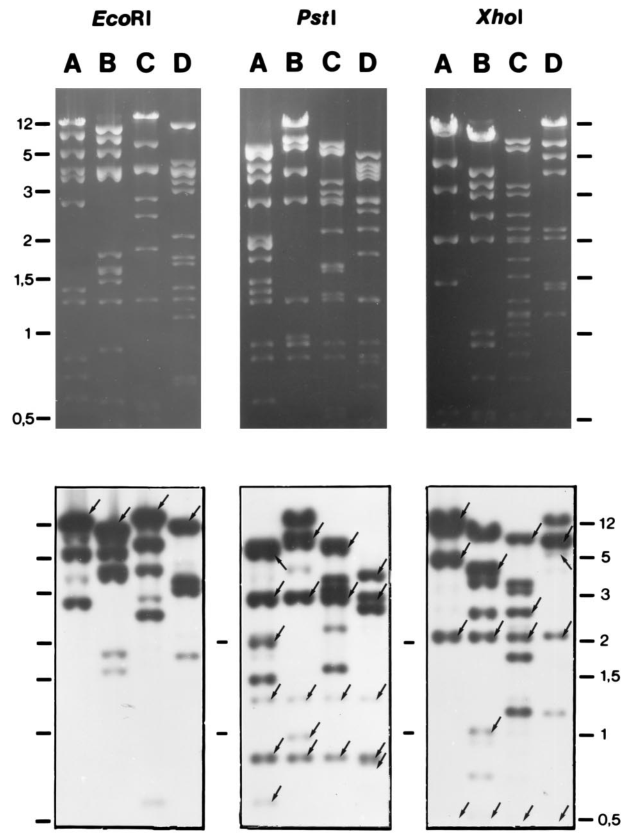


FIG. 4. (Top) *Eco*RI-, *Pst*I-, and *Xho*I-restricted cosmid DNAs of pXB807, pXB826, pXB953, and pXB1539 (lanes A to D, respectively). (Bottom) Corresponding Southern blots probed with ^{32}P-labeled *Sau*3AI fragments of NGR234 purified by subtractive hybridization against total DNA of *R. fredii* USDA257. Restriction fragments that are part of NGRRS-1 are marked with arrows. Molecular markers are given in kilobases.

labeled and PCR-amplified subtracted sequences, confirmed close physical linkage between NGRRS-1 repeats and fragments specific to NGR234 (Fig. 4). Similar clustering of the remaining NGRIS elements was observed when probing was extended to all cosmids listed in Table 4 (data not shown). Moreover, subtracted sequences hybridized to most restriction fragments carrying NGRRS-1, suggesting that NGRIS-10 and NGRIS-4 are not present in USDA257. Confirmation of this result was obtained by probing *Eco*RI-restricted genomic DNA of various *Rhizobium*, *Bradyrhizobium*, and *Agrobacterium* strains with PCR-amplified and ^{32}P-labeled fragments internal to NGRRS-1. Only weak hybridization signals were detected in the genomes of *R. meliloti* 2011 and of *A. rhizogenes* R1600, whereas many fragments ranging from 6 to ca. 20 kb were identified in NGR234 and ANU265 (data not shown). Despite the molecular evidence of close phylogenetic relationships between NGR234 and USDA257 symbiotic plasmids (4, 39–41), it seems that none of the IS elements which form NGRRS-1 are present in the genome of USDA257.

DISCUSSION

Molecular analysis of the four copies of NGRRS-1 identified three different insertion-like sequences, NGRIS-2, NGRIS-4, and the reconstructed NGRIS-10. There are nine copies of NGRIS-4, two of which are on pNGR234*a*. Interestingly, all nine NGRIS-4 elements appear to have transposed into other ISs. This is also true of the two copies of NGRIS-2 carried by the symbiotic plasmid. Hybridization data confirmed that all the known IS elements are closely linked to the sequences enriched by subtractive DNA hybridization. Since many of these unique sequences represent fragments of IS and transposons (39), there is a nonrandom distribution of transposable elements in NGR234. Although comparable data on IS of other *Rhizobium* strains are minimal, integration of IS*Rm3* within IS*Rm5* of *R. meliloti* IZ450 (30) suggests that clustering of IS-like sequences may be a general feature of rhizobial genomes. It is unclear whether accumulation of transposable elements in several islands in the genome of NGR234 is due to the intrinsic nature of the IS involved or to some selection process which eliminates mutants within other coding regions. A direct consequence of the preferential insertion of ISs into former transposable elements is to limit the chance of disrupting vital cellular functions, which would be detrimental to their dissemination.

Similarly, distribution of IS elements within the different replicons of the genome is nonlinear. Compared with the rest of the genome (39), the higher proportion of subtracted DNA sequences in pNGR234*a* correlates well with the findings that IS elements occur more frequently in plasmids than in chromosomes (18). Similar observations were made with various *R. leguminosarum* strains (33, 51) as well as with the archaeon *Haloferax volcanii* (14). Presumably, transposition into plasmid sequences is less likely to disrupt vital functions, especially if a large portion of the replicon consists of IS islands. Since many plasmids, including pNGR234*a* (17), are transmissible, they probably shuttle transposable elements into new genomic backgrounds. Once within their new host, IS elements may transpose again and disseminate to the rest of the genome. The 26 polymorphic nucleotide positions in the 880 bases that comprise the 5′- and 3′-border regions of NGRRS-1 (11 bases of 475 at the 5′ end and 15 bases of 405 at the 3′ end [Fig. 2]) suggest that the copy carried by pSym (NGRIS-10*a*) is the oldest of the four NGRIS-10 elements. In contrast, the remaining three chromosomal copies, which differ only by single-point mutations, probably result from more recent transpositions. In this respect, analysis of polymorphic bases together with the 61-bp deletion found at the 5′ end of NGRIS-10*c* (Fig. 2-A) indicates that NGRIS-10*b* and -10*d* are the most recent of these elements.

Interestingly, 108 of the 115 nucleotide differences between NGRIS-10*a* and -10*b* occur within predicted coding regions, and 92 of them are in the third codon position. Only 11 polymorphic nucleotides have changed amino acid compositions, seven base changes occur outside putative genes, and a single nucleotide replacement resulted in a nonsense mutation within *orf1* of NGRIS-10*a*. This strong bias in favor of same-sense mutations suggests that most substitutions took place before the insertion of NGRIS-4 into the NGRIS-10 repeats. Unfortunately, the mechanism by which so many silent mutations have been selected and established within the four copies of NGRIS-10 is unknown. In contrast, the DNA sequences of the two NGRIS-2 elements are identical, while those of two of the three characterized copies of NGRIS-4 differ by only a single nucleotide replacement (NGRIS-4*a* and NGRIS-4*c*). Thus, it seems that unlike the case for NGRIS-10, transposition of NGRIS-4 and NGRIS-2 took place much more recently and within in a short period of time.

NGRIS-2 is specific to pNGR234*a*. RFRS9, one of the nine copies of repetitive sequences found in *R. fredii* USDA257, is a partial homolog of NGRIS-2 and maps to symbiotic plasmids of *R. fredii* strains (28). Comparison of the nucleotide sequences of many symbiotic loci as well as of the chromosome-borne copy of the *recA* and 16S rDNA genes in NGR234 and USDA257 confirmed that most symbiotic functions were probably acquired by lateral gene transfer long after these two bacteria started to diverge (40). The absence of close homologs to NGRIS-4 and NGRIS-10 in the genome of USDA257 raises the question of the more distant origin of both elements. Unlike NGR234, which was first isolated from *Lablab purpureus* nodules in Papua New Guinea (50), USDA257 was isolated from a wild soybean plant (*Glycine soja*) growing near Wuking, China (26, 27). Depending on the overall sense of genetic transmission, NGRIS-4 and NGRIS-10 were either lost or picked up during the plasmid exchanges that produced the current NGR234 and USDA257 strains.

Complete sequencing of pNGR234*a* showed that IS and transposon-like elements make up 18% of the symbiotic plasmid (17). Although the proportion of ISs in the rest of NGR234 genome is unknown, a significant number of ISs are clustered in several islands outside pNGR234*a*, probably forming large and complex structures. We are currently cataloguing the various classes of IS and transposon-like sequences in NGR234. In this way, the proportion of the whole genome which is composed of transposable sequences will be assessed and compared to that of the symbiotic plasmid. At 6 kb, NGRRS-1 is the largest repeat identified in NGR234, but RFLP data suggest that it is not unique. The effects of such large and almost perfectly conserved repeats on the structure and plasticity of the NGR234 genome remain unclear, but it is likely that these duplications promote major rearrangements (42) and confer a very dynamic structure on the rhizobial genome.

ACKNOWLEDGMENTS

We thank R. Fellay and D. Gerber for their help in many aspects of this work.

Financial support for this project was provided by the Fonds National Suisse de la Recherche Scientifique (grants 31-36454.92 and 31-45921.95) and the Université de Genève.

REFERENCES

1. **Alam, J., J. M. Vrba, Y. Cai, J. A. Martin, L. J. Weislo, and S. E. Curtis.** 1991. Characterization of the IS*895* family of insertion sequences from the cyanobacterium *Anabaena* sp. strain PCC 7120. J. Bacteriol. **173:**5778–5783.
2. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
3. **Badenoch-Jones, J., T. A. Holton, C. M. Morrison, K. F. Scott, and J. Shine.** 1989. Structural and functional analysis of nitrogenase genes from the broad host-range *Rhizobium* strain ANU240. Gene **77:**141–153.
4. **Balatti, P. A., L. G. Kovacs, H. B. Krishnan, and S. G. Pueppke.** 1995. *Rhizobium* sp. NGR234 contains a functional copy of the soybean cultivar specificity locus, *nolXWBTUV*. Mol. Plant-Microbe Interact. **8:**693–699.
5. **Beringer, J. E.** 1974. R-factor transfer in *Rhizobium leguminosarum*. J. Gen. Microbiol. **84:**188–198.
6. **Bjourson, A. J., C. E. Stone, and J. E. Cooper.** 1992. Combined subtraction hybridization and polymerase chain reaction amplification procedure for isolation of strain-specific *Rhizobium* DNA sequences. Appl. Environ. Microbiol. **58:**2296–2301.
7. **Borodovsky, M., E. V. Koonin, and K. E. Rudd.** 1994. New genes in old sequence: a strategy for finding genes in the bacterial genome. Trends Biochem. Sci. **19:**309–313.
8. **Britten, R. J., and D. E. Kohne.** 1968. Repeated sequences in DNA. Science **161:**529–540.
9. **Broughton, W. J., N. Heycke, H. Meyer z.A., and C. E. Pankhurst.** 1984. Plasmid linked *nif* and *nod* genes in fast-growing rhizobia that nodulate *Glycine max*, *Psophocarpus tetragonolobus*, and *Vigna unguiculata*. Proc. Natl. Acad. Sci. USA **81:**3093–3097.
10. **Broughton, W. J., C. H. Wong, A. Lewin, U. Samrey, H. Myint, H. Meyer z.A., D. N. Dowling, and R. Simon.** 1986. Identification of *Rhizobium* plasmid sequences involve in recognition of *Psophocarpus*, *Vigna*, and other legumes. J. Cell Biol. **102:**1173–1182.
11. **Bult, J. C., O. White, G. J. Olsen, and J. C. Venter.** 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. Science **273:**1058–1072.
12. **Cami, B., and P. Kourilsky.** 1978. Screening of cloned recombinant DNA in bacteria by in situ colony hybridization. Nucleic Acids Res. **5:**2381–2390.
13. **Chua, K. J., C. E. Pankhurst, P. E. MacDonald, D. H. Hopcroft, B. D. W. Jarvis, and D. B. Scott.** 1985. Isolation and characterization of Tn*5*-induced symbiotic mutants of *Rhizobium loti*. J. Bacteriol. **162:**335–343.
14. **Cohen, A., W. L. Lam, R. L. Charlebois, W. F. Doolittle, and L. C. Schalkwyk.** 1992. Localizing genes on the map of the genome *Haloferax volcanii*, one of the Archaea. Proc. Natl. Acad. Sci. USA **89:**1602–1606.
15. **Dusha, I., S. Kovalenko, Z. Banfalvi, and A. Kondorosi.** 1987. *Rhizobium meliloti* insertion element IS*Rm2* and its use for identification of the *fixX* gene. J. Bacteriol. **169:**1403–1409.
16. **Freiberg, C., X. Perret, W. J. Broughton, and A. Rosenthal.** 1996. Sequencing the 500 kb GC-rich symbiotic replicon of *Rhizobium* sp. NGR234 using dye terminators and a thermostable "sequenase": a beginning. Genome Res. **6:**590–600.
17. **Freiberg, C., R. Fellay, A. Bairoch, W. J. Broughton, A. Rosenthal, and X. Perret.** 1997. Molecular basis of symbiosis between *Rhizobium* and legumes. Nature **387:**394–401.
18. **Galas, D. J., and M. Chandler.** 1989. Bacterial insertion sequences, p. 109–162. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
19. **Gibson, T. J., A. R. Coulson, J. E. Sulston, and P. F. R. Little.** 1987. Lorist2, a cosmid with transcriptional terminators insulating vector genes from interference by promoters within the insert: effect on DNA yield and cloned insert frequency. Gene **53:**275–281.
20. **Hanahan, D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166:**557–580.
21. **Herridge, D. F., and R. J. Roughley.** 1975. Variation in colony characteristics and symbiotic effectiveness in *Rhizobium*. J. Appl. Bacteriol. **38:**19–27.
22. **Heron, D. S., and S. G. Pueppke.** 1984. Mode of infection, nodulation specificity, and indigenous plasmids of 11 fast-growing *Rhizobium japonicum* strains. J. Bacteriol. **160:**1061–1066.
23. **Jouanin, L., J. Tourneur, and F. Casse-Delbart.** 1986. Restriction maps and homologies of the three plasmids of *Agrobacterium rhizogenes* strain A4. Plasmid **16:**124–134.
24. **Judd, A. K., and M. J. Sadowsky.** 1993. The *Bradyrhizobium japonicum* serocluster 123 hyperreiterated DNA region, HRS1, has DNA and amino acid sequence homology to IS*1380*, an insertion sequence from *Acetobacter pasteurianus*. Appl. Environ. Microbiol. **59:**1656–1661.
25. **Kaluza, K., M. Hahn, and H. Hennecke.** 1985. Repeated sequences similar to insertion elements clustered around the *nif* region of the *Rhizobium japonicum* genome. J. Bacteriol. **162:**535–542.
26. **Keyser, H. H., B. B. Bohlool, T. S. Hu, and D. F. Weber.** 1982. Fast-growing rhizobia isolated from root nodules of soybean. Science **215:**1631–1632.
27. **Keyser, H. H., and R. F. Griffin.** 1987. Beltsville *Rhizobium* culture collection catalog. Agricultural Research Service, U.S. Department of Agriculture, Beltsville, Md.
28. **Krishnan, H. B., and S. Pueppke.** 1993. Characterization of RFRS9, a second member of the *Rhizobium fredii* repetitive sequence family from the nitrogen-fixing symbiont *R. fredii* USDA257. Appl. Environ. Microbiol. **59:**150–159.
29. **Kuykendall, L. D., and B. Saxena.** 1992. Genetic diversity in *Bradyrhizobium japonicum* Jordan 1982 and a proposal for *Bradyrhizobium elkanii* sp. nov. Can. J. Microbiol. **38:**501–505.
30. **Laberge, S., A. T. Middleton, and R. Wheatcroft.** 1995. Characterization, nucleotide sequence, and conserved genomic locations of insertion sequence IS*Rm5* in *Rhizobium meliloti*. J. Bacteriol. **177:**3133–3142.
31. **Lewin, A., P. Rochepeau, X. Perret, E. Cervantes, and W. J. Broughton.** 1988. Determinants of broad host-range in *Rhizobium* spp. NGR234, p. 53–54. *In* R. Palacios and D. P. S. Verma (ed.), Molecular genetics of plant-microbe interactions—1988. American Phytopathological Society, St. Paul, Minn.
32. **Martinez, E., D. Romero, and R. Palacios.** 1990. The *Rhizobium* genome. Crit. Rev. Plant Sci. **9:**59–93.
33. **Mazurier, S. I., L. Rigottier-Gois, and N. Amarger.** 1996. Characterization, distribution, and localization of IS*RI2*, and insertion sequence element isolated from *Rhizobium leguminosarum* bv. viciae. Appl. Environ. Microbiol. **62:**685–693.
34. **Morrison, N. A., C. Y. Hau, M. J. Trinick, J. Shine, and B. G. Rolfe.** 1983. Heat curing of a Sym plasmid in a fast-growing *Rhizobium* sp. that is able to nodulate legumes and the nonlegume *Parasponia* sp. J. Bacteriol. **153:**527–531.
35. **Ogawa, J., H. L. Brierley, and S. Long.** 1991. Analysis of *Rhizobium meliloti* nodulation mutant WL131: novel insertion sequence IS*Rm3* in *nodG* and altered *nodH* protein product. J. Bacteriol. **173:**3060–3065.
36. **Osteras, M., J. Stanley, and T. M. Finan.** 1995. Identification of rhizobium-specific intergenic mosaic elements within an essential two-component regulatory system of *Rhizobium* species. J. Bacteriol. **177:**5485–5494.
37. **Perret, X., W. J. Broughton, and S. Brenner.** 1991. Canonical ordered cosmid library of the symbiotic plasmid of *Rhizobium* species NGR234. Proc. Natl. Acad. Sci. USA **88:**1923–1927.
38. **Perret, X.** 1992. Physical and genetic mapping of the genome of *Rhizobium*

species NGR234. Ph.D. thesis 2489. University of Geneva, Geneva, Switzerland.

39. **Perret, X., R. Fellay, A. J. Bjourson, J. E. Cooper, S. Brenner, and W. J. Broughton.** 1994. Subtraction hybridization and shot-gun sequencing: a new approach to identify symbiotic loci. Nucleic Acids Res. **22:**1335–1341.
40. **Perret, X., and W. J. Broughton.** Rapid identification of *Rhizobium* strains by targeted PCR fingerprinting. *In* G. Hardarson and W. J. Broughton (ed.), Molecular biology in soil microbial ecology, in press.
41. **Relić, B., X. Perret, M. T. Estrada-Garcia, J. Kopcinska, W. Golinowski, H. B. Krishnan, S. G. Pueppke, and W. J. Broughton.** 1994. *Nod* factors of *Rhizobium* are a key to the legume door. Mol. Microbiol. **13:**171–178.
42. **Romero, D., J. Martinez-Salazar, L. Girard, S. Brom, G. Davila, R. Palacios, M. Flores, and C. Rodriguez.** 1995. Discrete amplifiable regions (amplicons) in the symbiotic plasmid of *Rhizobium etli* CFN42. J. Bacteriol. **177:**973–980.
43. **Rosenberg, C., P. Boistard, J. Dénarié, and F. Casse-Delbart.** 1981. Genes controlling early and late functions in symbiosis are located on a megaplasmid in *Rhizobium meliloti*. Mol. Gen. Genet. **184:**326–333.
44. **Ruvkun, G. B., S. R. Long, H. M. Meade, R. C. van den Bos, and F. M. Ausubel.** 1982. IS*Rm1*: a *Rhizobium meliloti* insertion sequence that transposes preferentially into nitrogen fixation genes. J. Mol. Appl. Genet. **1:**405–418.
45. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
46. **Sanger, F., S. Nicklen, and R. A. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**1757–1761.
47. **Sobral, B. W., R. J. Honeycutt, A. G. Atherly, and M. McClelland.** 1991. Electrophoretic separation of the three *Rhizobium meliloti* replicons. J. Bacteriol. **173:**5173–5180.
48. **Soto, M. J., A. Zorzano, J. Olivares, and N. Toro.** 1992. Sequence of IS*Rm4* from *Rhizobium meliloti* strain GR4. Gene **120:**125–126.
49. **Stanley, J., D. N. Dowling, and W. J. Broughton.** 1988. Cloning of *hemA* from *Rhizobium* sp. NGR234 and symbiotic phenotype of a gene-directed mutant in diverse legume genera. Mol. Gen. Genet. **215:**32–37.
50. **Trinick, M. J.** 1980. Relationships amongst the fast-growing rhizobia of *Lablab purpureus*, *Leucaena leucocephala*, *Mimosa* spp., *Acacia farnesiana* and *Sesbania grandiflora* and their affinities with the other rhizobial groups. J. Appl. Bacteriol. **49:**39–53.
51. **Ulrich, A., and A. Pühler.** 1994. The new class II transposon Tn*163* is plasmid-borne in two unrelated *Rhizobium leguminosarum* biovar *viciae* strains. Mol. Gen. Genet. **242:**505–516.
52. **van Slooten, J. C., T. V. Bhuvanasvari, S. Bardin, and J. Stanley.** 1992. Two C_4-dicarboxylate transport system in *Rhizobium* sp., NGR234: rhizobial dicarboxylate transport is essential for nitrogen fixation in tropical legume symbioses. Mol. Plant-Microbe Interact. **5:**179–186.
53. **Wheatcroft, R., and R. J. Watson.** 1988. Distribution of insertion sequence IS*Rm1* in *Rhizobium meliloti* and other gram-negative bacteria. J. Gen. Microbiol. **134:**113–121.
54. **Wheatcroft, R., and S. Laberge.** 1991. Identification and nucleotide sequence of *Rhizobium meliloti* insertion sequence IS*Rm3*: similarity between the putative transposase encoded by IS*Rm3* and those encoded by *Staphylococcus aureus* IS256 and *Thiobacillus ferrooxidans* IST2. J. Bacteriol. **173:**2530–2538.