Applied and Environmental Microbiology, June 2009, p. 4035–4045 Vol. 75, No. 12
0099-2240/09/$08.00+0 doi:10.1128/AEM.00515-09

Rhizobium sp. Strain NGR234 Possesses a Remarkable Number of Secretion Systems▽†

Christel Schmeisser,[1]‡ Heiko Liesegang,[2]‡ Dagmar Krysciak,[1]‡ Nadia Bakkou,[3] Antoine Le Quéré,[3]§ Antje Wollherr,[2] Isabelle Heinemeyer,[4] Burkhard Morgenstern,[4] Andreas Pommerening-Röser,[1] Margarita Flores,[5] Rafael Palacios,[5] Sydney Brenner,[6] Gerhard Gottschalk,[2] Ruth A. Schmitz,[7] William J. Broughton,[3]* Xavier Perret,[3] Axel W. Strittmatter,[2]‖ and Wolfgang R. Streit[1]*

Biozentrum Klein Flottbek, Abteilung für Mikrobiologie und Biotechnologie, Universität Hamburg, Ohnhorststrasse 18, 22609 Hamburg, Germany[1]; Laboratorium für Genomanalyse, Universität Göttingen, 37077 Göttingen, Germany[2]; Département de Botanique et Biologie Végétale, Université de Genève, 30 Quai Ernest-Ansermet, 1211 Geneva, Switzerland[3]; Department of Bioinformatics, Universität Göttingen, 37077 Göttingen, Germany[4]; Centro de Ciencas Genomicas, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, Mexico[5]; Salk Institute for Biological Studies Crick-Jacobs Center, P.O. Box 85800, San Diego, California 92186-5800[6]; and Institut für Allgemeine Mikrobiologie, Christian-Albrechts-Universität zu Kiel, Am Botanischen Garten 1-9, 24118 Kiel, Germany[7]

Received 2 March 2009/Accepted 8 April 2009

***Rhizobium* sp. strain NGR234 is a unique alphaproteobacterium (order *Rhizobiales*) that forms nitrogen-fixing nodules with more legumes than any other microsymbiont. We report here that the 3.93-Mbp chromosome (cNGR234) encodes most functions required for cellular growth. Few essential functions are encoded on the 2.43-Mbp megaplasmid (pNGR234*b*), and none are present on the second 0.54-Mbp symbiotic plasmid (pNGR234*a*). Among many striking features, the 6.9-Mbp genome encodes more different secretion systems than any other known rhizobia and probably most known bacteria. Altogether, 132 genes and proteins are linked to secretory processes. Secretion systems identified include general and export pathways, a twin arginine translocase secretion system, six type I transporter genes, one functional and one putative type III system, three type IV attachment systems, and two putative type IV conjugation pili. Type V and VI transporters were not identified, however. NGR234 also carries genes and regulatory networks linked to the metabolism of a wide range of aromatic and nonaromatic compounds. In this way, NGR234 can quickly adapt to changing environmental stimuli in soils, rhizospheres, and plants. Finally, NGR234 carries at least six loci linked to the quenching of quorum-sensing signals, as well as one gene (*ngrI*) that possibly encodes a novel type of autoinducer I molecule.**

Diverse soil bacteria interact with plants in ways that range from symbiotic to pathogenic. Symbiotic *Eubacteria* (both alpha- and betaproteobacteria, collectively called rhizobia) form nitrogen-fixing associations of tremendous environmental importance (41, 66). Although some rhizobia are able to reduce atmospheric nitrogen to ammonia under saprophytic, free-living conditions, the reduced oxygen tensions found within the intracellular environment of specialized organs called nodules, maximizes this process (16). As legume roots penetrate the soil, they come in contact with rhizobia. Symbiotic interactions are initiated by the exchange of diverse molecules between the partners. Among them, plants liberate flavonoids into the rhizosphere that upregulate rhizobial genes. As a result, lipo-chito-oligo-saccharidic Nod factors are produced that trigger the nodulation pathway in susceptible legumes. Then, in many hosts, rhizobia enter the roots through root hairs, make their way to the cortex, multiply and fill the intracellular spaces of mature nodules. Centripetal progression of rhizobia into the plant and their maturation into nitrogen-fixing symbiosomes depends on the continued exchange of diverse signals. Many, but not all of these signals have been identified; one sure way to take stock of what is necessary for effective symbiosis is to sequence the partners. We began this work by assembling overlapping sets of cosmids (contigs) of the microsymbiont *Rhizobium* sp. strain NGR234 (hereafter NGR234) (63), which enabled us to elucidate the nucleotide sequence of the symbiotic (pNGR243*a*) plasmid (29). Similar techniques permitted the assembly of sections of the extremely large megaplasmid pNGR234*b* (86), and some snapshot genome information was made available earlier (91); however, the use of pyrosequencing methods greatly facilitated this process. We report here the genome sequence of NGR234 that is able to nodulate more than 120

* Corresponding author. Mailing address for W. J. Broughton: Laboratoire de Biologie Moléculaire des Plantes Supérieures, Université de Genève, 30 Quai Ernest-Ansermet, 1211 Geneva, Switzerland. Phone: (41) 22-3793108. Fax: (41) 22-3793009. E-mail: william.broughton@unige.ch. Mailing address for W. R. Streit: Biozentrum Klein Flottbek, Abteilung für Mikrobiologie und Biotechnologie, Universität Hamburg, Ohnhorststrasse 18, 22609 Hamburg, Germany. Phone: (49) 40-42816463. Fax: (49) 40-42816-459. E-mail: wolfgang.streit@uni-hamburg.de.

‡ C.S., H.L., and D.K. contributed equally to this study.

§ Present address: Laboratoire des Symbioses Tropicales et MéditerranéennesTA A-82/J Campus International de Baillarguet, 34398 Montpellier Cedex 5, France.

‖ Present address: Eurofins MWG Operon, Anzinger Strasse 7a, D-85560 Ebersberg, Germany.

† Supplemental material for this article may be found at http://aem.asm.org/.

▽ Published ahead of print on 17 April 2009.

genera of legumes and the nonlegume *Parasponia andersonii* (69). It seems likely that the vast richness of secretory systems might be a major key to the broad host range.

MATERIALS AND METHODS

Bacterial strains and growth conditions. *Escherichia coli* was grown at 37°C in Luria-Bertani medium (76) supplemented with appropriate antibiotics, and NGR234 was grown in TY (0.5% tryptone, 0.25% yeast extract, 10 mM $CaCl_2$ [pH 7.0]).

Sequencing, gap closure, annotation, and bioinformatic tools. DNA used to prepare shotgun libraries of NGR234 was isolated under standard conditions by using a MasterPure DNA purification kit (Biozym Scientifc GmbH, Oldendorf, Germany) and fragmented to between 2.0 and 4.0 kbp either by mechanical shearing (Nebulizer; Invitrogen, Carlsbad, CA) or by partial enzymatic digestion. After end repair, DNA fragments were separated by gel electrophoresis and cloned as described previously (86) using the pTZ19R vectors (Amersham Pharmacia Biotech, Essex, United Kingdom) or pCR4.1-TOPO (TOPO-TA cloning kit for sequencing; Invitrogen). About 30,000 plasmids were isolated by using two BioRobots8000 (Qiagen, Hilden, Germany), and 60,597 sequences were automatically analyzed on ABI Prism models 377-96 and 3730XL (Applied Biosystems, Darmstadt, Germany). PCR-based techniques were used to close the remaining gaps using both the shotgun library and the ordered cosmid library clones as templates. Finally, a single 454 sequencing run that generated 110 Mbp of raw sequences was used to close the gaps. The 454 sequencing was done according to the manufacturer's protocols (Roche Applied Science, Mannheim, Germany). Sanger and 454 contigs were manually coassembled and curated. All manual editing steps were performed by using the GAP4 software package v4.5 and v4.6 (79). Coding sequences (CDS) and open reading frames (ORFs) were predicted with YACOP (87) using the ORF finders Glimmer, Critica, and Z-curve. All CDS were manually curated and verified by comparison with the publicly available databases SwissProt, GenBank, ProDom, COG, and Prosite using the annotation software ERGO (61).

Symbiotic plasmid-borne ORFs are identified as NGR_a (e.g., NGR_a04220 [*traI*]), those on the megaplasmid are identified as NGR_b (NGR_b22000 [ribulose-bis-phosphate carboxylase; EC 4.1.1.39]), while those on the chromosome are designated NGR_c (NGR_c26660 [*bioC*]). The DNA sequences of the three replicons have been deposited at GenBank. An update of the original pNGR234*a* sequence NC_000914 can be found under U00090, the pNGR234*b* replicon carries the number CP000874, and the bacterial chromosome is listed under the number CP001389.

To detect small RNA (sRNA) genes, we combined comparative sequence analysis with structure prediction. Then, we compared the genome sequence of NGR234 with five phylogenetically related bacteria. Since we expected sRNA genes and regulatory elements to be within intergenic regions at or near the start of operons, this technique was only applied to extended intergenic regions rather than the complete genomic sequence. Inter- and intraorganism homologies were obtained by applying the BLAST program (1) to all extended intergenic regions. Overlapping BLAST homologies were extended both upstream and downstream of the overlap. All extracted sequences were aligned with CLUSTAL W (88). Structural conservation and stability within multiple alignments was used to predict candidate regions for sRNA genes using RNAz (92). This search was refined by using the software tool INFERNAL (20). INFERNAL scans candidate sequences against covariance models of known RNA structures. In our study, we used the covariance models of sRNA families in the database Rfam (33).

Comparative genomics. Putative orthologous genes were identified by using bidirectional BLAST comparisons among a representative set of whole-genome protein datasets using a variant of the BLAST method developed by A. Wollherr and H. Liesegang (unpublished data).

Construction of cosmid libraries. Cosmid libraries of partially Sau3AI restricted genomic DNA of NGR234 were established in Lorist2 and pWE15 (Stratagene/Agilent, Foster City, CA) as described previously (23, 63). Initially, 2,000 Lorist2 inserts were grouped into sets of overlapping clones (contigs) using a combination of BamHI, EcoRI, and HindIII fingerprints, as well as hybridization and sequencing data (63). Cosmid DNA was end sequenced using the ForLor (5′-GCTTGTACATATTGTCGTTAGAACGCGG-3′) or RevLor (5′-TCTCCGGGAGCTGCATGTGTCAGAGG-3′) primers that are complementary to the Lorist2 border sequences and T7 (5′-TAATACGACTCACTATAGGG-3′) or T3 (5′-ATTAACCCTCACTAAAGGGA-3′) promoter primers that are complementary to the pWE15 border sequences.

TABLE 1. General features of the NGR234 chromosome and plasmids pNGR234*a* and pNGR234*b*[a]

Characteristic/function	pNGR234*a*	pNGR234*b*	cNGR234	Genome
Size (Mbp)	0.54	2.43	3.93	6.90
GC content (avg mol%)	58.4	62.3	62.8	61.2
No. of rRNA operons	0	0	3	3
No. of tRNAs	1	0	52	53
No. of ncRNAs	3	17	22	42
Total no. of features	418	2,343	3,633	6,394
No. of hypothetical CDS	138	644	943	1,725
No. of transposases	114	142	73	329
No. of Nod/Nif	46	8	5	59
No. of regulatory proteins	22	207	232	461
No. of sigma subunits	0	3	11	14
No. of transport proteins	29	362	291	682
No. of secretory proteins	32	57	43	132
No. of other proteins	37	920	2,035	2,992

[a] The data for pNGR234*a* were in part extracted from a previous study (29); the data for the secretion systems include those involved in the assembly of type I to VI secretion machines, the type IV attachment pili, those involved in constructing the SRP, and the twin arginine and general export (Sec) pathways. Flagellar genes have not been included.

RESULTS AND DISCUSSION

General features of the NGR234 genome. NGR234 contains three replicons totaling 6,891,900 bp (Table 1). The 3,925,702-bp chromosome encodes 3,633 ORFs (gene density, 1.06 genes per kbp) with a mean G+C content of 62.8%. The 2,430,033-bp pNGR234*b* replicon comprises 2,342 ORFs (gene density, 1.03 genes per kbp), with a mean G+C content of 62.3%. The previously published pNGR234*a* replicon (536,165 bp) (29) has a significantly lower G+C content (58.4%) and a lower gene density (1.27 genes per kbp). The overall G+C content of NGR234 is 61.2%, which is similar to the previously determined value (6). The distribution of the ORFs on the forward and the reverse strand is almost identical on all three replicons (Fig. 1).

At 9.1 Mbp, the genome of *Bradyrhizobium japonicum* USDA110 (44) is the largest rhizobial genome sequenced, and that of *Azorhizobium caulinodans* ORS571 (5.4 Mbp) (52) is one of the smallest. The NGR234 genome (6.9 Mbp) that spreads over three replicons is thus average for rhizobia and similar in structure to that of *Sinorhizobium meliloti* strain 1021 (referred to hereafter as strain 1021), a 1.4-Mbp symbiotic plasmid (pSymA), a megaplasmid of 1.7 Mbp (pSymB), and a chromosome of 3.7 Mbp (total 6.7 Mbp) (30).

Most essential genes are carried on the chromosome. A few are located on pNGR234*b*, and there are none on pNGR234*a*. The absence of essential genes on the symbiotic plasmid was expected since strains deprived of pNGR234*a* by heat curing, grew as well as the wild-type NGR234 but failed to nodulate any of its hosts (59). This was also reported for the strain 1021 counterpart pSymA (60). Except for a second copy of the tRNA-Met (CAT anticodon) predicted on pNGR234*a*, the remaining 52 tRNAs are encoded by the chromosome. Similarly, of the 55 ribosomal protein genes that were identified, only a single gene that plays a role in the modification of tRNA was predicted on pNGR234*b* (Table 2). The presence

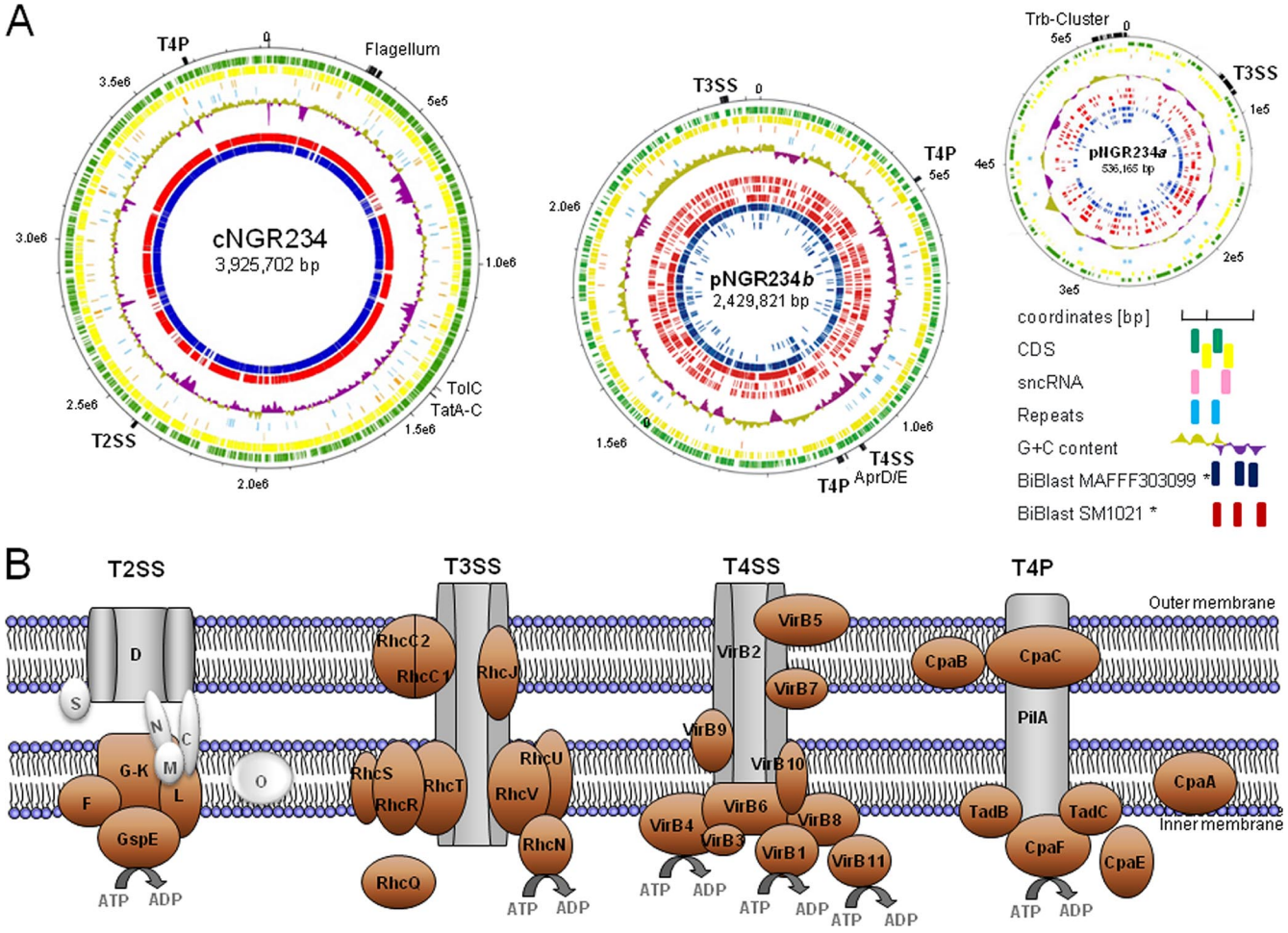

FIG. 1. (A) Map of the genome of *Rhizobium* sp. strain NGR234 chromosome (cNGR234, 3.93 Mbp), megaplasmid (pNGR234*b*, 2.43 Mbp), and sym plasmid (pNGR234*a*, 0.54 Mbp). Circles are described from outside to innermost circle. The outer circle shows the coordinates and the position of the secretion systems and several other selected genes; the second and the third outer circles indicate the ORFs on the leading and the lagging strands (yellow and green). The next inner circle indicates the sncRNAs (in light red), followed by a circle indicating the positions of the repeats (in light blue). The next inner circle shows the G+C content (dark yellow and purple). The following inner circles marked in red show the putative orthologues that are present on the NGR234 and on the strain 1021 chromosome. The circle marked in blue shows the putative orthologues that are present on the NGR234 and MAFF303099 chromosomes. This analysis is refined on the plasmids. The outer circle in red represents the putative orthologues shared with the chromosome of strain 1021, the next inner circle indicates the putative orthologues shared with pSymA, and the next red circle indicates the putative orthologues shared with pSymB. Similarly, the blue circles indicate the putative orthologues shared with the chromosome of MAFF303099, and the next inner circles indicate the putative orthologues shared with pMLa and pMLb, respectively. Larger views of the different replicons can be found in Fig. S1 to S3 in the supplemental material. (B) Hypothetical structures of the T2SS, T3SS, and T4SS and T4P of NGR234. Individual genes and proteins identified in NGR234 and their locus tags are indicated in Table S2 in the supplemental material. The positions of the various secretion systems are indicated on the outer circle; the GspC, GspS, and GspO proteins in light gray have not yet been identified in NGR234, whereas NGR_c23060 and NGR_c23050 share weak similarity to GspM and GspN, respectively. All essential genes and components of the various secretion apparatuses, which have been identified in NG234, are indicated in brown or dark gray.

on pNGR234*b* of gene copies coding for the ribosomal protein S21A (*rpsU1*, NGR_b17570) and the septum-associated component FtsK (81) (*ftsK1*, NGR_b21580), as well as the *minCDE* operon, raises the question of whether this replicon is essential for survival of the bacterium. In both NGR234 and strain 1021, the *minEDC* genes which are required for the placement of the cell division site are located on the megaplasmid, whereas in *R. etli* they are borne by the small p42e (10). In strain 1021, the phenotypes of a deletion mutant showed that the *minCDE* genes are not essential for cell viability, cell division, or symbiotic nitrogen fixation (10). Since the chromosome of NGR234 carries additional copies of *rpsU* (NGR_c29060) and *ftsK* (*ftsK2*, NGR_c31480), there is no firm indication that pNGR234*b* is essential for the survival of the strain in laboratory conditions. pNGR234*b* also encodes 3 of the 14 sigma factors (NGR_b21500, NGR_b18750, and NGR_b18130), along with the noncoding RNA (ncRNA) genes that are involved in plasmid maintenance. In summary, however, it appears that the majority of all genes encoded on pNGR234*b* are nonessential. This has already been shown in previous studies for its counterpart the pSymB (8).

Insertion sequences. Together, insertion sequence elements (ISs) and phage sequences represent 4.1% of the NGR234 genome, but their relative abundance and distribution varies between replicons (Table 1). In proportion to its size, pNGR234*a* carries more transposable elements (16.3%) than

TABLE 2. Comparative analysis of essential and additional functions encoded by pNGR234*b* and the pSymB of *S*M1021[a]

Functional class	Comparative function or characteristic	
	pNGR234*b*	pSymB[b]
Cell process	*ftsK* *minEDC* Ten chaperones None 30S ribosomal protein S21A	*ftsK* *minEDC* Four chaperones tRNA arginine (single copy) None
Amino acid biosynthesis	Asparagine synthetase (2×) *glnII,* glutamine synthetase II Shikimate dehydrogenases (3×)	Asparagine synthetase (2×) *glnII*, glutamine synthetase II Shikimate dehydrogenase (1×)
Cofactor biosynthesis	*thiD* and *thiCOGE*, plus two ORFs possible (*bioF* and *bioA*)	*thiD* and *thiCOGE*
EPS, lipopolysaccharide, and CPS biosynthesis	~195 genes	~188 genes

[a] No essential functions are encoded by pNGR234*a*.
[b] The genes identified on pSymB were derived from a previous study (27) and GenBank accession number AL591985.

pNGR234*b* (4.8%) or the chromosome (2.1%). This feature, which is consistent with the recombinatorial nature and foreign origin of pNGR234*a*, was proposed earlier (29). Unlike the NGR234 chromosome, both plasmids seem prone to DNA rearrangements and include segments of up to 12 kb exclusively constituted of several layers of ISs that transposed into one another, however. ISs that range in size from 623 to 3,316 bp, were grouped into 36 families, the largest of which includes 10 copies of the NGRIS-4 transposable element (67). Of these 36 types of ISs, five families are chromosome specific, six are pNGR234*a* specific, and seven are pNGR234*b* specific.

Catabolic and metabolic functions. All major genes of the pentose phosphate and the Entner-Doudoroff pathways are present on the chromosome, suggesting that these are probably the main routes of carbon metabolism under aerobic and free-living growth conditions. Genes encoding functions associated with the degradation of larger polymers such as cellulose (NGR_b15380), glycogen (NGR_b11120), poly-β-hydroxybutyrate (NGR_b05400, NGR_c23850), pectin (NGR_b17700), and starch (NGR_b11110) are present on both the chromosome and pNGR234*b*. The cellulases and pectinases might play special roles during infection of the plant. Many genes involved in the catabolism of aromatic substrates were identified, including more than 30 ORFs linked to mono- and dioxygenases. It is possible these genes are used to modify aromatic compounds (i.e., flavonoids and phytoalexins) exuded by the plant (11). It is also noteworthy that a gene cluster containing nine genes (NGR_b20090 to NGR_b20170) was identified that is probably linked to 4-phenylacetate metabolism, along with genes involved in catabolism of opines (*ooxA*, *ooxB*, and *agaE*; NGR_b06440 to NGR_b06470, NGR_b10720).

Some *B. japonicum* symbionts of soybeans that possesses ribulose-bis-phosphate carboxylase (EC 4.1.1.39) are capable of autotrophic growth (53). Autotrophy seems unlikely in NGR234 since the genome only encodes the two copies of the large subunit of the ribulose-bis-phosphate carboxylase (NGR_b22000 and NGR_c06470) but lacks the smaller subunit.

Twenty-five putative decarboxylases and four biotin-dependent carboxylases were identified, mostly on pNGR234*b*. Interestingly, pNGR234*b* encoded the acetyl coenzyme A (CoA) carboxylase, the propionyl-CoA carboxylase, and the 3-methylcrotonyl-CoA carboxylase, while an additional copy of acetyl-CoA carboxylase is carried on the chromosome. Most probably these enzymes are linked to fatty acid synthesis (12), as well as the synthesis of polyketides derived from the carboxylation of acetyl-, butyryl-, or propionyl-CoA (39).

Vitamin biosynthesis. Genes linked to the biosynthesis of the cobalamin (vitamin B_{12}) and riboflavin (vitamin B_2), along with three B_{12} riboswitches (NGR_c16930, NGR_c21900, and NGR_c26330), are located on the chromosome (see Table S1 in the supplemental material). One of these is located directly upstream of the putative cobalt transporter genes, *cbtAB* (NGR_c21910 and NGR_c21920). Interestingly, a possible vitamin B_{12} riboswitch is also carried on pNGR234*b* (NGR_b02500). Relatively few genes that encode essential pathways for cofactors are present on pNGR234*b*, but among those found include some coding for pyrolloquinolone, pyridoxal phosphate, and thiamine synthesis. Altogether, four thiamine synthesis genes, *thiEGOC* (NGR_b01980 and NGR_b02900 to NGR_b02930), along with the related *thiD* (NGR_b18410), are present. Genes linked to thiamine transport are carried by the chromosome (NGR_c32610, NGR_c32620, and NGR_c32630) however. Some of these are essential for bacterial colonization of the rhizosphere (85). A minimum of six to seven genes are generally required for biotin biosynthesis in gram-negative bacteria and four genes—*bioADBF* (NGR_c25110 to NGR_c25140)—are clustered on the chromosome but separate from the fifth gene, *bioC* (NGR_c26660). Two additional genes possibly linked to synthesis of biotin are present on pNGR234*b*: one encodes an 8-amino-7-oxononanoate synthase (NGR_b10670, a possible *bioF* homologue), while the other, a possible *bioA* homologue (NGR_b06270), encodes an adenosylmethionine-8-amino-7-oxononanoate aminotransferase. Furthermore, genes involved in biotin transport (*bioMNY* [NGR_c05050, NGR_c05060, and NGR_c05040]) and BioS (NGR_c13770) a sensor/regulator protein (36, 37) were identified. Intriguingly, BioS proteins have only been found in strains NGR234 and 1021 (37). Perhaps this indicates that biotin has a special regulatory function in these microbes. As a cofactor of carboxylases, biotin (a B

vitamin) has profound effects on the metabolism of rhizobia (21, 22, 38) and root colonization (85).

pNGR234*b* carries many transporters. Since ~12% of all genes in *B. japonicum*, *M. loti*, *S. meliloti*, and other plant-associated bacteria encode transporters (especially of small molecules), the 682 genes (10.5% of all genes) found in NGR234 is slightly below average. The majority specify ABC transporters, but 19 genes encode MFS-like transporters. As with the distribution pattern seen in strain 1021 (27), 291 transporter genes are present on the chromosome, 362 are present on pNGR234*b*, and 29 are present on pNGR234*a* (Table 1). Since 50% of all transporters are carried by pNGR234*b*, this replicon is obviously of great importance for growth and the survival of the microbe under a wide range of environmental and nutritional conditions.

Regulatory elements. Altogether, 466 ORFs (7.2%, Table 1) encode regulatory proteins, particularly of the LysR (64 gene copies) and GntR (42 gene copies) families. After activation by inducers, LysR proteins modulate the activity of regulons with diverse functions (77), while members of the GntR family bind to promoters and downregulate transcription (34). Genes encoding other regulators, including the AraC (38 gene copies), ArsR (7 gene copies), AsnC (28 gene copies), LacI (33 gene copies), and TetR (24 gene copies) familes, as well as the DeoR, LuxR, MarR, and MerR families, were found, many of which are probably involved in the regulation of antibiotic synthesis, efflux pumps, and resistance to metals (70). A further 63 ORFs encode two component regulators. Although the majority of the regulator genes are found on the chromosome (6.4% [Table 1]), the overall percentage is higher on pNGR234*b* (8.7%).

We used bioinformatics' tools to search for genes that encode phylogenetically conserved, stable noncoding, and regulatory RNAs (so-called ncRNAs [33]) (Table 1). Many were found, especially the well-known genes for signal recognition particles (SRP), RNase P, and 6S RNA that are present in all known genomes. Other genes typical of the alphaproteobacteria such as *suhB* are also present. A summary of the most prominent regulatory RNAs and riboswitches is given in Table S1 in the supplemental material.

Quorum sensing and cell-cell communication. Quorum sensing is a cell density-dependent system of gene regulation and cell-cell communication in prokaryotes (62). Bacterial populations sense increases in cell density via signal molecules called autoinducers that, together with dedicated regulators, modulate gene expression accordingly. Many quorum-sensing mechanisms involve *N*-acyl-homoserine lactones (N-AHLs) in gram-negative bacteria and modified oligopeptides in gram-positive bacteria (62). Plasmid replication (rep) and conjugal transfer (tra) functions, including homologues of the Ti plasmid quorum-sensing regulators TraI (NGR_a04220) and TraR (NGR_a04090), are present on pNGR234*a* (29). TraI synthesizes an acyl-HSL that is probably 3-oxo-C8-HSL, but *traI* mutants and a pNGR234*a*-cured derivative produce low levels of a similar acyl-HSL along with another, more hydrophobic signal molecule (35). Thus, the discovery of a chromosome bound *traI* homologue (NGR_c16900), together with its possible regulator (NGR_c16890) (designated *ngrI* and *ngrR*, respectively), probably explains these observations (Fig. 2).

NGR234 also carries many genes involved in the degradation of N-AHLs, including three putative lactonases (NGR_b01930, NGR_b22150, and NGR_c03800). Using a recently published protocol (80) to screen a cosmid library of NGR234, we identified at least six loci (three on the chromosome and three on pNGR234*b*) that are actively involved in N-AHL degradation (51). Overexpression, purification and biochemical characterization of two of the proteins showed that one is similar to a metal-dependent β-lactamase (NGR_b16870), while the other resembles a bacterial dienelactone hydrolase (NGR_b22150) (51).

Polysaccharide synthesis. Various rhizobial exopolysaccharides (EPS), capsular polysaccharides (K antigens, also referred to as KPS), lipopolysaccharides, and smaller periplasmic glucans (4, 7, 83) participate in nodule formation. Most genes linked to surface polysaccharide synthesis are located on megaplasmids in both strains NGR234 (~195 genes) and 1021 (~188 genes) (Table 2). Overall, the genetic organization of the *exo* clusters, as well as the constituent genes (>80% identities) is very similar in strains NGR234 and 1021 (9, 86). In NGR234, more than 25 genes stretching from *exoU* (NGR_b19500) to *exsI* (NGR_b01520) are involved in the synthesis of low-molecular-weight EPS that are essential for nodule invasion on some plants (84). Notable differences in homology occur on both sides of the conserved *exoI* region, however. ORFs corresponding to *exoH* and *exoTWV* are not present in the pNGR234*b* cluster (86), although the separate *exoH* (NGR_b13480) is flanked by transposases, suggesting that it has recently moved there. Since the acidic EPS of NGR234 are not succinylated (84), it seems likely that this copy of *exoH* is either nonfunctional or has a distinct physiological role. A possible *exoT* homologue (NGR_b13540) is located several kilobases downstream of *exoH*, where it lies together with two glycosyltransferases (NGR_b13520 and NGR_b13530). Since *exoW* of strain 1021 encodes a glycosyltransferase (5), perhaps these ORFs represent *exoW* homologues. ExoV replaces the terminal glucose of the strain 1021 succinoglycan subunit with a pyruvyl group. Although the NGR234 *exo* cluster lacks *exoV*, the nonreducing galactose of the subunit is also pyruvylated (15, 84), suggesting that an unidentified pyruvyl transferase must exist in NGR234. Since a clear homologue of *exoV* was not found in the NGR234 genome, another enzyme must be responsible for the pyruvylation of galactose.

K antigens are tightly associated with the rhizobial outer membrane and are thus distinct from the loosely adhering EPS. Rhizobial K antigens are strain-specific antigens that are structurally analogous to the group II K antigens of *E. coli* (45, 74). In NGR234 the major K antigens consist of polymeric 5,7-diacetamido-3,5,7,9-tetradeoxy-non-2-ulosonic acid (54). As such, it is almost identical to the K_R5 antigen of *S. meliloti* strain Rm41 (75). Three clusters of genes (*rkp-1*, *rkp-2*, and *rkp-3*) are involved in the production of the K_R5 antigen (4). Chromosomally located *rkp-1* and *rkp-2* clusters are probably responsible for the production of a specific lipid carrier necessary for the synthesis of K antigens and in the metabolism of nucleotide diphospho sugars, respectively. In contrast, *rkp-3* is carried by pSymB of Rm41 and pNGR234*b*. As expected from similarities in their K antigen structures, the coding regions of Rm41 and NGR234 are similar, containing genes that are required for the synthesis of the strain-specific sugar precursors (*rkpL*, *rkpM*, *rkpN*, *rkpP*, and *rkpQ*), as well as those

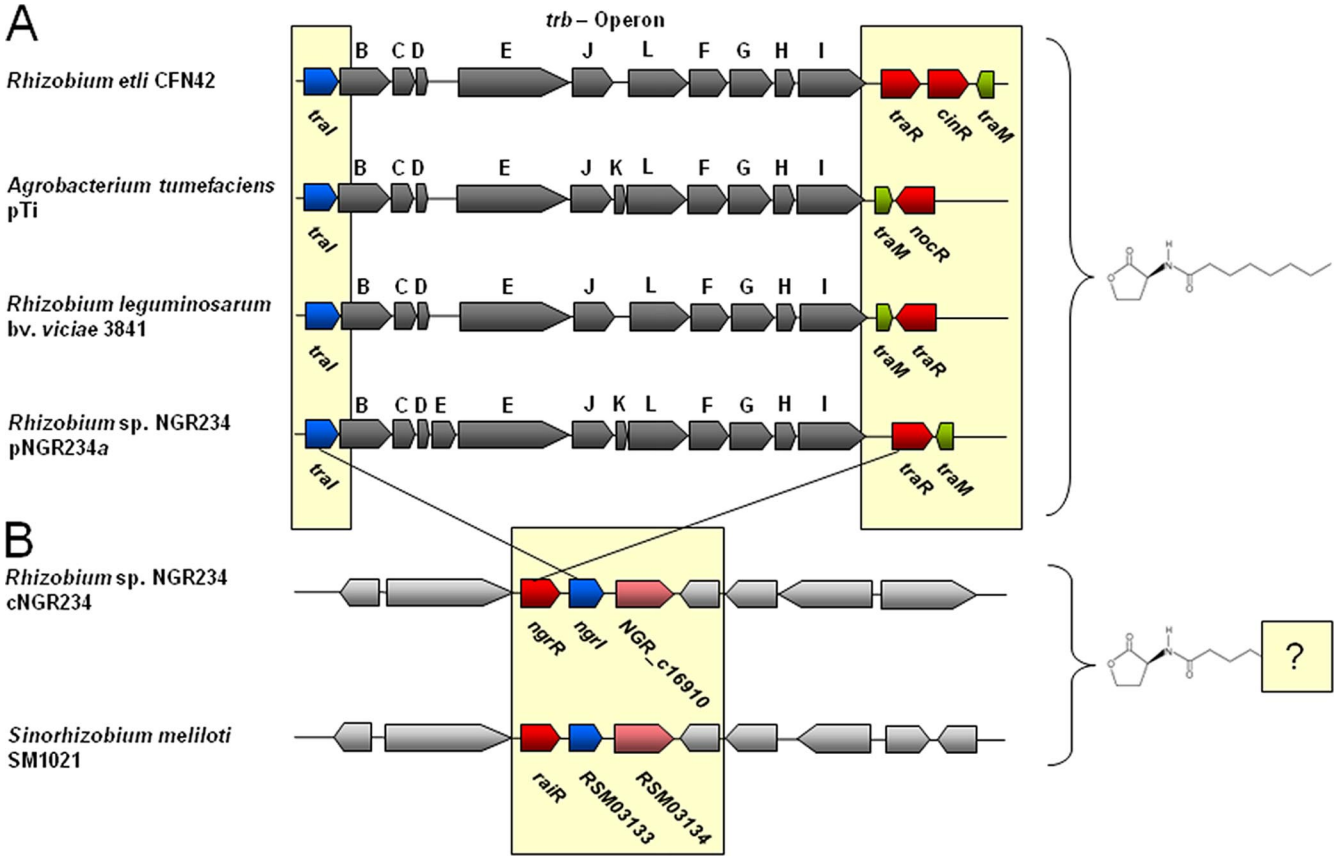

FIG. 2. Conserved clusters of genes linked to the synthesis of AHL-based quorum sensing molecules. (A) Comparison of the TraI/TraR system, together with conserved clusters of genes identified in other rhizobial species. The AHL synthase TraI (NGR_a04220, blue) directs the synthesis of 3-oxo-C8 homoserine lactone, which associates with the response regulator TraR (NGR_a04090, red) and activates transcription. TraM (NGR_a04080, green) functions as a suppressor, preventing TraR from activating target genes under noninducing conditions. The *trb* genes (*trbB* to *trbI*) (NGR_a04210 to NGR_a04100) involved in the conjugal plasmid transfer are shaded dark gray. (B) An additional quorum-sensing system identified on the *Rhizobium* sp. strain NGR234 chromosome, composed of NgrI (NGR_c16900, blue)/NgrR (NGR_c16890, red) (LuxI/LuxR homologs) and a hypothetical protein (NGR_c16910, light red), possibly linked to quorum sensing and autoinducer synthesis, is also shown. The yellow box highlights the regions with conserved gene organization.

involved in their export (*rkpR*, *rkpS*, and *rkpT*) and polymerization (*rkpO* and *rkpZ*) (46, 54). On the other hand, neither *rkpP*, which is probably involved in the acylation of the K_R5 antigen of Rm41, nor 3-hydroxybutyrate substitutions of the K antigen are present or occur in NGR234 (54).

NGR234 encodes many secretion-related proteins. Proteins secreted by bacteria play an important role in the infection of eukaryotes. In *R. leguminosarum* biovar viciae strain 3841, NodO is secreted via a type I secretion system (T1SS) encoded by *prsDE* (19, 28, 78). Mutation of *tatC* that encodes a Sec-independent protein translocase in *R. leguminosarum*, produces only white, non-nitrogen-fixing nodules when inoculated onto *Pisum sativum* (48). Sequencing pNGR234*a* (29) revealed a type III secretion system (T3SS-I) that affects nodulation of many hosts (2, 57, 82, 90). Since then, various symbiotic protein secretion systems have been found in diverse rhizobia (13). A detailed inventory of these systems in NGR234 was thus called for and is presented below and in Table 3. Individual ORFs linked to secretion systems are summarized in Table S2 in the supplemental material.

(i) Type I secretion genes. T1SSs comprise three proteins that transport targeted proteins across both bacterial membranes to the extracellular space. An ATPase of the ATP-binding cassette (ABC) family, which spans the periplasm, thus linking the inner and outer membranes, is needed, together with an outer membrane protein (18). Both the chromosome and the megaplasmid (but not the symbiotic plasmid) carry T1SS homologues, including NGR_c30050, NGR_c30060, NGR_c30070, *aprD* (NGR_b10690), *aprE* (NGR_b10700), and the TolC-like protein (NGR_c13520). Copies of T1SS are also present in most other rhizobial species (Table 3).

(ii) Type II-linked protein secretion systems: general secretion pathway, general export pathway, twin arginine pathway, and SRP. Type II secretion systems (T2SSs), which are broadly conserved in gram-negative bacteria, translocate exoproteins (e.g., cellulases, lipases, etc.) from the bacterial periplasm into the surrounding media (56) and are encoded by a set of 12 to 16 proteins named GspA to GspS. In many ways, T2SSs resemble type IV pilus (T4P) assembly systems (26, 56). In NGR234 the *gsp* genes required for type II pilus assembly are organized in one large chromosomal cluster containing 13 ORFs (see Table S2 in the supplemental material). This cluster stretches from NGR_c22980 (a hypothetical protein that is unique to NGR234) to *gspD* (NGR_c23100). Since the NGR234 genome encodes a

TABLE 3. Genes and proteins involved in the synthesis of secretion systems in selected rhizobial genomes[a]

Secretion system and characteristic	No. of ORFs and genes linked to the formation of secretion systems in strain[b]:					
	NGR234	3841	1021	110*	42	99
Type I						
AprD/E	2	8	5	2	2	4
TolC	1		1	1		1
PrtD	1	1	1		1	1
PrtE, HlyD family	1	1	1	1		1
Type I protein	1	1	1			
Type II						
Gsp (general secretion pathway)	13			8		11
Sec pathway	7	12	5	5	7	7
Type III (Hrp, Rhc)						
Rhc pili	42			21	21	21
Type IV						
F-type (conjugation, Vir, Trb)	23	10	10	13	31	19
P-type (Flp and attachment)	35	10	17	11	10	9
Type V						
Autotransporter (Aut)		3	2	2		1
TAT (twin arginine), TatA/B/C	3	3	3	3	3	3
SRP	3	3	3	3	3	3
Total	132	52	49	70	78	81

[a] Data for the analysis were extracted from the respective genome projects and the corresponding GenBank files. Column subheadings: NGR234, *Rhizobium* sp. strain NGR234; 3841, *R. leguminosarum* bv. viciae 3841 (48, 94); 1021, *S. meliloti* strain 1021 (30); 110, *B. japonicum* strain USDA110 (44); 42, *R. etli* strain CFN42 (32); 99, *M. loti* strain MAFF303099 (43).

[b] *, *B. japonicum* sp. strain BTAi1 (NC_009485 and NC_009475) and *B. japonicum* sp. strain ORS278 (NC_009445) each carry 21 *gsp* genes (31). The data indicate the absolute number of ORFs identified in the genome linked to a particular secretion system. Genes identified in NGR234 are listed in Table S2 in the supplemental material, together with the respective ORF designations. Flagellum genes were not included in the analysis, and no ncRNAs were included in the SRP analysis.

number of exoproteins, it is likely that the T2SS is functional. Interestingly, *gsp* genes seem to be confined to broad-host-range rhizobia. A similar gene cluster (containing 11 genes and an inversion) is found in *M. loti* (Fig. 3). *B. japonicum* strain USDA110 carries a truncated cluster of eight genes, while the two *B. japonicum* strains BTAi1 and ORS278 each carry two complete *gsp* clusters (Table 3). The narrow-host-range rhizobia *R. etli* CFN42, *R. leguminosarum* bv. viciae 3841, and strain 1021 do not possess *gsp* genes (Table 3).

Transport of exported proteins from the cytoplasm into the periplasm however, is usually managed by the general export pathway, GEP (Sec) and/or the twin arginine translocase (TAT) pathway. While the Sec pathway exports unfolded proteins, the TAT pathway is believed to export only folded proteins. At least three proteins, encoded by *tatABC* are required for TAT-dependent transport (14). ORFs (NGR_c13710, NGR_c13720, and NGR_c13730) corresponding to these genes were identified in a conserved cluster on the NGR234 chromosome (Fig. 1). Sec-dependent transport requires at least five proteins, the secretion-dedicated chaperone SecB, the Sec translocase (a multimeric membrane protein complex composed of a highly conserved protein-conducting channel comprising SecYEG), and a peripherally bound ribosome or ATP-dependent motor protein SecA (17). All are encoded by the NGR234 chromosome: *secA* (NGR_c26720), *secB* (NGR_c33550), *secD* (NGR_c02010 and NGR_c13810), *secE* (NGR_c11760), *secD/F* (NGR_c13810), and *secY* (NGRc_12100) (Table 3; see also Table S2 in the supplemental material).

The signal recognition particle (SRP) mediates membrane targeting of translating ribosomes displaying a signal anchor sequence. In *E. coli*, SRP consists of 4.5S RNA and a protein, Ffh, that recognizes the signal peptide emerging from the ribosome and the SRP receptor at the membrane, FtsY (55). The SRP-docking protein is encoded by FtsY (NGR_c32300) and the possible Ffh protein by NGR_c32250. The signal peptidase I is encoded by NGR_c08280. Altogether, three ncRNA genes were linked to SRP: NGR_c22640 encoded 4.5S RNA, while NGR_b21180 and NGR_c05600 were also linked to the SRP (see Table S2 in the supplemental material).

(iii) Type III and type IV protein secretion systems. Many gram-negative bacteria use specialized secretion machines to direct effector proteins into the cytoplasm of their eukaryotic hosts. In both animal and plant pathogens these secretion systems are important components of bacterial virulence. First identified in NGR234 (29), gene clusters encoding the major and conserved components of T3SSs have subsequently been found in diverse and distantly related rhizobia, including *B. japonicum* strains USDA110, BTAi1, and ORS278 (31, 44); *Mesorhizobium loti* MAFF303099 (43); and *S. fredii* strain USDA257 (50) (Table 3). None were identified in strain 1021 (30) and *R. leguminosarum* bv. viciae strain 3841 (94), however.

The flavonoid-dependent and NodD1-SyrM1-NodD2-TtsI-dependent regulatory cascade (47, 57) controls the activity of the T3SS-I locus (90) and modulates the nodulation of many hosts (2, 57, 82, 90). Surprisingly, pNGR234*b* carries a second T3SS cluster (T3SS-II) comprising 22 genes (NGR_b22800 to NGR_b23010) (see Table S2 in the supplemental material). Analysis of the promoter regions that control expression of the T3SS-II cluster in NGR234 failed to detect symbiotic regulatory elements such as *nod* or *ttsI* boxes, making it difficult to predict the role and regulation of these genes. Using another approach to test whether this cluster is symbiotically active, a polar mutant called NGRΔT3SS-II was constructed by deleting NGR_b22890 to NGR_b22950. This locus is predicted to encode five conserved components of a type III secretion machine, as well as two hypothetical proteins, and was replaced with the Omega-Km^r interposon. On several hosts of NGR234 tested, significant differences were not found between plants inoculated with NGR234 or NGRΔT3SS-II (3).

Type IV-related transport systems have been identified in many plant-associated microbes and the best studied system is the conjugative pilus (T4SS) of *Agrobacterium tumefaciens.* Most T4SS are formed by 12 proteins, VirB1 to VirB11, along with VirD4. pNGR234*b* carries the complete set of *virB1* to *virB11* genes (NGR_b10250 to NGR_b10360) next to the *traA-G* genes (NGR_b10550 to NGR_b10520) (Fig. 1). In strain 1021, a similar locus was shown to be required for the conjugation and transfer of pSymA but is not involved in symbiosis (42). This indicates that *virB1 to virB11* are probably part of a conjugation system and that pNGR234*b* might have evolved on a transferable plasmid. *virD* and *virE* are not present, however, suggesting that this system does not transfer DNA to plants. Interestingly, pNGR234*b* might be transmissi-

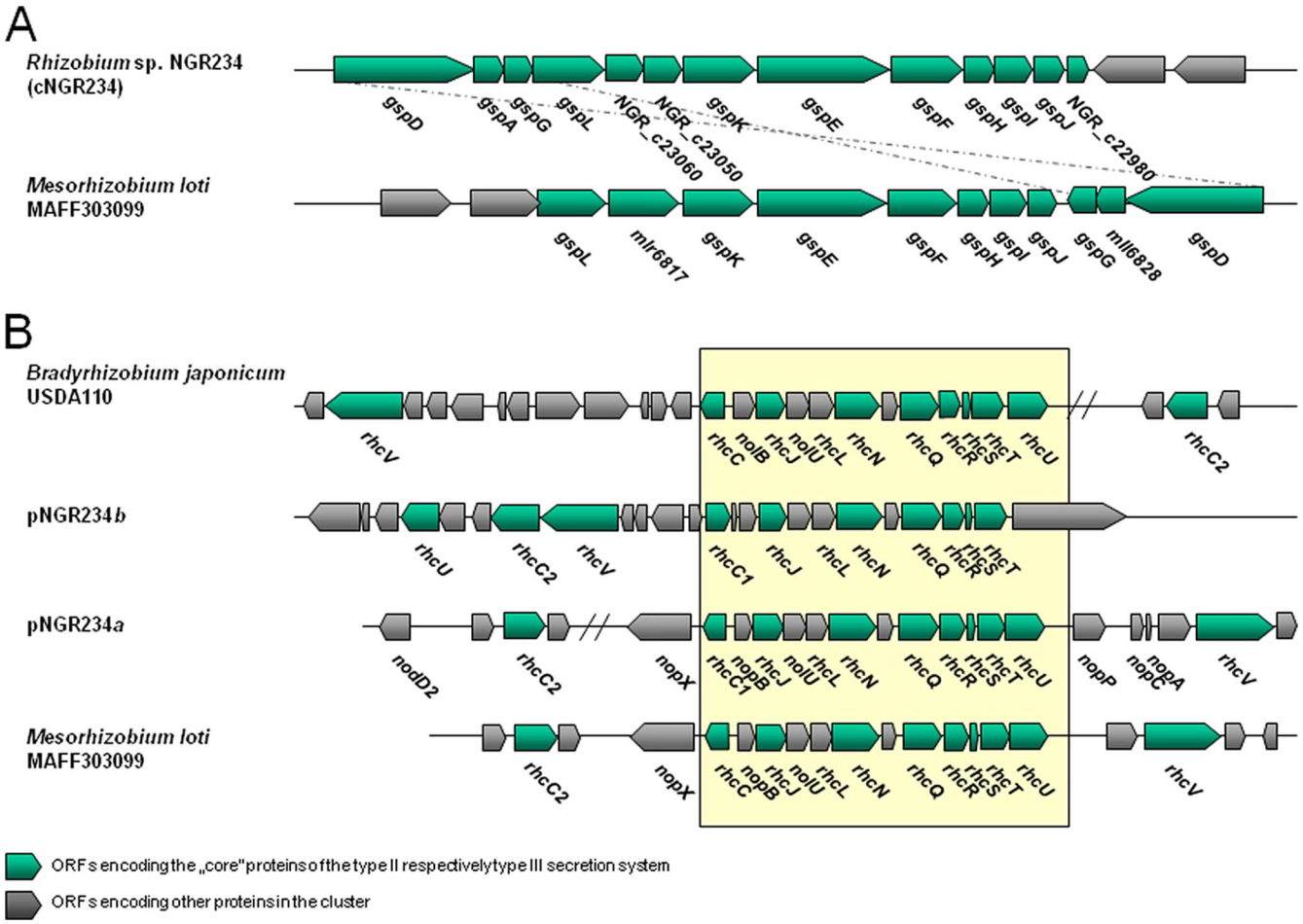

FIG. 3. Physical maps of the gene clusters of the single type II secretion systems (A), as well as the two copies of the type III pili (B) that have been identified in NGR234 and related bacteria. The yellow box highlights the regions with conserved gene organizations.

ble since a putative *oriT* region is located within *traA-G*. Since pNGR234*a* was also shown to be transferable via conjugation (29), both plasmids of NGR234 are apparently capable of being transferred to recipient strains.

Two clusters of 9 (NGR_b03670 to NGR_b03750) and 13 (NGR_b10770 to NGR_b10890) genes, plus the *tadG* locus, which is located elsewhere in pNGR234*b* (NGR_b18990), that encode polar T4P are present on pNGR234*b*, along with another cluster of 12 genes on the chromosome (NGR_c34610 to NGR_c34720) (Fig. 1). Such pili are involved in motility, attachment to surfaces, biofilm formation, twitching motility, and virulence (58, 89). Interestingly, the content and order of the 42-kb region encompassing T4P, T4SS, and *traA-G* genes on pNGR234*b* are almost identical to that on pSymA of strain 1021.

(iv) Type V and VI transporters. After being transported to the periplasm, proteins secreted via type V secretion systems (T5SS; also termed autotransporters), find their own way across the outer membrane. Using the AutA (RL1927), AutB (RL1196) and the AutC (RL1069) proteins from *R. leguminosarum* to query the NGR234 sequence did not reveal T5SS homologues, suggesting that this pathway is probably absent in NGR234. Similarly, type VI transport systems were not found. Nevertheless, 132 genes spread over the three replicons are involved in some form of protein secretion in NGR234.

NGR234-specific genes and possible links to host range. According to the Gold genome database (http://www.genomesonline.org/), almost 5,000 microbial genomes have either been sequenced or their sequences are being established. Compared to the 17 entries for *E. coli* genomes, only 12 of the ~1,000 completed genomes concern diverse rhizobia. The genome of *M. loti* MAFF303099 was the first to be deciphered (43), and since then those of *S. meliloti* strain 1021 (30), *B. japonicum* strains USDA110, BTAI1, and ORS278 (31, 44), *Azorhizobium caulinodans* ORS571 (52), *Mesorhizobium* sp. strain BNC1 (NC_008254), and *R. etli* strains CIAT652 (CP001074 [unpublished data]) and CFN42 (32), as well as *R. leguminosarum* bv. viciae strain 3841, have been completed (94) (NC_008380). The *Sinorhizobium medicae* genome has been analyzed (NC_009636) but not published. Even though the rhizobial genome data set is restricted, it is sufficient to permit certain comparisons.

To further determine NGR-specific genes, BiBlast searches were done using a subset of two published rhizobial genomes and the data obtained during the NGR genome project. BiBlast comparisons (each strain singly compared to the other two strains) of whole genomes of strain NGR234, strain 1021, and *R. etli* strain CFN42 suggests that they share a core genome of ~3,200 orthologous genes (Fig. 4), but that about 1,800 genes are unique to each strain. Strains 1021 and

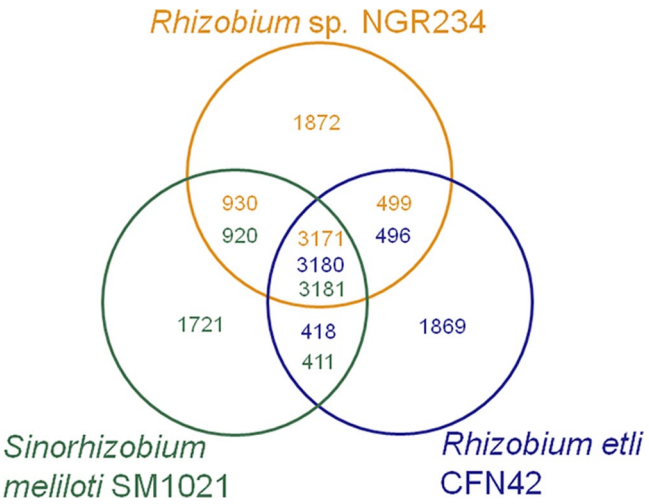

FIG. 4. BiBlast comparison (each single strain against the other two strains) of the complete genomes of strains NGR234, 1021, and CFN42. The Venn diagram shows the numbers of proteins shared or unique within a particular relationship for the three microbes. NGR-specific proteins are indicated in orange, strain 1021 proteins are indicated in green, and strain CFN42 proteins are indicated in blue. The three strains share a core genome of ~3,200 orthologous genes; all genomes contain about 1,800 genes that are unique for each strain. The 930 putative orthologous genes shared between strains 1021 and NGR234 (but which are absent in *R. etli*) are compared to 499 orthologues shared by *R. etli* and NGR234 (but which are absent in strain 1021) indicate that strain NGR234 is more closely related to strain 1021.

NGR234 share ~930 putative orthologous genes that are absent in *R. etli*. In turn, *R. etli* and NGR234 share 499 putative orthologous genes that are not present in strain 1021. In this sense, NGR234 is more closely related to strain 1021 but is also significantly different from it. This statement is also supported by a phylogenetic comparison based on the 16S rRNA gene diversity (see Fig. S4 in the supplemental material).

The distribution of putative orthologous versus nonorthologous genes on the different NGR234 replicons shows a clear bias. On the chromosome many of the NGR234-specific genes are located in regions which differ in their GC contents and contain significantly less repetitive elements compared to conserved parts of the chromosome. In contrast, pNGR234*b* is a patchwork plasmid that consists of regions which share partial blocks of orthologous genes that are similar to the different replicons of strain 1021. These concern especially the regions carrying genes for the T4 secretion systems. T4P and the T4SS systems share most orthologues to pSymB and the strain 1021 chromosome. The absolute number of shared orthologous genes is, however, highest between the strain 1021 pSymA and pNGR234*b* (Fig. 1).

Another important observation is that correlations seem to exist between the host range of rhizobia and the number of specialized protein secretion systems they carry. Classic narrow-host-range rhizobia such as *S. meliloti* and *R. leguminosarum* that nodulate a restricted group of Middle-Eastern and Northern European plants carry neither T3 nor T4 secretion systems. At the other extreme, NGR234 which has the greatest capacity to nodulate of all known rhizobia, carries almost twice as many, often duplicated, secretion systems (Table 3). Nonsymbiotic bacteria possess even fewer: *E. coli* K-12 only carries about 35 genes linked to secretion systems (not including the flagellum genes), the opportunistic pathogen *Pseudomonas aeruginosa* PA01 encodes ~80 genes and *Burkholderia thailandensis* about 90 genes involved in protein secretion.

Although formal proof that all of the many, diverse protein secretion systems extend the host range of NGR234 has yet to be furnished, another piece in the puzzle which began with the characterization of Nod-NGR234 factors seems to be falling into place. NGR234 secretes a large family of lipo-chito-oligosaccharidic Nod factors that are variously 3-O, 4-O, or 6-O carbamoylated, that are N methylated, and that carry a 2-*O*-methyl-fucose residue that may be either 3-O sulfated or 4-O acetylated (68). Since no other rhizobia synthesize such a large family of these lipo-oligo-saccharides that prepare the legume for nodulation (42) and allow the rhizobia to penetrate root hairs (71, 73), we speculated that Nod-factors themselves contribute to the broad host range of NGR234 (7, 66, 69). A second piece of the puzzle fell into place when we showed that NGR234 not only treats the legume root to a large palette of Nod factors but that their concentration is much higher than even very closely related rhizobia (72). Moreover, while we have not yet tested the effects of all protein secretion systems on the nodulation of NGR234 hosts, it is clear that the T3SS-I locus and the effectors it secretes have profound host range effects (13).

The diverse catabolic and metabolic functions encoded by NGR234 are another piece of the puzzle. Rhizobia have two lifestyles: one saprophytic and the other symbiotic and intracellular. A soil bacterium that is able to survive and grow under a wide range of nutritional conditions has more chances of surviving than strains that have very specific growth requirements. More importantly perhaps, to be primed for the invasion of legumes, rhizobia have to grow in the rhizosphere of approaching root systems. Bacteria that can metabolize virtually any carbon- and nitrogen-containing compound that emanates from plant roots will preferentially colonize them. Once established within the rhizosphere, NodD1 of NGR234 senses a wide variety of phenolic substances (24, 25, 49, 65). The NodD1-flavonoid complex then either directly or indirectly activates genes downstream of the many *nod* and *tts* boxes present in NGR234 (47, 93). In this way, the family of Nod-NGR234 factors are released, the T3SS-I begins to secrete proteins, and the spectrum of NGR234 hosts that are nodulated is enlarged.

Although the complete genome sequence of NGR234 has not revealed all of the mysteries of broad host range, it probably lays out a plan that can lead to them. Many more pieces of the broad-host-range puzzle will probably fall into place once the NGR234's closest relative, *Rhizobium* (*Sinorhizobium*) *fredii* strain USDA257, is sequenced (64). In extensive nodulation tests, USDA257 nodulated an exact subset of the NGR234 hosts (69), suggesting that rather simple differences between the two genomes (e.g., the presence or absence of a functional *nodS*, which encodes an *N*-methyl transferase, and of *nodU*, which encodes a 6-*O*-carbamoylase [40]) will help explain the differences in host range.

ACKNOWLEDGMENTS

This study was supported by the Swiss National Science Foundation (grants 3100-67977, 3100-63893, and 3100A0-116591) and the University of Geneva. Research in Göttingen and Hamburg was funded by the Genomik-Plus Network of the BMBF.

We thank Isabelle Saint Girons, Dora Gerber, and William J. Deakin for their unstinting help with many aspects of this work.

REFERENCES

1. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
2. **Ausmees, N., H. Kobayashi, W. J. Deakin, C. Marie, H. B. Krishnan, W. J. Broughton, and X. Perret.** 2004. Characterization of NopP, a type III secreted effector of *Rhizobium* sp. strain NGR234. J. Bacteriol. **186:**4774–4780.
3. **Bakkou, N., and X. Perret.** 2008. Functional analysis of a second type III secretion system in *Rhizobium* sp. NGR234, p. 209. *In* M. Holsters (ed.), Abstract book of the Eighth European Nitrogen Fixation Conference, Ghent, Belgium.
4. **Becker, A., N. Fraysse, and L. Sharypova.** 2005. Recent advances in studies on structure and symbiosis-related function of rhizobial K-antigens and lipopolysaccharides. Mol. Plant-Microbe Interact. **18:**899–905.
5. **Becker, A., A. Kleickmann, H. Kuster, M. Keller, W. Arnold, and A. Pühler.** 1993. Analysis of the *Rhizobium meliloti* genes *exoU*, *exoV*, *exoW*, *exoT*, and *exoI* involved in exopolysaccharide biosynthesis and nodule invasion: *exoU* and *exoW* probably encode glucosyltransferases. Mol. Plant-Microbe Interact. **6:**735–744.
6. **Broughton, W. J., M. J. Dilworth, and I. K. Passmore.** 1972. Base ratio determination using unpurified DNA. Anal. Biochem. **46:**164–172.
7. **Broughton, W. J., S. Jabbouri, and X. Perret.** 2000. Keys to symbiotic harmony. J. Bacteriol. **182:**5641–5652.
8. **Charles, T. C., and T. M. Finan.** 1991. Analysis of a 1600-kilobase *Rhizobium meliloti* megaplasmid using defined deletions generated in vivo. Genetics **127:**5–20.
9. **Chen, H., J. X. Gray, M. Nayudu, M. A. Djordjevic, M. Batley, J. W. Redmond, and B. G. Rolfe.** 1988. Five genetic loci involved in the synthesis of acidic exopolysaccharides are closely linked in the genome of *Rhizobium* sp. strain NGR234. Mol. Gen. Genet. **212:**310–316.
10. **Cheng, J., C. D. Sibley, R. Zaheer, and T. M. Finan.** 2007. A *Sinorhizobium meliloti minE* mutant has an altered morphology and exhibits defects in legume symbiosis. Microbiology **153:**375–387.
11. **Cooper, J. E., and J. A. Callow.** 2004. Multiple responses of rhizobia to flavonoids during legume root infection. Adv. Botan. Res. **41:**1–62.
12. **Cronan, J. E., and G. L. Waldrop.** 2002. Multi-subunit acetyl-CoA carboxylases. Prog. Lipid Res. **41:**407–435.
13. **Deakin, W. J., and W. J. Broughton.** 2009. Symbiotic use of pathogenic strategies: rhizobial protein secretion systems. Nat. Rev. Microbiol. **7:**312–320.
14. **De Buck, E., E. Lammertyn, and J. Anné.** 2008. The importance of the twin-arginine translocation pathway for bacterial virulence. Trends Microbiol. **16:**442–453.
15. **Djordjevic, S. P., B. Rolfe, M. Batley, and J. W. Redmond.** 1986. The structure of the exopolysaccharide from *Rhizobium* sp. strain ANU280 (NGR234). Carbohydr. Res. **148:**87–99.
16. **Downie, J. A.** 2005. Legume haemoglobins: symbiotic nitrogen fixation needs bloody nodules. Curr. Biol. **15:**R196–R198.
17. **Driessen, A. J. M., and N. Nouwen.** 2008. Protein translocation across the bacterial cytoplasmic membrane. Annu. Rev. Biochem. **77:**643–667.
18. **Duong, F., E. Bonnet, V. Géli, A. Lazdunski, M. Murgier, and A. Filloux.** 2001. The AprX protein of *Pseudomonas aeruginosa*: a new substrate for the Apr type I secretion system. Gene **262:**147–153.
19. **Economou, A., W. D. Hamilton, A. W. Johnston, and J. A. Downie.** 1990. The *Rhizobium* nodulation gene *nodO* encodes a Ca^{2+}-binding protein that is exported without N-terminal cleavage and is homologous to haemolysin and related proteins. EMBO J. **9:**349–354.
20. **Eddy, S. R.** 2002. A memory-efficient dynamic programming algorithm for optimal alignment of a sequence to an RNA secondary structure. BMC Bioinform. **3:**18.
21. **Encarnacion, S., M. Dunn, K. Willms, and J. Mora.** 1995. Fermentative and aerobic metabolism in *Rhizobium etli*. J. Bacteriol. **177:**3058–3066.
22. **Encarnacion, S., Y. Guzman, M. F. Dunn, M. Hernandez, M. del Carmen Vargas, and J. Mora.** 2003. Proteome analysis of aerobic and fermentative metabolism in *Rhizobium etli* CE3. Proteomics **3:**1077–1085.
23. **Entcheva, P., W. Liebl, A. Johann, T. Hartsch, and W. R. Streit.** 2001. Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia. Appl. Environ. Microbiol. **67:**89–99.
24. **Fellay, R., M. Hanin, G. Montorzi, J. Frey, C. Freiberg, W. Golinowski, C. Staehelin, W. J. Broughton, and S. Jabbouri.** 1998. *nodD2* of *Rhizobium* sp. NGR234 is involved in the repression of the *nodABC* operon. Mol. Microbiol. **27:**1039–1050.
25. **Fellay, R., X. Perret, V. Viprey, W. J. Broughton, and S. Brenner.** 1995. Organization of host-inducible transcripts on the symbiotic plasmid of *Rhizobium* sp. NGR234. Mol. Microbiol. **16:**657–667.
26. **Filloux, A.** 2004. The underlying mechanisms of type II protein secretion. Biochim. Biophys. Acta **1694:**163–179.
27. **Finan, T. M., S. Weidner, K. Wong, J. Buhrmester, P. Chain, F. J. Vorholter, I. Hernandez-Lucas, A. Becker, A. Cowie, J. Gouzy, B. Golding, and A. Pühler.** 2001. The complete sequence of the 1,683-kb pSymB megaplasmid from the N_2-fixing endosymbiont *Sinorhizobium meliloti*. Proc. Natl. Acad. Sci. USA **98:**9889–9894.
28. **Finnie, C., N. M. Hartley, K. C. Findlay, and J. A. Downie.** 1997. The *Rhizobium leguminosarum prsDE* genes are required for secretion of several proteins, some of which influence nodulation, symbiotic nitrogen fixation and exopolysaccharide modification. Mol. Microbiol. **25:**135–146.
29. **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.
30. **Galibert, F., T. M. Finan, S. R. Long, A. Pühler, P. Abola, F. Ampe, F. Barloy-Hubler, M. J. Barnett, A. Becker, P. Boistard, G. Bothe, M. Boutry, L. Bowser, J. Buhrmester, E. Cadieu, D. Capela, P. Chain, A. Cowie, R. W. Davis, S. Dreano, N. A. Federspiel, R. F. Fisher, S. Gloux, T. Godrie, A. Goffeau, B. Golding, J. Gouzy, M. Gurjal, I. Hernandez-Lucas, A. Hong, L. Huizar, R. W. Hyman, T. Jones, D. Kahn, M. L. Kahn, S. Kalman, D. H. Keating, E. Kiss, C. Komp, V. Lelaure, D. Masuy, C. Palm, M. C. Peck, T. M. Pohl, D. Portetelle, B. Purnelle, U. Ramsperger, R. Surzycki, P. Thebault, M. Vandenbol, F. J. Vorholter, S. Weidner, D. H. Wells, K. Wong, K. C. Yeh, and J. Batut.** 2001. The composite genome of the legume symbiont *Sinorhizobium meliloti*. Science **293:**668–672.
31. **Giraud, E., L. Moulin, D. Vallenet, V. Barbe, E. Cytryn, J.-C. Avarre, M. Jaubert, D. Simon, F. Cartieaux, Y. Prin, G. Bena, L. Hannibal, J. Fardoux, M. Kojadinovic, L. Vuillet, A. Lajus, S. Cruveiller, Z. Rouy, S. Mangenot, B. Segurens, C. Dossat, W. L. Franck, W.-S. Chang, E. Saunders, D. Bruce, P. Richardson, P. Normand, B. Dreyfus, D. Pignol, G. Stacey, D. Emerich, A. Vermeglio, C. Medigue, and M. Sadowsky.** 2007. Legumes symbioses: absence of *Nod* genes in photosynthetic *Bradyrhizobia*. Science **316:**1307–1312.
32. **Gonzalez, V., R. I. Santamaria, P. Bustos, I. Hernandez-Gonzalez, A. Medrano-Soto, G. Moreno-Hagelsieb, S. C. Janga, M. A. Ramirez, V. Jimenez-Jacinto, J. Collado-Vides, and G. Davila.** 2006. The partitioned *Rhizobium etli* genome: genetic and metabolic redundancy in seven interacting replicons. Proc. Natl. Acad. Sci. USA **103:**3834–3839.
33. **Griffiths-Jones, S., S. Moxon, M. Marshall, A. Khanna, S. R. Eddy, and A. Bateman.** 2005. Rfam: annotating non-coding RNAs in complete genomes. Nucleic Acids Res. **33:**D121–D124.
34. **Haydon, D. J., and J. R. Guest.** 1991. A new family of bacterial regulatory proteins. FEMS Microbiol. Lett. **63:**291–295.
35. **He, X., W. Chang, D. L. Pierce, L. O. Seib, J. Wagner, and C. Fuqua.** 2003. Quorum sensing in *Rhizobium* sp. strain NGR234 regulates conjugal transfer (*tra*) gene expression and influences growth rate. J. Bacteriol. **185:**809–822.
36. **Hebbeln, P., D. A. Rodionov, A. Alfandega, and T. Eitinger.** 2007. Biotin uptake in prokaryotes by solute transporters with an optional ATP-binding cassette-containing module. Proc. Natl. Acad. Sci. USA **104:**2909–2914.
37. **Heinz, E. B., D. A. Phillips, and W. R. Streit.** 1999. BioS, a biotin-induced, stationary-phase, and possible LysR-type regulator in *Sinorhizobium meliloti*. Mol. Plant-Microbe Interact. **12:**803–812.
38. **Heinz, E. B., and W. R. Streit.** 2003. Biotin limitation in *Sinorhizobium meliloti* strain 1021 alters transcription and translation. Appl. Environ. Microbiol. **69:**1206–1213.
39. **Hopwood, D. A., and D. H. Sherman.** 1990. Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. Annu. Rev. Genet. **24:**37–62.
40. **Jabbouri, S., R. Fellay, F. Talmont, P. Kamalaprija, U. Burger, B. Relić, J.-C. Promé, and W. J. Broughton.** 1995. Involvement of *nodS* in N-methylation and *nodU* in 6-*O*-carbamoylation of *Rhizobium* sp. NGR234 Nod factors. J. Biol. Chem. **270:**22968–22973.
41. **Jones, D. L., J. R. Healey, V. B. Willett, J. F. Farrar, and A. Hodge.** 2005. Dissolved organic nitrogen uptake by plants: an important N uptake pathway? Soil Biol. Biochem. **37:**413–423.
42. **Jones, K. M., J. Lloret, J. R. Daniele, and G. C. Walker.** 2007. The type IV secretion system of *Sinorhizobium meliloti* strain 1021 is required for conjugation but not for intracellular symbiosis. J. Bacteriol. **189:**2133–2138.
43. **Kaneko, T., Y. Nakamura, S. Sato, E. Asamizu, T. Kato, S. Sasamoto, A. Watanabe, K. Idesawa, A. Ishikawa, K. Kawashima, T. Kimura, Y. Kishida, C. Kiyokawa, M. Kohara, M. Matsumoto, A. Matsuno, Y. Mochizuki, S. Nakayama, N. Nakazaki, S. Shimpo, M. Sugimoto, C. Takeuchi, M. Yamada, and S. Tabata.** 2000. Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. DNA Res. **7:**331–338.
44. **Kaneko, T., Y. Nakamura, S. Sato, K. Minamisawa, T. Uchiumi, S. Sasamoto, A. Watanabe, K. Idesawa, M. Iriguchi, K. Kawashima, M. Kohara, M. Matsumoto, S. Shimpo, H. Tsuruoka, T. Wada, M. Yamada, and S. Tabata.** 2002. Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. DNA Res. **9:**189–197.
45. **Kannenberg, E. L., B. L. Reuhs, L. S. Forsberg, and R. W. Carlson.** 1998. Lipopolysaccharides and K-antigens, p. 119–154. *In* H. P. Spaink, A. Kondorosi, and P. J. Hooykaas, (ed.), The *Rhizobiaceae*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
46. **Kiss, E., A. Kereszt, F. Barta, S. Stephens, B. L. Reuhs, A. Kondorosi, and P. Putnoky.** 2001. The *rkp-3* gene region of *Sinorhizobium meliloti* Rm41 contains strain-specific genes that determine K antigen structure. Mol. Plant-Microbe Interact. **14:**1395–1403.
47. **Kobayashi, H., Y. Naciri-Graven, W. J. Broughton, and X. Perret.** 2004. Flavonoids induce temporal shifts in gene expression of *nod*-box controlled loci in *Rhizobium* sp. NGR234. Mol. Microbiol. **51:**335–347.
48. **Krehenbrink, M., and J. A. Downie.** 2008. Identification of protein secretion

systems and novel secreted proteins in *Rhizobium leguminosarum* bv. *viciae*. BMC Genomics **9:**55.

49. **Krishnan, H. B., A. Lewin, R. Fellay, W. J. Broughton, and S. G. Pueppke.** 1992. Differential expression of *nodS* accounts for the varied abilities of *Rhizobium fredii* USDA257 and *Rhizobium* sp. strain NGR234 to nodulate *Leucaena* spp. Mol. Microbiol. **6:**3321–3330.
50. **Krishnan, H. B., J. Lorio, W. S. Kim, G. Jiang, K. Y. Kim, M. DeBoer, and S. G. Pueppke.** 2003. Extracellular proteins involved in soybean cultivar-specific nodulation are associated with pilus-like surface appendages and exported by a type III protein secretion system in *Sinorhizobium fredii* USDA257. Mol. Plant-Microbe Interact. **16:**617–625.
51. **Krysciak, D., A. Strittmatter, H. Liesegang, C. Schmeisser, R. Schmitz, G. Gottschalk, W. J. Broughton, X. Perret, and W. Streit.** 2008. The complete genome sequence of *Rhizobium* sp. NGR234: a large gene pool for secretion and communication mechanisms, p. 217. *In* M. Holsters (ed.), Abstract book of the Eighth European Nitrogen Fixation Conference, Ghent, Belgium.
52. **Lee, K. B., P. De Backer, T. Aono, C. T. Liu, S. Suzuki, T. Suzuki, T. Kaneko, M. Yamada, S. Tabata, D. M. Kupfer, F. Z. Najar, G. B. Wiley, B. Roe, T. T. Binnewies, D. W. Ussery, W. D'Haeze, J. D. Herder, D. Gevers, D. Vereecke, M. Holsters, and H. Oyaizu.** 2008. The genome of the versatile nitrogen fixer *Azorhizobium caulinodans* ORS571. BMC Genomics **9:**271.
53. **Lepo, J. E., F. J. Hanus, and H. J. Evans.** 1980. Chemoautotrophic growth of hydrogen-uptake-positive strains of *Rhizobium japonicum*. J. Bacteriol. **141:**664–670.
54. **Le Quéré, A. J., W. J. Deakin, C. Schmeisser, R. W. Carlson, W. R. Streit, W. J. Broughton, and L. S. Forsberg.** 2006. Structural characterization of a K-antigen capsular polysaccharide essential for normal symbiotic infection in *Rhizobium* sp. NGR234: deletion of the *rkpMNO* locus prevents synthesis of 5,7-diacetamido-3,5,7,9-tetradeoxy-non-2-ulosonic acid. J. Biol. Chem. **281:**28981–28992.
55. **Luirink, J., and I. Sinning.** 2004. SRP-mediated protein targeting: structure and function revisited. Biochim. Biophys. Acta Mol. Cell Res. **1694:**17–35.
56. **Maria, S.** 2001. Biology of type II secretion. Mol. Microbiol. **40:**271–283.
57. **Marie, C., W. J. Deakin, T. Ojanen-Reuhs, E. Diallo, B. Reuhs, W. J. Broughton, and X. Perret.** 2004. TtsI, a key regulator of *Rhizobium* species NGR234 is required for type III-dependent protein secretion and synthesis of rhamnose-rich polysaccharides. Mol. Plant-Microbe Interact. **17:**958–966.
58. **Mattick, J. S.** 2002. Type IV pili and twitching motility. Annu. Rev. Microbiol. **56:**289–314.
59. **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.
60. **Oresnik, I. J., S. L. Liu, C. K. Yost, and M. F. Hynes.** 2000. Megaplasmid pRme2011a of *Sinorhizobium meliloti* is not required for viability. J. Bacteriol. **182:**3582–3586.
61. **Overbeek, R., N. Larsen, T. Walunas, M. D'Souza, G. Pusch, E. Selkov, Jr., K. Liolios, V. Joukov, D. Kaznadzey, I. Anderson, A. Bhattacharyya, H. Burd, W. Gardner, P. Hanke, V. Kapatral, N. Mikhailova, O. Vasieva, A. Osterman, V. Vonstein, M. Fonstein, N. Ivanova, and N. Kyrpides.** 2003. The ERGO genome analysis and discovery system. Nucleic Acids Res. **31:**164–171.
62. **Park, S.-Y., S. J. Lee, T.-K. Oh, J.-W. Oh, B.-T. Koo, D.-Y. Yum, and J.-K. Lee.** 2003. AhlD, an *N*-acylhomoserine lactonase in *Arthrobacter* sp., and predicted homologues in other bacteria. Microbiology **149:**1541–1550.
63. **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.
64. **Perret, X., R. Fellay, A. J. Bjourson, J. E. Cooper, S. Brenner, and W. J. Broughton.** 1994. Subtraction hybridisation and shot-gun sequencing: a new approach to identify symbiotic loci. Nucleic Acids Res. **22:**1335–1341.
65. **Perret, X., C. Freiberg, A. Rosenthal, W. J. Broughton, and R. Fellay.** 1999. High-resolution transcriptional analysis of the symbiotic plasmid of *Rhizobium* sp. NGR234. Mol. Microbiol. **32:**415–425.
66. **Perret, X., C. Staehelin, and W. J. Broughton.** 2000. Molecular basis of symbiotic promiscuity. Microbiol. Mol. Biol. Rev. **64:**180–201.
67. **Perret, X., V. Viprey, C. Freiberg, and W. J. Broughton.** 1997. Structure and evolution of NGRRS-1, a complex, repeated element in the genome of *Rhizobium* sp. strain NGR234. J. Bacteriol. **179:**7488–7496.
68. **Price, N. P., B. Relić, F. Talmont, A. Lewin, D. Prome, S. G. Pueppke, F. Maillet, J. Dénarié, J.-C. Promé, and W. J. Broughton.** 1992. Broad-host-range *Rhizobium* species strain NGR234 secretes a family of carbamoylated, and fucosylated, nodulation signals that are O-acetylated or sulphated. Mol. Microbiol. **6:**3575–3584.
69. **Pueppke, S. G., and W. J. Broughton.** 1999. *Rhizobium* sp. strain NGR234 and *R. fredii* USDA257 share exceptionally broad, nested host ranges. Mol. Plant-Microbe Interact. **12:**293–318.
70. **Ramos, J. L., M. Martinez-Bueno, A. J. Molina-Henares, W. Teran, K. Watanabe, X. Zhang, M. T. Gallegos, R. Brennan, and R. Tobes.** 2005. The TetR family of transcriptional repressors. Microbiol. Mol. Biol. Rev. **69:**326–356.
71. **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.
72. **Relić, B., C. Staehelin, R. Fellay, S. Jabbouri, T. Boller, and W. J. Broughton.** 1994. Do Nod-factor levels play a role in host specificity?, p. 69–75. *In* G. B. Kiss and G. Endre (ed.), Proceedings of the First European Congress on Nitrogen Fixation. Officina Press Szed, Szeged, Hungary.
73. **Relić, B., F. Talmont, J. Kopcinska, W. Golinowski, J.-C. Promé, and W. J. Broughton.** 1993. Biological activity of *Rhizobium* sp. NGR234 Nod-factors on *Macroptilium atropurpureum*. Mol. Plant-Microbe Interact. **6:**764–774.
74. **Reuhs, B. L., R. W. Carlson, and J. S. Kim.** 1993. *Rhizobium fredii* and *Rhizobium meliloti* produce 3-deoxy-D-manno-2-octulosonic acid-containing polysaccharides that are structurally analogous to group II K antigens (capsular polysaccharides) found in *Escherichia coli*. J. Bacteriol. **175:**3570–3580.
75. **Reuhs, B. L., D. P. Geller, J. S. Kim, J. E. Fox, V. S. K. Kolli, and S. G. Pueppke.** 1998. *Sinorhizobium fredii* and *Sinorhizobium meliloti* produce structurally conserved lipopolysaccharides and strain-specific K antigens. Appl. Environ. Microbiol. **64:**4930–4938.
76. **Sambrook, J., and D. W. Russell.** 2001. Molecular cloning: a laboratory manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
77. **Schell, M. A.** 1993. Molecular biology of the LysR family of transcriptional regulators. Annu. Rev. Microbiol. **47:**597–626.
78. **Scheu, A. K., A. Economou, G. F. Hong, S. Ghelani, A. W. Johnston, and J. A. Downie.** 1992. Secretion of the *Rhizobium leguminosarum* nodulation protein NodO by haemolysin-type systems. Mol. Microbiol. **6:**231–238.
79. **Schiex, T., J. Gouzy, A. Moisan, and Y. de Oliveira.** 2003. FrameD: a flexible program for quality check and gene prediction in prokaryotic genomes and noisy matured eukaryotic sequences. Nucleic Acids Res. **31:**3738–3741.
80. **Schipper, C., C. Hornung, P. Bijtenhoorn, M. Quitschau, S. Grond, and W. R. Streit.** 2009. Metagenome-derived clones encoding for two novel lactonase family proteins involved in biofilm inhibition in *Pseudomonas aeruginosa*. Appl. Environ. Microbiol. **75:**224–233.
81. **Sivanathan, V., M. D. Allen, C. de Bekker, R. Baker, L. K. Arciszewska, S. M. Freund, M. Bycroft, J. Lowe, and D. J. Sherratt.** 2006. The FtsK gamma domain directs oriented DNA translocation by interacting with KOPS. Nat. Struct. Mol. Biol. **13:**965–972.
82. **Skorpil, P., M. M. Saad, N. M. Boukli, H. Kobayashi, F. Ares-Orpel, W. J. Broughton, and W. J. Deakin.** 2005. NopP, a phosphorylated effector of *Rhizobium* sp. strain NGR234, is a major determinant of nodulation of the tropical legumes *Flemingia congesta* and *Tephrosia vogelii*. Mol. Microbiol. **57:**1304–1317.
83. **Skorupska, A., M. Janczarek, M. Marczak, A. Mazur, and J. Krol.** 2006. Rhizobial exopolysaccharides: genetic control and symbiotic functions. Microb. Cell Fact **5:**7.
84. **Staehelin, C., L. S. Forsberg, W. D'Haeze, M. Y. Gao, R. W. Carlson, Z. P. Xie, B. J. Pellock, K. M. Jones, G. C. Walker, W. R. Streit, and W. J. Broughton.** 2006. Exo-oligosaccharides of *Rhizobium* sp. strain NGR234 are required for symbiosis with various legumes. J. Bacteriol. **188:**6168–6178.
85. **Streit, W. R., C. M. Joseph, and D. A. Phillips.** 1996. Biotin and other water-soluble vitamins are key growth factors for alfalfa root colonization by *Rhizobium meliloti* 1021. Mol. Plant-Microbe Interact. **9:**330–338.
86. **Streit, W. R., R. A. Schmitz, X. Perret, C. Staehelin, W. J. Deakin, C. Raasch, H. Liesegang, and W. J. Broughton.** 2004. An evolutionary hot spot: the pNGR234*b* replicon of *Rhizobium* sp. strain NGR234. J. Bacteriol. **186:**535–542.
87. **Tech, M., and R. Merkl.** 2003. YACOP: enhanced gene prediction obtained by a combination of existing methods. In Silico Biol. **3:**441–451.
88. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673–4680.
89. **Tomich, M., P. J. Planet, and D. H. Figurski.** 2007. The *tad* locus: postcards from the widespread colonization island. Nat. Rev. Microbiol. **5:**363–375.
90. **Viprey, V., A. Del Greco, W. Golinowski, W. J. Broughton, and X. Perret.** 1998. Symbiotic implications of type III protein secretion machinery in *Rhizobium*. Mol. Microbiol. **28:**1381–1389.
91. **Viprey, V., A. Rosenthal, W. J. Broughton, and X. Perret.** 2000. Genetic snapshots of the *Rhizobium* species NGR234 genome. Genome Biol. **1:** RESEARCH0014.
92. **Washietl, S., I. L. Hofacker, and P. F. Stadler.** 2005. Fast and reliable prediction of noncoding RNAs. Proc. Natl. Acad. Sci. USA **102:**2454–2459.
93. **Wassem, R., H. Kobayashi, K. Kambara, A. J. Le Quéré, G. C. Walker, W. J. Broughton, and W. J. Deakin.** 2008. TtsI regulates symbiotic genes in *Rhizobium* species NGR234 by binding to *tts* boxes. Mol. Microbiol. **68:**736–748.
94. **Young, J. P., L. Crossman, A. Johnston, N. Thomson, Z. Ghazoui, K. Hull, M. Wexler, A. Curson, J. Todd, P. Poole, T. Mauchline, A. East, M. Quail, C. Churcher, C. Arrowsmith, I. Cherevach, T. Chillingworth, K. Clarke, A. Cronin, P. Davis, A. Fraser, Z. Hance, H. Hauser, K. Jagels, S. Moule, K. Mungall, H. Norbertczak, E. Rabbinowitsch, M. Sanders, M. Simmonds, S. Whitehead, and J. Parkhill.** 2006. The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. Genome Biol. **7:**R34.