Research

Genetic snapshots of the *Rhizobium* species NGR234 genome

Virginie Viprey*†, André Rosenthal‡, William J Broughton* and Xavier Perret*

Addresses: *Laboratoire de Biologie Moléculaire des Plantes Supérieures, Université de Genève, chemin de l'Impératrice, 1292 Chambésy, Genève, Switzerland. ‡Institut für Molekulare Biotechnologie, Abteilung Genomanalyze, Beutenbergstrasse, 07745 Jena, Germany. †Present address: John Innes Centre, Colney Lane, Norwich NR4 7UH, UK.

Correspondence:William J Broughton. E-mail: broughtw@sc2a.unige.ch

Published: 4 December 2000

Genome ***Biology*** 2000, **1(6)**:research0014.1–0014.17

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2000/1/6/research/0014

Received: 17 April 2000
Revised: 25 October 2000
Accepted: 31 October 2000

Abstract

Background: In nitrate-poor soils, many leguminous plants form nitrogen-fixing symbioses with members of the bacterial family Rhizobiaceae. We selected *Rhizobium* sp. NGR234 for its exceptionally broad host range, which includes more than 112 genera of legumes. Unlike the genome of *Bradyrhizobium japonicum*, which is composed of a single 8.7 Mb chromosome, that of NGR234 is partitioned into three replicons: a chromosome of about 3.5 Mb, a megaplasmid of more than 2 Mb (pNGR234*b*) and pNGR234*a*, a 536,165 bp plasmid that carries most of the genes required for symbioses with legumes. Symbiotic loci represent only a small portion of all the genes coded by rhizobial genomes, however. To rapidly characterize the two largest replicons of NGR234, the genome of strain ANU265 (a derivative strain cured of pNGR234*a*) was analyzed by shotgun sequencing.

Results: Homology searches of public databases with 2,275 random sequences of strain ANU265 resulted in the identification of 1,130 putative protein-coding sequences, of which 922 (41%) could be classified into functional groups. In contrast to the 18% of insertion-like sequences (ISs) found on the symbiotic plasmid pNGR234*a*, only 2.2% of the shotgun sequences represent known ISs, suggesting that pNGR234*a* is enriched in such elements. Hybridization data also indicate that the density of known transposable elements is higher in pNGR234*b* (the megaplasmid) than on the chromosome. *Rhizobium*-specific intergenic mosaic elements (RIMEs) were found in 35 shotgun sequences, 6 of which carry RIME2 repeats previously thought to be present only in *Rhizobium meliloti*. As non-overlapping shotgun sequences together represent approximately 10% of ANU265 genome, the chromosome and megaplasmid may carry a total of over 200 RIMEs.

Conclusions: 'Skimming' the genome of *Rhizobium* sp. NGR234 sheds new light on the fine structure and evolution of its replicons, as well as on the integration of symbiotic functions in the genome of a soil bacterium. Although most putative coding sequences could be distributed into functional classes similar to those in *Bacillus subtilis*, functions related to transposable elements were more abundant in NGR234. In contrast to ISs that accumulated in pNGR234*a* and pNGR234*b*, the hundreds of RIME elements seem mostly attributes of the chromosome.

Background

Many different Gram-negative bacteria colonize the nutrient-rich rhizospheres of plant roots. Some bacteria are pathogenic, whereas others form beneficial associations. In nitrate-poor soils, strains of *Azorhizobium*, *Bradyrhizobium*, *Mesorhizobium* and *Rhizobium* (collectively known as rhizobia), form nitrogen-fixing symbioses with leguminous plants. In compatible interactions, invading rhizobia penetrate their

hosts through infection threads, which develop centripetally. At the same time, new structures called nodules develop from meristems induced in the cortex of infected roots. When infection threads reach nodule cells, rhizobia are released as symbiosomes into the cytoplasm of infected cells where they eventually enlarge and differentiate into nitrogen-fixing bacteroids. Continuous exchange of chemical signals between the two symbionts coordinates expression of bacterial and plant genes required for a symbiotic development. Flavonoids released by legume roots are amongst the first signals exchanged in this molecular dialog. By interacting with rhizobial regulators of the NodD family, flavonoids trigger the expression of nodulation genes (*nod*, *noe* and *nol*). In turn, most nodulation genes participate in the synthesis and secretion of a family of lipochito-oligosaccharide molecules, the Nod factors that are required for bacterial entry into root hairs. Little is known about how rhizobia migrate inside the infection threads, although it seems likely that genetic determinants of both partners are again involved (see [1,2]). Once within the cortex, the rhizobia differentiate into bacteroids where low free-oxygen tensions help coordinate the expression of genes involved in nitrogen fixation (*nif* and *fix*) [3].

Taxonomic proposals based on DNA sequences of highly conserved genes indicate that rhizobia are a group of genetically diverse soil bacteria [4]. Other data suggest that in populations of soil bacteria, natural genetic mechanisms exist which can transform isolates with widely different chromosomal backgrounds into nodulating bacteria (that is, rhizobia) (for review see [1]). Comparisons of genomes of soil bacteria will help define the pools of symbiotic genes. Unfortunately, genomic studies of this kind have been hindered by the relatively large size of rhizobial genomes (6.5 to 8.7 Mb for *R. meliloti* and *B. japonicum*, respectively). Instead, as many symbiotic loci are often clustered on large plasmids in *Rhizobium* strains, or in chromosomal 'symbiotic islands' as in *B. japonicum* [5] and *M. loti* [6], physical and genetic analyzes of symbiotic plasmids or 'islands' prevailed. *Rhizobium* sp. NGR234 was selected for its exceptionally broad host range, which includes more than 112 genera of legumes in addition to the non-legume *Parasponia andersonii* [7,8]. As in *R. meliloti*, the genome of NGR234 is partitioned into three replicons, a chromosome of about 3.5 Mb, a megaplasmid of more than 2 Mb (pNGR234*b*) and pNGR234*a*, a 536 kb symbiotic plasmid [9-11]. Although various experiments have shown that most symbiotic genes are amongst the 416 open reading frames (ORFs) identified in the complete sequence of pNGR234*a* [9,12,13], others are carried by the chromosome and/or the mega-plasmid [10,14].

Many ways of finding genes exist, but with the rapid advances in genomics, among the most effective are those that involve sequencing parts of or entire genomes. Although contiguous sequences of several symbiotic islands/plasmids will be released in the near future, *R. meliloti* strain 1021 as well as the phytopathogens *Ralstonia solanacearum* and *Xanthomonas citri* are the only plant-interacting microbes currently being sequenced [15-17]. The cost of sequencing a complete genome is still well beyond the capability of most laboratories, however. Nevertheless, extensive information on the structure and content of genomes can be gained by randomly sequencing libraries made from total DNA [18-21]. Here, we have used this approach to analyze the megaplasmid and chromosome of NGR234. A total of 2,275 individual shotgun sequences of ANU265 (a derivative strain of NGR234 cured of its symbiotic plasmid [22]) were searched for protein and/or DNA homologies, and putative coding sequences were grouped into 28 classes according to their putative function. In addition, clones carrying various *Rhizobium*-specific repeated elements such as RIME1 and RIME2 were also analyzed.

Results and discussion

Random sequencing of the ANU265 genome

Total genomic DNA of ANU265 was used to construct an M13 library with inserts ranging in size from 0.9 to 1.5 kb. Of the 2,856 random clones analyzed, 80% (2,275) produced high-quality DNA sequence with an average read length of 253 bp (Table 1). In this way, more than 575 kb of total nucleotide sequence was collected, which corresponds to approximately 10% of the ANU265 genome [11]. At 61.2 mol%, the mean G+C content of these sequences is similar to that found for the entire genome [23], but is also significantly higher than the value of 58.5 mol% calculated for pNGR234*a* [9]. This pool of 2,275 sequences was then screened for redundancy. A total of 381 overlapping sequences were identified, and grouped into 195 contigs (sets of overlapping sequences) of two to four elements each: 154 contigs represent pairs of clones, whereas the remaining 73 sequences belong to 23 groups of three elements and one of four clones. Because of the many highly conserved sequences repeated throughout

Table 1

Major characteristics of the ANU265 shotgun library

M13 insert size (range in bp)	900-1,500	
Number of forward-sequencing reactions (M13 primers)	2,856	
Number of sequences stored in database	2,275	
Average length of the edited reads (bp)	253	
Homology searches		
Total number of sequences	2,275	(100%)
Sequences matching:		
rDNA	3	(0.1%)
tRNA	4	(0.1%)
repeated sequences / intergenic elements only	29	(1.3%)
protein-coding-genes of: known function	922	(40.5%)
unknown function	208	(9.0%)
No database match (pioneer sequences)	1,109	(49.0%)

the NGR234 genome [9,11,24], it was not possible to determine if overlapping clones represent contiguous sequences or DNA fragments from distinct repeats. Nevertheless, truly unique sequences represent 92% of the total number of clones. With an average insert size of 1.2 kb, clones tagged with non-overlapping sequences represent more than 40% (2.5 Mbp) of the ANU265 genome.

RIME- and IS-like sequences

Homology searches against nucleotide databases (BLASTN [25]) showed that 35 ANU265 sequences carried *Rhizobium*-specific intergenic mosaic elements (RIMEs). First identified in *R. meliloti*, *R. leguminosarum* bv. *viciae* and NGR234, RIME1 elements are 108 bp repeats characterized by two large palindromes, whereas RIME2 sequences are 109 bp repeats thought to be present only in *R. meliloti* [26]. RIMEs have many features of the short interspersed repeated elements that are non-coding, intercistronic sequences of less than 200 bp found in many prokaryotic genomes [27]. Of the 2,275 shotgun sequences of ANU265 collected, 29 contained RIME1 elements and 6 carried RIME2 repeats. Although Southern hybridizations indicated that approximately 20 copies of RIME1 were present in the genomes of *R. meliloti* and NGR234 [26], our data indicate that there are many more. Among the 29 clones with RIME1 sequences, most (23) carry repeats that are very similar to the consensus ([26] and Figure 1). In another six (Figure 1, clones 27d06, 29g08, 01f01, 11b07, 25e07 and 13c06), only one of the two large palindromic structures is conserved, however. This suggests that, in some cases, individual palindromes constitute independent repeats, not necessarily associated to form RIME1 elements. In the eight clones that code for putative proteins (Figure 1), RIME1 sequences are found immediately downstream of predicted ORFs (data not shown), indicating that these elements are probably confined to intergenic regions. Surprisingly, no RIME2 and a single RIME1 repeat were found on pNGR234*a* [9,11]. If these elements were regularly distributed throughout the NGR234 genome, more than a single RIME1 would have been expected on the 536 kb of pNGR234*a*. Thus, current data suggest that RIMEs preferentially accumulate on specific replicons, and that NGR234 carries possibly as many as 200 RIME-like elements.

In contrast to pNGR234*a*, which carries many IS sequences, only 2.2% (51) of the 2,275 ANU265 sequences were predicted to encode transposon-related functions. Although several clones that did not match database homologs may also carry sequences of yet uncharacterized IS elements, these results suggest that in proportion to their size, chromosome and megaplasmid carry fewer transposable elements than pNGR234*a*. Nevertheless most of the 51 clones

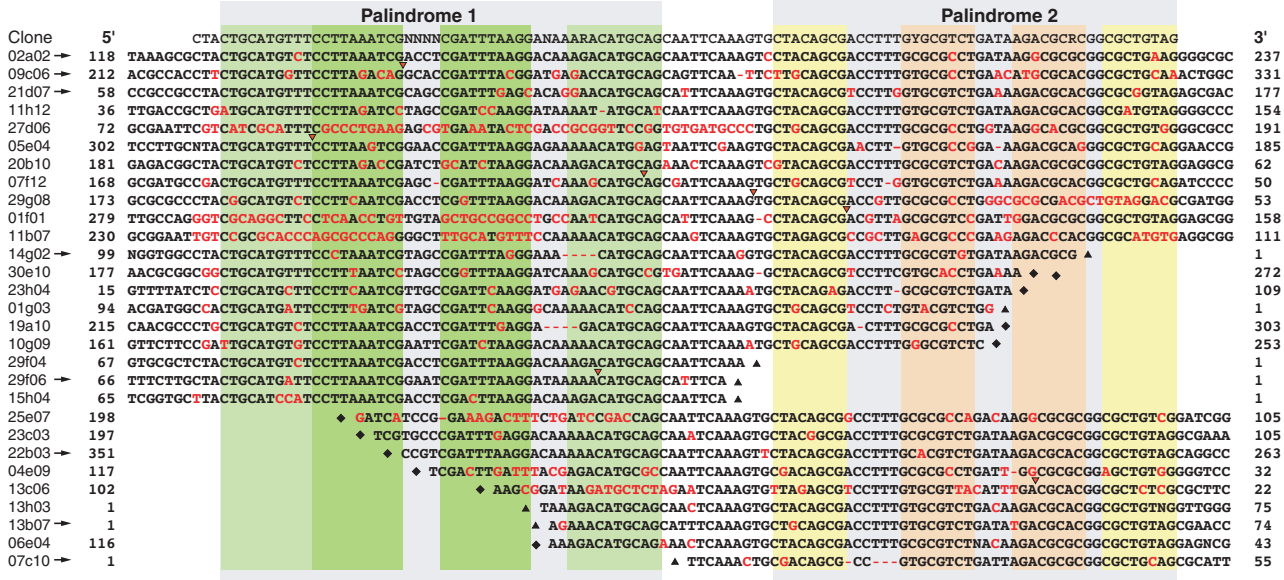

Figure 1
ANU265 clones that carry RIME1 repeats. ANU265 clones are numbered in bold in the first column and the RIME1 repeat consensus sequence is shown in bold on top row of the alignment. Positions in the ANU265 sequences of the initial and final bases in each alignment are given in the 5' and 3' columns, respectively. Partial alignments represent RIME repeats covering either the end (filled diamonds) or the beginning (filled triangle) of the established sequence. The two palindromic structures of RIME1 are shaded in gray. Internal inverted repeats are shown in matching colors. Gaps (marked with red hyphens) and single-nucleotide deletions (inverted red triangles) were introduced for optimal alignment. Base mismatches are colored in red. Arrows mark clones that encode putative proteins. N, any base; Y indicates C or T, R indicates A or G.

(70%) matched ISs that were first identified in pNGR234*a* [9]. For example, ten sequences highly homologous to NGRIS-4 were found. This 3,316 bp element is duplicated in pNGR234*a* [9], whereas chromosome and megaplasmid carry two and five copies of NGRIS-4 respectively [11,24].

Identification of putative genes

To assign putative functions to the cloned DNA fragments, sequences were compared to protein and nucleotide databases [25,28]. BLAST analyses showed that about 50% (1,130) of the 2,275 sequences matched protein-coding ORFs, three were homologous to rDNA and four to tRNA loci (see Table 1). Of the 1,130 putative protein-coding sequences, 208 (or 9% of the 2,275 sequences) were similar to hypothetical genes with no known function (pioneer sequences) of rhizobia and other organisms. Thus, together with the 1,109 clones which showed no significant similarity to entries in nucleotide and amino-acid databases (see Table 1), functions could not be assigned to 58% of the shotgun sequences. To provide an overview of the genetic organization of the ANU265 genome, predicted protein-coding sequences were grouped into various classes according to their putative function (Table 2).

A genetic snapshot of the ANU265 genome

In total, 922 of the 2,275 sequences were grouped into 28 functional categories (Table 2). Interestingly, comparison of this data with that derived from the complete sequence of the *Bacillus subtilis* genome [29] showed a similar distribution of genes in both organisms. Although *B. subtilis* is a Gram-positive bacterium, it is commonly found in soil, water sources and in associations with plants. Thus, with the exception of one homolog of a sporulation gene (which was not expected in rhizobia), the comparative analysis presented in Table 2 suggests that the number of shotgun sequences is probably sufficiently large to form a representative selection of ANU265 loci. All 1,130 sequences for which significant matches were found in database searches are classified by function in Table 3.

As in other bacterial genomes, such as that of *Escherichia coli* [30], the largest functional class represents transport and binding proteins (see Tables 2 and 3). A number of essential genes, including those required for replication, transcription and translation as well as those linked to primary metabolism, were also found. As expected of a soil-borne prokaryote, many loci (18%) involved in carbon and nitrogen metabolism were identified (encoding enzymes for the assimilation of nitrate/ammonia, the tricarboxylic acid cycle, or transporters of dicarboxylic acids, and so on). In *B. subtilis*, 19% of the protein-coding genes are devoted to the metabolism of carbohydrates, amino acids and related molecules (Table 2). This is in contrast to microorganisms such as *Haemophilus influenzae* and *M. genitalium* that are not able to grow on many nitrogen and carbon sources (only 10% of their predicted genes code for such metabolic functions [31]). Interestingly, homologs of various chaperones such as GroES/GroEL, DnaJ, and other small heat-shock proteins (sHsps), were identified (Table 3, clones 308 to 318). The presence of multiple sHsps is not common in prokaryotes, but was shown to be widespread in rhizobia [32].

Obviously, the ability of rhizobia to respond to plant compounds that stimulate their growth contributes to successful colonization of the root [33] and absence of vitamins often limits the growth or rhizobia. Furthermore, the ability to either take up or synthesize vitamins is thought to be an essential characteristic of rhizobia [33]. For these reasons, it is not surprizing that several ANU265 sequences matched genes for biotin and thiamine utilization, such as that coding for a homolog of *bioS* (clone 745), a biotin-regulated locus of *R. meliloti* [34]. In *R. meliloti*, *bioS* is part of an operon which includes the *surE* and *lppB/nlpD* genes that are also found in ANU265 (clones 744 and 183). Homologs of thiamine biosynthetic genes *thiCG* of *R. etli* (clones 512 and 513) were also found. Miranda-Rios *et al.* [35] reported a direct correlation between the expression of *thiC* and the production of the symbiotic terminal oxidase *cbb*3, which is required for bacteroid respiration under conditions of low oxygen.

Putative symbiotic genes include loci involved in exopolysaccharide (EPS) biosynthesis and/or export, which are encoded by pNGR234*b* [10], as well as genes involved in the elaboration of acidic capsular polysaccharides (K-antigens), lipopolysaccharides and cyclic β-glucans (Table 3, clones 245 to 270). A sequence homologous to *fixN* of *R. meliloti* was also identified (clones 208 and 209). The chromosomal *fixNOPQ* locus encodes an oxidase complex that is probably active during nitrogen fixation. Although sequences of the regulatory *fixK* genes [3] were identified (clone 683), no significant match to the oxygen-responsive system encoded by *fixLJ* was found. Members of other symbiotic two-component regulatory systems were detected in ANU265, however, including homologs of the sensor histidine kinase *exoS* (clone 200) and the response regulator *chvI* (clone 717). Both are necessary for regulating production of succinoglycans that are important in *R. meliloti-Medicago sativa* symbioses [36]. Similarly, the *nwsA* locus (clone 202) encodes a putative sensor kinase that is involved in the expression of nodulation genes in *Bradyrhizobium* strains [37].

It has been postulated that genes responsible for the synthesis (*mos*) and catabolism (*moc*) of rhizopines confer a competitive advantage on their host rhizobia [38]. Rhizopines are synthesized in nodules of *M. sativa* inoculated with *R. meliloti* strain L5-30, and can be used as growth substrates by certain rhizobia. Although *mos* and *moc* genes were thought to be limited to *R. meliloti* strains [39], homologs of *mocABC*, and *mosA* genes were also found in ANU265 (clones 543 to 549). Propagation of rhizobia in the

Table 2

Comparison of sequences encoding probable cellular functions of *Rhizobium* sp. NGR234 strain ANU265 with functional classes of proteins of *Bacillus subtilis*

Functional categories	ANU265		*B. subtilis**	
Cell envelope and cellular processes				
Cell wall	17	(1.9)	93	(3.9)
Transport/binding proteins and lipoproteins	184	(20.0)	381	(16.0)
Sensors (signal transduction)	21	(2.3)	38	(1.6)
Membrane bioenergetics (electron transport and ATP synthase)	49	(5.3)	78	(3.3)
Surface polysaccharides biosynthesis and export	25	(2.7)	16	(0.7)
Sporulation	1	(0.1)	139	(5.8)
Germination/transformation	0		43	(1.8)
Mobility and chemotaxis	26	(2.8)	55	(2.3)
Cell division	5	(0.5)	21	(0.9)
Protein secretion	13	(1.4)	18	(0.8)
Chaperones/heat-shock proteins	12	(1.3)	15	(0.6)
Cell death	8	(0.9)	5	(0.2)
Intermediary metabolism				
Carbohydrates and related molecules	69	(7.5)	261	(11.0)
Amino acids and related molecules	91	(9.9)	202	(8.5)
Nucleotides and nucleic acids	11	(1.2)	83	(3.5)
Lipids	19	(2.1)	77	(3.2)
Cofactors/prosthetic groups	37	(4.0)	99	(4.2)
Phosphate	3	(0.3)	9	(0.4)
Opine-like compounds	8	(0.9)	3	(0.1)
Sulphur	2	(0.2)	8	(0.3)
Information pathways				
DNA replication, restriction, modification and repair	26	(2.8)	61	(2.6)
DNA segregation, recombination and transfer	10	(1.1)	27	(1.1)
RNA synthesis and modification	19	(2.1)	50	(2.1)
Protein synthesis and modification	63	(6.8)	123	(5.2)
Regulatory functions	68	(7.4)	213	(8.9)
Other categories				
Adaptation to atypical conditions and protection	27	(2.9)	147	(6.2)
Transposon-related functions	51	(5.5)	10	(0.4)
Phage-related functions	5	(0.5)	83	(3.5)
Miscellaneous	52	(5.6)	21	(0.9)
Total	922	(100)	2,379	(100)

*The functional classification of the *B. subtilis* protein-coding genes was adapted from Kunst *et al.* [29]. The number of sequences and of genes in each category is listed for ANU265 and *B. subtilis*, respectively. The percent of the putatively identified genes devoted to each functional group is indicated in brackets.

soil, and hence their symbiotic efficiency, probably also depends on their tolerance to osmotic changes. It is thus notable that homologs of the *R. meliloti betABC* genes, which are involved in the osmoregulatory choline-glycine betaine pathway [40], were also found (clones 726 to 730).

Other putative symbiotic loci include homologs of the *phbC* and *prsDE* genes of *R. meliloti,* which encode a poly-3-hydroxybutyrate synthase [41] and a type I secretion system [42] (clones 741 to 743, and 298 to 301, respectively). Interestingly, PrsD and PrsE of *R. meliloti* are involved in the secretion of enzymes that modify succinoglycans [43], whereas a similar type I secretion system seems to be responsible for the export of the nodulation-signaling protein NodO in *R. leguminosarum* bv. *viciae* [44,45]. Although the role of these *prsDE* homologs in NGR234 is not clear, it is possible that more than one type of protein secretion system has a symbiotic role in this bacterium [46].

Table 3

Classification of putative protein-coding genes of *Rhizobium* sp. NGR234 cured of its symbiotic plasmid (= ANU265)

Clone No.	Name	Homolog description
Cell envelope and cellular processes		
Cell wall		
1	01d07	*N*-acetylmuramoyl-L-alanine amidase
2	06g04	*N*-acetylglucosamine-1-phosphate uridyl Tase
3	26b06	UDP-*N*-acetylenolpyruvoylglucosamine RDase MurB
4	28f02	UDP-*N*-acetylmuramate-alanine ligase MurC
5	22b10	UDP-*N*-acetylmuramoylalanine-D-glutamate ligase MurD
6	29g07	UDP-*N*-acetylmuramyl-tripeptide synthetase MurE
7	08h05	UDP-*N*-acetylmuramyl-tripeptide synthetase MurE
8	12f10	outer membrane prot.
9	17a08	outer membrane prot. Omp28
10	21h07	group 1 outer membrane prot. OMP1 precursor60
11	18f12	penicillin-binding prot. 1B
12	19h11	penicillin-binding prot. 1A
13	29f10	D-alanyl-D-alanine carboxypeptidase
14	19d01	monofunctional biosynthetic peptidoglycan TGase MtgA
15	28f09	lysozyme M1 precursor Acm
16	03f07	acriflavine resistance prot. E precursor AcrE
16a	*09h01*	*overlaps clone 03f07*
Transport/ binding proteins and lipoproteins		
17	22a03	sugar-binding transp. ATP-binding prot.
18	08h08	ABC transp. ATP-binding prot.
19	29c10	sugar ABC transp., permease prot.
20	03f05	inner membrane prot.of trehalose/ maltose transp.
21	02h10	transp. permease prot.
21a	*09f05*	*overlaps clone 02h10*
22	11c04	ABC transp. permease prot.
23	12d03	inner membrane ABC transp.
24	18h08	sugar ABC transp. ATP-binding prot.
25	21b03	ATP-binding transp. prot.
26	01h04	ATP-binding transp. prot.
27	21b11	ATP-binding transp. prot.
28	26g01	maltose/ maltodextrin transp. ATP-binding prot.
29	18g11	sugar ABC transp. ATP-binding prot.
30	24e03	ABC transp. ATP-binding prot.
31	02c09	ABC transp. ATP-binding prot.
31a	*12f05*	*overlaps clone 02c09*
32	28g12	ABC transp. ATP-binding prot.
32a	*01a11*	*overlaps clone 28g12*
32b	*09h06*	*overlaps clone 28g12*
33	01a10	ribose transp. ATP-binding prot.
34	06h03	D-ribose-binding periplasmic prot. precursor
35	14e04	sugar transp. system permease prot.
36	18d02	sugar transp. system permease prot.
37	04e03	sugar transp. system permease prot.
38	14f09	sugar transp. system permease prot.
39	08g05	sugar transp. system permease prot.
40	06b09	sugar transp. system permease prot.
41	19e12	sugar transp. system permease prot.
42	16f12	membrane-spanning permease
43	26d10	sugar transp. system permease prot.
44	23b06	sugar transp. system permease prot.
45	27h12	ABC transp. integral membrane prot.
46	22d10	ribose ABC transp. permease prot.
47	25a11	sugar transp.
48	21a06	sugar transp. ATP-binding prot.
49	21b02	sugar transp. ATP-binding prot.
50	24d10	galactoside transp. ATP-binding prot. MglA
51	24e10	lactose transp. system permease prot. LacF
52	11f10	sugar transp. prot.
53	26d04	sugar transp. ATP-binding prot.
54	26h11	sugar transp. ATP-binding prot.
55	05b03	sugar transp. system permease prot.
56	30b08	sugar transp. system permease prot.
57	26a07	sugar ABC transp., ATP-binding prot.
58	14e06	xylose transp. permease prot.
59	03f04	xylose transp. permease prot.
60	19a04	maltose binding prot.
61	11c09	membrane bound sugar transp. prot.
62	18a12	sugar transp. ATP-binding prot.
63	28e01	sugar transp. ATP-binding prot.
64	30f09	sugar transp. prot.
65	01c10	galactoside transp. system permease prot.
66	19g07	galactoside transp. ATP-binding prot.
67	21e06	branched-chain amino acid transp.
68	29a02	amino-acid ABC transp. permease prot.
69	27b03	amino-acid ABC transp. permease prot.
70	05c04	ABC transp. permease prot.
71	07b05	amino-acid ABC transp. ATP-binding
72	19h08	amino-acid ABC transp. ATP-binding
73	06a09	amino-acid ABC transp. ATP-binding
74	11d03	glutamate/ aspartate transp. system permease prot.
75	21d10	high-affinity branched-chain amino acid transp.
75a	*24a05*	*overlaps clone 21d10*
76	02c01	amino acid ABC transp.
77	08f08	branched-chain amino acid transp. prot.
78	10e04	branched-chain amino acid transp. prot.
79	17b01	branched-chain amino acid transp. prot.
80	08g08	branched-chain amino acid transp. prot.
81	04g02	branched-chain amino acid transp. permease prot.
82	28h02	high-affinity branched-chain amino acid transp.
82a	*08a04*	*overlaps clone 28h02*
82b	*11g01*	*overlaps clone 28h02*
83	16c12	periplasmic dipeptide transp. prot. precursor
84	03h07	dipeptide ABC transp.
85	10d11	peptide ABC transp. permease prot.
86	15e08	ABC transp. ATP-binding prot.
87	04a06	peptide ATP-bind. transp.
88	09a09	peptide ABC transp. permease prot.
89	12c06	peptide ABC transp. permease prot.
90	03f10	ABC transp.
91	22c01	ABC transp. ATP-binding prot.
92	19b12	ABC transp. ATP-binding prot.
93	23g09	peptide ABC transp. ATP-binding prot.
94	18e05	ABC transp. ATP-binding prot.
95	14d12	oligopeptide ABC transp.
96	19c05	oligopeptide binding prot.
97	21g02	peptide ABC transp.
98	22h09	dipeptide transp. ATP-binding prot.
99	25c02	oligopeptide transp. ATP-binding prot.
99a	*23c05*	*overlaps clone 99*
100	27b09	oligopeptide ABC transp. permease prot.
101	27c09	oligopeptide-binding prot. precursor
102	30a09	oligopeptide transp. ATP-binding prot
103	03b02	ABC transp., y4wM pNGR234*a*
104	03g02	ABC transp., y4wM pNGR234*a*
105	07c06	ABC transp., y4wM pNGR234*a*
106	30e02	ATP-binding prot.
107	05c03	ATP-binding prot.
108	20e03	ATP-binding prot.

Table 3 (continued)

Clone No.	Name	Homolog description
109	19f02	ABC transp. ATP-binding prot.
110	19d07	ABC transp. ATP-binding prot.
111	16a07	ATP-dependent transp.
112	17f05	ABC transp. ATP-binding prot.
113	09d03	ABC transp. ATP-binding prot.
114	17h02	putrescine transp. system permease prot.
115	22e01	inner membrane prot.
116	02g07	spermidine/ putrescine transmembrane prot.
116a	*06d04*	*overlaps clone 02g07*
116b	*24b12*	*overlaps clone 06d04*
117	13b12	putrescine transp. prot.
118	06d07	putrescine transp. permease prot.
119	09b07	putrescine transp. permease prot.
119a	*06b04*	*overlaps clone 09b07*
120	24a01	glycine betaine transp. system permease prot
121	24f03	glycine betaine transp. system permease prot
122	28c03	glycine betaine / proline transp. prot. ProV
123	10h02	inner membrane prot.
124	03b03	aquaporin Z (bacterial nodulin-like intrinsic prot.)
125	03c04	arginine / ornithine antiporter
126	03e12	glycerol-3-phosphate-binding periplasmic prot.
127	22g07	glycerol-3-phosphate transp. prot.
128	05a06	acriflavine resistance lipoprot. A precursor
129	29h11	acriflavine resistance prot. B
129a	*15d06*	*overlaps clone 29h11*
130	14b09	acriflavine resistance prot.
131	14c06	antibiotic resistance prot
131a	*05c12*	*overlaps clone 14c06*
132	04d08	Leu/ Ile/ Val/ (Thr/Ala)-binding prot. precursor
133	07e02	cytoplasmic prot. CeoB
134	01c07	NolH (AcrB/AcrD/AcrF family prot.)
135	27a10	FixI; E1-E2 type cation ATPase
136	12b12	heavy-metal transp.ing P-type ATPase
137	29f07	cation-transp. ATPase PacS
138	11e02	H^+/Ca^{2+} exchanger
139	01g05	tonB-dependent outer membrane heme receptor HemR
140	02b10	inner membrane prot., energy transducer TonB
141	27h11	TonB-dependent transp. ExbD
142	02b11	nitrite extrusion prot.
143	08f10	nitrate transp. permease prot. nrtB
144	16d07	nitrate transp. prot. NrtD
145	09g09	phosphate transp. prot. PhoE
146	27h09	phosphate transp. prot. PhoT
146a	*11g03*	*overlaps clone 27h09*
147	17e11	phosphate transp. prot. PhoT
148	17c11	phosphate transp. prot. Pit
149	21a10	phosphate transp. prot. Pit linked to RIME 2
150	04d06	Pit accessory protein orfA
151	12d06	macrolide-efflux determinant
152	13d04	cation efflux system prot.
153	17e08	cation efflux system prot.
154	21e03	ferric siderophore receptor
154a	*14c02*	*overlaps clone 21e03*
155	29b02	ferric siderophore receptor
155a	*18g03*	*overlaps clone 29b02*
156	14e12	potassium uptake prot. Kup
157	14f07	phosphoenolpyruvate-prot. phosphoTase
158	15f09	ABC transp. ATP-binding prot.
159	16d04	molybdenum transp. prot.
160	16g11	periplasmic sulphate binding prot. Sbp
161	04f12	periplasmic sulphate binding prot. Sbp
162	18b09	drug efflux pump (AcrB/AcrD/AcrF family)
163	18d12	tartrate transp. TtuB
164	01h09	L-asparagine permease AnsP
165	29d05	C4-dicarboxylate transp. prot. DctA1 pNGR234*a*
166	20a10	C4-dicarboxylate transp. prot. DctA1 pNGR234*a*
167	20c09	chelated iron ABCtransp. ATP-binding prot.
168	29f01	chelated iron ABCtransp. ATP-binding prot.
169	03c12	chelated iron transp. system membrane prot.
170	19a03	chelated iron transp. system membrane prot.
171	11d07	chelated iron transp. system membrane prot.
172	26g08	iron transp. prot.
173	20e11	phosphoenolpyruvate-prot. phosphoTase
174	20e12	Na+/H+-exchanging prot. system component
175	22b06	mannopine-binding periplasmic prot. motA
175a	*21g06*	*overlaps clone 22b06*
176	21h10	sulfate transp. system permease prot.
177	29h02	taurin-binding periplasmic prot
178	22d09	cytoplasmic membrane prot. CeoB
179	28f12	integral membrane prot. (sodium:sulfate symporter)
180	23g03	sulphate transp. system permease prot. CysT
181	25h07	transp. prot., *y*4xM pNGR234*a*
182	24h07	periplasmic binding prot.
183	08c01	lipoprot. LppB/NlpD
183a	*10c07*	*overlaps clone 08c01*
184	24f12	lipoprot.
185	23h12	lipoprot.
186	06a01	outer membrane lipoprot.
186a	*11a05*	*overlaps clone 06a01*
186b	*21d02*	*overlaps clone 11a05*
Sensors (signal transduction)		
187	12f09	sensor histidine kinase ExsG
187a	*15d09*	*overlaps clone 12f09*
188	18f03	sensor histidine kinase ExsG
189	06a07	sensor prot. TctD
190	16d05	sensor prot. for potassium transp. KdpD
191	06d10	sensor prot. for potassium transp. KdpD
191a	*25f03*	*overlaps clone 06d10*
191b	*26d08*	*overlaps clone 06d10*
192	09c11	two-component sensor histidine kinase
192a	*26a04*	*overlaps clone 09c11*
193	10f06	C4-dicarboxylate sensor prot. DctB
194	13b09	C4-dicarboxylate sensor prot. DctB
195	14c01	sensor of two-component system FlhS
196	01g04	sensor of two-component system FlhS
197	15f11	prokaryotic sensory transduction prot.
198	15g02	sensory transduction histidine kinase
199	19a06	sensory transduction histidine kinase
199a	*22d04*	*overlaps clone 19a06*
200	22g10	histidine kinase sensory prot. ExoS
201	23e05	histidine prot. kinase ActS
202	29f03	sensor kinase NwsA
Membrane bioenergetics (electron transport, etc)		
203	09f04	pyridine nucleotide transhydrogenase sub. αPntA
204	02h05	pyridine nucleotide transhydrogenase sub. α PntA
205	09d11	pyridine nucleotide transhydrogenase sub. α PntA
206	20h04	pyridine nucleotide transhydrogenase sub. β PntB
207	25g07	pyridine nucleotide transhydrogenase sub. β PntB
208	13b05	FixN cytochrome CBB3 sub. I
209	01b04	FixN cytochrome CBB3 sub. I
210	08a07	FixS cbb3-type cytochrome oxidase formation prot.
211	24d07	cytochrome-c oxidase chain IIIB CoxP
212	05f03	cytochrome BB3 sub. I CoxN
212a	*03h10*	*overlaps clone 05f03*

Table 3 (continued)

Clone No.	Name	Homolog description
212b	*12g07*	*overlaps clone 05f03*
213	05f06	cytochrome C oxidase assembly prot. CoxZ
214	05e03	cytochrome C-type biogenesis prot. CycJ/K
215	06d08	cytochrome C-type biogenesis prot. CycH
216	11f02	cytochrome c-type biogenesis prot. CcdA
217	12e03	cytochrome oxidase δ, sub. II
218	11g05	ubiquinol-cytochrome C RDase iron-sulfur sub.
219	15e06	cytochrome o ubiquinol oxidase sub. III
220	29e05	cytochrome c small sub. of nitric oxide RDase
221	06h06	glycolate oxidase iron-sulfur sub. (Fe-S prot.)
221a	*07e07*	*overlaps clone 06h06*
222	08h10	ATP synthase α-chain
222a	*14d10*	*overlaps clone 08h10*
222b	*22c06*	*overlaps clone 14d10*
223	17f07	ATP synthase γ-chain
224	09a12	NADH-ubiquinone oxidoRDase (CI-40 kDa)
225	09h02	NADH-ubiquinone oxidoRDase (CI-51kDa)
226	13d12	cyanide insensitive terminal oxidase CioAB
227	01g02	NADH DHase (ubiquinone), sub. I
228	19b04	NADH DHase
229	12a01	NADH ubiquinone oxidoRDase
230	20c11	NADH ubiquinone oxidoRDase
231	06e08	NADH ubiquinone oxidorRDase
232	26e10	NADH ubiquinone oxidoRDase
233	08g12	NADH ubiquinone oxidoRDase
234	21h02	NADH ubiquinone oxidoRDase
235	22a06	NADH ubiquinone oxidoRDase
236	27c06	NADH ubiquinone oxidoRDase
237	24b11	electron transfer flavoprotein-ubiquinone oxidoRDase
238	25a08	glutathione RDase
239	01h05	thioredoxin
240	29a01	thioredoxin RDase
241	16d03	ferredoxin [2Fe-2S] II
242	19f07	ferredoxin-type prot. [4Fe-4S]
243	09c06	ferredoxin [3Fe-4S][4Fe-4S]
244	26a10	ferredoxin, [2Fe-2S]
244a	*05c08*	*overlaps clone 26a10*
244b	*18a04*	*overlaps clone 26a10*
Surface polysaccharide - biosynthesis and export		
245	01c04	ExoB UDP-galactose 4-epimerase
246	03a05	ExoB UDP-galactose 4-epimerase
247	06a08	ExoN UDP-glucose pyrophosphorylase
248	07g05	ExoN UDP-glucose pyrophosphorylase
249	10e08	ExoF exopolysaccharide production prot. precursor
250	10c12	ExoK endo-β-1,3-1,4-glucanase
251	18b08	ExoP succinoglycan transp. prot.
252	22e05	ExoU succinoglycan biosynthesis glycosylTase
253	27h01	ExoU succinoglycan biosynthesis glycosylTase
254	18h12	ExoY exopolysaccharide production prot.
255	25h10	ExoL succinoglycan biosynthesis glycosylTase
256	17b10	exopolysaccharide production prot. Pss
257	11a12	KPS production, fatty acid synthase RkpC
258	27f08	KPS modification / export prot. RkpI
259	11d09	KPS modification / export prot. RkpJ
260	17h04	polysialic acid transp. prot. KpsM
261	21b04	specialized small acyl carrier prot. (lipid A)
262	10c11	*N*-acetylglucosamine deacetylase (lipid A)
263	10b05	3-deoxy-D-manno-octulosonic-acid (Kdo) Tase KdtA
264	18f02	3-deoxy-D-manno-octulosonic-acid (Kdo) Tase KdtA
265	26h01	3-deoxy-manno-octulosonate cytidylylTase KpsU
266	12c02	membrane bound galactosylTase RfpB
267	08h09	O-antigen acetylase
268	13h11	capsular polysaccharide biosynthesis prot.
269	08c05	spore coat polysaccharide biosynthesis prot.E
270	14d02	β-(1,2)-glucans production inner-membrane prot. NdvB
Mobility / chemotaxis		
271	17a10	(MCP)-glutamate methylesterase CheB
272	05c06	flagellar basal-body (proximal) rod prot. FlgB
273	29g09	flagellar biosynthetic prot. FlhB
273a	*08g10*	*overlaps clone 29g09*
273b	*24g01*	*overlaps clone 08g10*
274	13b08	flagellar biosynthetic prot. FliQ
275	12f06	flagellar basal-body MS-ring prot. FliF
276	14e10	flagellar hook prot. FlgE
277	14f11	flagellar basal-body (distal) rod prot. FlgG
278	17c05	flagellar C-ring prot. FliG
279	26d03	flagellar biosynthetic prot. FliP
280	18g10	flagellar basal-body rod prot. FlgF
281	27a08	flagellin prot. FlaD
282	20h07	new class of flagellar prot. FlmD
282a	*19b08*	*overlaps clone 20h07*
282b	*30a08*	*overlaps clone 20h07*
283	07h07	chemotactic transducer for amino acids
284	13g03	methyl-accepting chemotaxis prot
285	11b02	methyl-accepting chemotaxis prot
286	11c06	methyl-accepting chemotaxis prot
287	15b10	methyl-accepting chemotaxis prot
288	16b03	methyl-accepting chemotaxis prot
289	16f10	methyl-accepting chemotaxis prot
290	28h04	methyl-accepting chemotaxis prot
291	28a12	methyl-accepting chemotaxis prot
292	27f01	chemotaxis prot. methylTase CheR
292a	*15e07*	*overlaps clone 27f01*
Cell division		
293	05f02	cell division prot. FtsH
294	08d04	cell division prot. FstK
295	16b04	cell division prot. FtsK
296	16e07	septum formation prot. Maf
297	25f06	cell division inhibitor MinC
Protein secretion		
298	14c09	ABC transp. type I PrsD
298a	*06a04*	*overlaps clone 14c09*
299	24e06	ABC transp. type I PrsD
300	11h03	membrane fusion prot. type I PrsE
301	29e03	membrane fusion prot. type I PrsE
302	06h10	general secretion pathway prot. D precursor XpsD
302a	*03e03*	*overlaps clone 06h10*
303	12h06	general secretion prot. F XcpS
303a	*10c03*	*overlaps clone 12h06*
304	29d12	general secretion prot. D GspD
304a	*12e11*	*overlaps clone 29d12*
304b	*10e05*	*overlaps clone 12e11*
305	22g03	general secretion prot. E GspE
306	28g09	pNGR234*a*, probable translocation prot. RhcT
307	29b12	preprotein translocase SecA sub.
Chaperones		
308	10d03	heat shock prot. cnp60 GroEL
309	21a09	heat shock prot. cnp60 GroEL
310	26d02	heat shock prot. cnp10 A GroES
311	21a03	heat shock prot. cnp60 GroEL
312	07c04	small heat shock prot. HspF
313	14e03	small heat shock prot. (hsp20-2)

Table 3 (continued)

Clone No.	Name	Homolog description
313a	*18g01*	*overlaps clone 14e03*
314	28f11	18 kd antigen2 (small heat shock prot. Hsp20 family)
315	15g09	small heat shock prot. HspE
316	17b02	DnaJ-like heat shock chaperone prot.
317	01e06	heat shock prot. 90 HtpG
318	23g01	heat shock prot. X HtpX
Cell death		
319	08h11	hemolysin-like prot. TlyC
319a	*05b05*	*overlaps clone 08h11*
320	04h12	cyclolysin (haemolysin-adenylate cyclase toxin)
321	22b07	cyclolysin (haemolysin-adenylate cyclase toxin)
322	10b08	cyclolysin (haemolysin-adenylate cyclase toxin)
323	21g08	cyclolysin (haemolysin-adenylate cyclase toxin)
324	12b07	iron-regulated prot. (cytotoxins Ca^{2+} binding domain)
325	20c02	hemolysin
Intermediary metabolism		
Metabolism of carbohydrates and related molecules		
326	01b09	glucose-6-phosphate isomerase
327	24c09	glucose-6-phosphate isomerase
328	18g07	glucose-6-phosphate isomerase
328a	*20e09*	*overlaps clone 18g07*
329	09a03	glyoxylate carboligase
330	01c12	α-ketoglutarate DHase
330a	*24d12*	*overlaps clone 01c12*
331	16g02	acetoin:DCPIP oxidoRDase α
332	02e09	acetoin:DCPIP oxidoRDase β
333	02e11	succinyl-coA synthetase β chain
334	03e11	ribulose-bisphosphate carboxylase, large sub.
335	03h09	citrate synthase
336	05b01	L-xylulose kinase
337	06c08	dihydoxyacetone kinase
338	18g08	dihydoxyacetone kinase
339	06g05	lipoamide DHase E3 subunit of α-ketoacid DHase complex.
340	04g03	alcohol DHase(acceptor) precursor
341	04h06	malate DHase
342	09d09	malate DHase
343	07e09	glycogen operon protein (glycosyl hydrolases family)
344	08b07	alcohol DHase
345	08e03	2-hydroxyhepta-2,4-diene-1,7-dioate isomerase
346	13a05	glycolate oxidase sub
347	09d08	glycolate oxidase sub
348	10c06	acetyl-CoA synthetase
349	11b03	aconitate hydratase (citrate hydro-lyase)
350	25a09	aconitate hydratase (citrate hydro-lyase)
351	01b12	2-keto-3-deoxygluconokinase
352	15a09	ribitol kinase
353	11c02	glucose DHase (pyrroloquinoline-quinone)
354	25b01	formaldehyde DHase (glutathione)
354a	*11c11*	*overlaps clone 25b01*
355	09a10	β-glucosidase (cellulose degradation)
356	07a05	β-glucosidase (cellulose degradation)
357	18b12	β-glucosidase (cellulose degradation)
357a	*12b03*	*overlaps clone 18b12*
357b	*07f11*	*overlaps 12b03*
358	12c07	NADP-dependent malic enzyme
359	12g01	phosphogluconate DHase
360	28c02	glutathione-dependent formaldehyde DHase
361	04h08	glycerol-3-phosphate DHase
362	14g01	glycerol-3-phosphate DHase
363	30f10	glycerol-3-phosphate DHase
364	15e12	dTDP-glucose 4-6-DTase
365	21f02	mannonate DTase
366	16g05	alcohol DHase
367	17c02	phosphomannomutase AlgC
368	17f03	glycogen phosphorylase
369	18d11	phosphoglucomutase
370	18f01	L-ribulose-P-4-epimerase (AraD/FucA family)
371	18h09	triosephosphate isomerase
372	19d03	starch (bacterial glycogen) synthase
373	20d06	zinc-type alcohol DHase-like prot
374	20d08	glutathione-dependent formaldehyde DHase
375	20e05	succinate DHase (iron-sulfur prot.)
376	20h10	tartrate DHase
377	21b06	lactaldehyde DHase
378	02d12	D-lactate DHase
379	22b08	D-lactate DHase
380	22c08	dihydrolipoamide acetylTase
381	28b09	dihydrolipoamide DHase
382	30a11	dihydrolipoamide DHase
383	23f06	transketolase
384	23h07	α-glucosidase
385	28a07	D-mannonate oxidoreductase
386	28c08	glutathione-independent formaldehyde DHase
387	29a04	y4uC, pNGR234*a*, aldehyde-DHase-like prot
388	29h05	fumarate hydratase
389	30b10	mannitol 2-DHase
390	05d12	isocitrate lyase
Metabolism of amino acids and related molecules		
391	27a11	α-isopropylmalate synthase LeuA
392	27d07	α-isopropylmalate synthase LeuA
393	14d11	α-isopropylmalate synthase LeuA
394	14b07	3-isopropylmalate dehydratase small sub.LeuD
395	25g10	aspartate ammonia-lyase AspA
396	02c06	aspartate ATase (AspC family)
397	02h01	5-methyltetrahydrofolate-homocysteine Tase MetH
398	06d09	3-dehydroquinate synthase AroD
399	03c03	3-dehydroquinate synthase AroD
400	18c04	shikimate 5-dehydrogenase AroE
401	28a08	shikimate 5-dehydrogenase AroE
402	02h02	3-dehydroquinate DTase AroQ
402a	*20f02*	*overlaps clone 02h02*
403	03b05	aspartate aminoTase B
404	29b06	aspartate aminoTase B
405	26d07	aspartate aminoTase
406	11e06	aspartate aminoTase
407	03e05	y4sL pNGR234*a*, D-amino-acid DHase
408	03g10	adenylosuccinate Sase (IMP-aspartate ligase) PurA
409	04a04	glutamate 5 -kinase
410	05d06	*N*-acyl-L-aminoacid amidohydrolase (aminoacylase)
411	05e10	assimilatory nitrate RDase sub. NirB
411a	*05h01*	*overlaps clone 05e10*
412	05g02	3-isopropylmalate DTase large sub. LeuC
413	05g05	class III pyridoxal-phosphate-dependent ATase
414	06a10	threonine deaminase IlvA
415	26f02	threonine deaminase IlvA
416	13g01	acetolactate Sase (acetohydroxy-acid Sase) IlvB
417	01b11	acetolactate Sase (acetohydroxy-acid Sase) IlvB
418	08b10	dihydroxy-acid DTase IlvD
419	18c03	dihydroxy-acid DTase IlvD
420	06b12	histidinol DHase HisD
421	07a04	*N*-acetylornithine ATase
422	07b06	low specificity D-threonine aldolase
423	08f03	branched-chain α-keto acid DHase component E1

Table 3 (continued)

Clone No.	Name	Homolog description
424	30a04	serine acetylTase (CysE/LacA/LpxA/NodL family)
424a	*10c10*	*overlaps clone 30a04*
425	10e02	anthranilate synthase (tryptophan biosynthesis) TrpE/G
426	30e09	anthranilate synthase (tryptophan biosynthesis) TrpE/G
427	11e05	serine hydroxymethylTase GlyA
428	29h12	tryptophan synthase TrpA
429	11f03	homoserine DHase
430	11h02	5,10-methylenetetrahydrofolate RDase MetF
430a	*15a08*	*overlaps clone 11h02*
431	11h04	proline DHase PutA
432	18d05	proline DHase PutA
433	25c05	proline DHase PutA
434	12b02	glutaryl-CoA DHase (acyl-coA DHase. family)
435	12f03	glycine acetylTase
436	13h10	homoserine DHase
437	14f01	ethanolamine ammonia-lyase heavy chain EutB
438	15b01	2-oxoisovalerate DHase α sub.
439	15d01	methionine gamma-lyase MegI
440	19g12	methionine gamma-lyase MegL
441	16h06	4-hydroxyphenylpyruvate dioxygenase
442	19g03	arginine deiminase ArcA
443	29h06	arginine deiminase ArcA
444	29c04	ornithine cyclodeaminase ArcB
444a	*17g07*	*overlaps clone 29c04*
445	23c08	ornithine cyclodeaminase ArcB
446	18e07	hydroxypyruvate RDase
447	19d12	asparagine synthetase
448	19e03	agmatine ureohydrolase SpeB
449	19h06	alanine racemase
450	20d07	ornithine/acetylornithine aminoTase
451	21e02	urocanate hydratase HutU
452	21f08	adenosylhomocysteinase
453	22d08	adenosylhomocysteinase
454	09b12	phosphoglycerate DHase SerA
455	25b12	D-3-phosphoglycerate DHase
456	22e04	aminotripeptidase PepT
457	22e08	4-hydroxybenzoate hydroxylase PobA
458	22h10	chorismate mutase / prephenate dehydratase PheA
459	25c08	diaminopimelate decarboxylase LysA
460	06f04	sarcosine oxidase α SoxA
461	25e04	sarcosine oxidase α SoxA
462	01e05	sarcosine oxidase β SoxB
463	03c10	sarcosine oxidase δ SoxD
464	21h08	sarcosine DHase
465	26a01	sarcosine DHase
465a	*07f10*	*overlaps clone 26a01*
466	21b09	ferredoxin-dependent glutamate Sase GltB
467	24d04	NADH-glutamate synthase small sub. GltD
468	13e08	NADPH dependent glutamate synthase small sub. GltX
469	01h06	glutamine synthetase II GlnII
470	26f03	dihydrodipicolinate synthase DapA
471	27g06	malyl-coA lyase
472	28b05	argininosuccinate synthase ArgG
473	30a12	urease accessory prot. (UreD homolog)
474	12h11	4-aminobutyrate aminoTase
475	30e05	w-aminoTase-like prot
476	15h11	uridylyltransferase/uridylyl-removing enzyme GlnD
476a	*08e06*	*overlaps clone 15h11*
477	01e08	hydantoin racemase HyuE
Metabolism of nucleotides and nucleic acids		
478	02b08	uracil phosphoribosylTase Upp
479	02e04	formyltetrahydrofolate synthase FthfS
480	12f08	adenylate kinase (ATP-AMP transphosphorylase)
480a	*09b05*	*overlaps clone 12f08*
481	11c12	deoxyuridine 5' triphosphate nucleotidohydrolase
482	14g07	cytosine deaminase CodA
483	05e12	phosphoribosylformylglycinamidine synthase PurQ
484	17b06	phosphoribosylformylglycinamidine synthase PurQ
485	15f05	formyltetrahydrofolate deformylase-like prot. PurU
486	23e07	phosphoribosylformylglycinamidine PurL
487	29g10	thymidine kinase
Metabolism of lipids		
488	09a01	Nod Factor fatty acyl chain modification NodG
489	17c10	3-hydroxydecanoyl-acyl-carrier-prot. DTase FabA
490	03c07	fatty oxidation complex α sub. FadB
491	05h03	3-oxoacyl-acyl-carrier-prot. synthase I FabB
492	19g02	malonyl CoA-acyl carrier prot. transacylase FabD
493	10b03	3-oxoacyl-acyl carrier prot.synthase II FabF
494	22d03	3-oxoacyl-acyl carrier prot. synthase II FabF / NodE
495	15c04	3-oxoacyl-acyl-carrier-prot. synthase III FabH
496	30f01	enoyl-acyl-carrier-prot. Rdase (NADH) FabI
497	01c09	enoyl-CoA hydratase
498	05d10	3-hydroxyisobutyrate DHase
499	10h05	long-chain-fatty-acid--CoA ligase RpfB
500	05g09	acyl-coA DHase
501	15e01	acyl-coA DHase
502	17h07	3-hydroxybutyryl-CoA DHase
503	19c09	3-hydroxybutyryl-CoA DHase
504	21h04	sulfolipid biosynthesis prot. SqdA
505	22d05	sub. A of acetyl-coA carboxylase
506	01h11	acetyl-CoA carboxylTase β-sub.
Metabolism of cofactors / prosthetic groups		
507	02c03	coenzyme PQQ synthesis prot. C
508	05e02	coenzyme PQQ synthesis prot. E
508a	*03a08*	*overlaps clone 05e02*
509	03c06	DNA / panthotenate metabolism flavoprot.
510	03e04	dihydroxybenzoate DHase EntA
510a	*10a12*	*overlaps clone 03e04*
511	03g11	glutathione Tase
512	05b08	thiamine biosynthesis prot. ThiC
513	03g09	thiamine biosynthesis prot. ThiG
514	05b12	S-adenosylmethionine: 2-demethylmenaquinonemethylTase
515	12e09	cobyrinic acid a,c-diamide synthase CobB
516	11b11	precorrin isomerase CobH
517	02b03	cobalamin synthesis prot. CobN
518	05d05	cobalamin/porphyrin biosynthesis prot. CobS
519	28e08	cobalamin synthesis prot. CobT
520	10d09	glutathione S-Tase Gst
521	21d03	glutathione synthetase GshB
521a	*06b03*	*overlaps clone 21d03*
522	06e11	ferrochelatase (protoheme ferro-lyase) HemH
523	10f10	γ-glutamyltranspeptidase precursor
524	10g02	NH (3)-dependent NAD+ Sase NadE
525	11e08	riboflavin synthase, β sub. RibH
526	13e09	pu. amino acid oxidase flavoprot. ThiO
527	13e11	1-deoxyxylulose-5-phosphate Sase
528	14d08	4-hydroxybenzoate octaprenylTase (polyprenylTase)
529	18a08	7,8-diamino-pelargonic acid ATase (DAPA ATase) BioA
530	19b10	dihydroneopterin aldolase (DHNA) FolB
531	14g12	porphobilinogen deaminase precursor HemC
532	20a12	uroporphyrinogen decarboxylase HemE
533	30c03	NH(3)-dependent NAD(+) Sase NadE
533a	*26c07*	*overlaps clone 30c03*

Table 3 (continued)

Clone No.	Name	Homolog description
534	28d07	NH(3)-dependent NAD(+) Sase NadE
535	22a09	pyridoxal phosphate biosynthetic prot. PdxA
536	06a02	pyridoxamine kinase
537	24d01	glutamate 1-semialdehyde 2,1-aminomutase
538	24g03	coenzyme F390 synthetase II
539	26a02	molybdopterin biosynthesis prot.
540	29e04	pantothenate synthetase PanC
Metabolism of phosphate		
541	04f07	inorganic pyrophosphatase Ppa
541a	*29c05*	*overlaps clone 04f07*
542	25h01	phosphonate utilization PhnJ
Metabolism of rhizopine		
543	05a11	MocA oxidoreductase
544	07d03	MocB rhizopine-binding prot. precursor
545	15b08	MocB rhizopine-binding prot. precursor
546	19a12	MocB rhizopine-binding prot. precursor
547	15b06	MocC rhizopine catabolism
548	18c11	MosA rhizopine biosynthesis (dihydrodipicolinate Sase)
548a	*04a05*	*overlaps clone 18c11*
549	11g06	MocB opine catabolism (phosphogluconate DTase)
Metabolism of sulphur		
550	29b11	phospho-adenylylsulfate sulfoTase CysH
551	17a09	sulfite reductase (hemoprot. sub.) CysI
Information pathways		
DNA replication, restriction, modification and repair		
552	02f12	ribonuclease H RnhA
553	05e01	DNA polymerase III α sub. DnaE
554	28f04	DNA polymerase ε chain DnaQ
555	06d12	DNA polymerase III sub. gamma and tau DnaZX
556	08e04	DNA topoisomerase IV sub. A ParC
557	11d02	primosomal replication factor Y PriA
558	08g09	DNA gyrase sub. A (DNA topoisomerase II) GyrA
559	14e01	DNA gyrase sub. A GyrA
560	12b05	DNA gyrase sub. A GyrA
561	23b07	DNA gyrase sub. B GyrB
562	29f12	DNA gyrase sub. B GyrB
563	12g08	replication prot. RepB
563a	*02d09*	*overlaps clone 12g08*
564	24g07	DNA polymerase I. PolA
564a	*19a01*	*overlaps clone 24g07*
565	01b07	excinuclease ABC sub. A (DNA repair prot.) UvrA
566	13d07	excinuclease ABC sub. A UvrA
567	21e12	excinuclease ABC sub. A UvrA
568	25d05	excinuclease ABC sub. C UvrC
569	18g04	excinuclease ABC sub. C UvrC
570	02a02	excinuclease ABC sub. C UvrC
571	18g06	transcription-repair coupling factor Mfd
572	04d05	uracil-DNA glycosylase Ung
573	07g08	uracil-DNA glycosylase Ung
574	17d05	type I restriction-modification enzyme M sub. HsdM
575	23g05	type I restriction-modification enzyme M sub. HsdM
576	21f03	type I restriction enzyme S sub. HsdS
DNA segregation, recombination and transfer		
577	10d01	integrase/recombinase
578	11a04	integrase/recombinase
579	14b05	integrase/recombinase (y4qK pNGR234*a*)
580	30a10	recombination prot. RecA
581	19f08	conjugal transfer prot. TraA
582	21d09	secretory prot. kinase sim. to TrbB pNGR234*a*
582a	*05d11*	*overlaps clone 21d09*
582b	*24b02*	*overlaps clone 21d09*
583	15d08	VirB4-like prot., sim. to TrbeB pNGR234*a*
584	06h12	DNA- binding prot. HRm / HU (histone-like prot.)
RNA synthesis and modification		
585	07b10	transcription elongation factor GreA
586	27d08	transcription elongation factor GreA
587	27e10	RNA polymerase α sub. RpoA
588	17b03	ribonuclease HII RnhB
589	02a01	RNA polymerase β sub RpoB
590	03e09	RNA polymerase β sub RpoB
591	06d05	RNA polymerase β sub RpoB
592	22h12	RNA polymerase β sub RpoB
593	28d10	RNA polymerase β sub RpoB
594	16b02	RNA polymerase β' sub RpoC
595	04h05	RNA polymerase primary sigma-70 factor RpoD
596	03a11	RNA polymerase sigma-E factor SigH
597	21b10	RNA polymerase sigma-E factor SigC
598	27f09	RNA polymerase sigma-32 factor RpoH
599	12c03	probable sigma factor SigI
600	25f08	probable sigma factor
600a	*25h04*	*overlaps clone 25f08*
601	27c10	transcription accessory prot. Tex
602	25d11	VacB ribonuclease II family
603	24a12	reverse transcriptase/maturase
Protein synthesis and modification		
604	29h08	GTP-binding prot. (protease) HflX
605	01b01	GTP-binding prot. LepA
605a	*20h09*	*overlaps clone 01b01*
605b	*25g09*	*overlaps clone 01b01*
606	01b02	alanyl-tRNA synthetase AlarS
607	06a11	cystein- tRNA ligase CysS
608	26h03	glycyl tRNA-synthetase chain GlyQ
609	20a08	histidyl-tRNA synthetase HisS
610	21g03	leucyl-tRNA synthase LeuS
611	16h08	lysyl-tRNA synthetase LysS
612	19g05	phenylalanyl-tRNA synthetase chain PheS
613	25f01	seryl-tRNA synthetase SerS
614	29d09	tryptophan- tRNA ligase TrprS
615	03g06	tryptophan- tRNA ligase TrprS
616	10f04	tyrosyl-tRNA synthetase TyrS
617	22f04	valyl-tRNA synthetase ValS
617a	*01f12*	*overlaps clone 22f04*
618	10b10	50S ribosomal prot. L2 RplB
619	10e10	50S ribosomal prot. L4 RplD
620	03h08	50S ribosomal prot. L7/ L12 RplL
621	18e03	50S ribosomal prot. L9 RplI
622	12h01	50S ribosomal prot. L13 RplM
623	06a03	50S ribosomal prot. L14 RplN
623a	*20e04*	*overlaps clone 06a03*
624	23h06	50S ribosomal prot. L17 RplQ
625	06c05	50S ribosomal prot. L21 RplU
626	12b06	50S ribosomal prot. L22 RplV
627	29f09	50S ribosomal prot. L33 RpmG
628	17a03	30S ribosomal prot. S1 RpsA
629	25d09	30S ribosomal prot. S1 RpsA
630	13c09	30S ribosomal prot. S2 RpsB
631	29h04	30S ribosomal prot. S3 RpsC
632	17h03	30S ribosomal prot. S5 RpsE
633	13c05	30S ribosomal prot. S12 RpsL
634	22f07	30S ribosomal prot. S15 RpsQ
635	02b05	30S ribosomal prot. S17 RpsQ

Table 3 (continued)

Clone No.	Name	Homolog description
636	12a11	30S ribosomal prot. S21 RpsU
637	10f09	30S ribosomal prot. S21 RpsU
638	15f06	ribosomal prot. L11 methylTase PrmA
639	02c08	y4tL pNGR234*a*; hydrolase/ peptidase
640	27e11	clp protease ATP binding sub.
641	06h11	ATP-dependent Clp protease binding sub. ClpA
642	02e07	ATP-dependent Clp protease binding sub. ClpA
643	22f12	ATP-dependent Clp protease binding sub. ClpA
644	02f09	ATP-dependent protease regulatory ATPase sub. ClpB
645	13a07	ATP-dependent protease regulatory ATPase sub. ClpB
646	26f08	ATP-dependent protease regulatory ATPase sub. ClpB
647	04a02	serine protease, heat shock HtrA like -prot.
648	10a05	*N*-carbamyl-L-amino acid amidohydrolase AmaB
649	18e01	*N*-carbamyl-L-amino acid amidohydrolase AmAB
650	26f09	peptide chain release factor 1 PrfA
651	12a02	peptide chain release factor 3 PrfC
652	14f04	O-sialoglycoprotein endopeptidase
653	15d11	periplasmic endopeptidase RmDEGP
654	05a07	ATP-dependent RNA helicase HelO
655	16f11	ATP-dependent RNA helicase HelO
656	21f01	translation elongation factor EF-Tu
657	29d04	translation elongation factor EF-Tu
658	23f05	translation elongation factor EF-G
659	28d12	translation elongation factor EF-Ts
660	23b05	ATP-dependent Lon protease
661	01h03	L-isoaspartyl protein carboxyl methylTase
662	26c02	aminoacyl-histidine dipeptidase PepD
Regulatory functions		
663	05e05	exoenzyme regulatory prot. AepA
664	07e12	MucR transcriptional regulatory prot.; Ros/mucR family
664a	*16e11*	*overlaps clone 07e12*
665	29a08	SyrB (*syrM* repressor, sim.to y4aN, pNGR234*a*)
666	07f06	sugar fermentation stimulation prot.
667	08f09	two-component response regulator
668	25g06	transcript. regulator; XylS/AraC family
669	15g01	transcript. regulator; XylS/AraC family
670	09g10	transcript. regulator of NodD3, sim. to y4fK pNGR234*a*
671	19c04	transcript. regulator of NodD3, sim. to y4fK pNGR234*a*
672	11c01	transcript. regulator GlxA; XylS/AraC family
672a	*17g08*	*overlaps clone 11c01*
673	10d05	DNA-binding prot., sim. to y4wC/y4aM pNGR234*a*
674	06c07	adenylate cyclase CyaA
675	10h04	adenylate cyclase CyaA
675a	*17b12*	*overlaps clone 10h04*
676	27b01	transcript. regulator of sensory transduction systems
676a	*13f09*	*overlaps clone 27b01*
677	02e03	phosphate regulatory prot. PhoB
678	01d05	phosphate regulatory prot. PhoU
679	03f08	pyruvate Dhase complex repressor
680	03g07	regulatory prot. PcaR
681	05a10	catabolite control prot. A CcpA
682	08d09	acetate repressor prot.; IclR family
683	09b11	FixK regulator
684	28b10	GacA (FixJ-like) response regulator; LuxR/UhpA family
685	14c12	two-component response regulator; LuxR family
686	17g10	transcript. regulator y4qH pNGR234*a*, LuxR family
687	28c10	NifR3-like regulator
688	14d06	transcript. regulatory prot. (two-component system)
689	14h07	transcript. activator prot.
690	15c11	transcript. regulator
691	20d01	transcript. regulator; GntR family
691a	*15f07*	*overlaps clone 20d01*
692	14c08	gluconate utilization system repressor; lacI family GntR
693	07e11	transcript. regulatory prot. NtaR; GntR family
694	08h12	hydrogen peroxide-inducible activator ; lysR family OxyR
695	08c08	ribitol operon repressor; lacI family
696	14f06	transcript. repressor CytR; lacI family
697	24g11	transcript. repressor; LacI family
698	22e09	transcript. repressor; LacI family
699	28g11	transcript. repressor; LacI family
700	22b04	transcript. repressor; LacI family
701	17g03	catabolite control prot.; LacI family
702	18b10	extragenic suppressor prot. SuhB
703	22b12	extragenic suppressor prot. SuhB
704	25d07	transcript. regulator HexA; LysR family
704a	*20a07*	*overlaps clone 25d07*
705	07c07	transcript. regulator HexA; LysR family
706	10d07	transcript. regulator GstR; LysR family
707	18h01	transcript. regulator GstR; LysR family
708	11f05	transcript. regulator; LysR family
708a	*10d08*	*overlaps clone 11f05*
709	05e08	transcript. regulator y4mQ pNGR234*a*; LysR family
709a	*16c06*	*overlaps clone 05e08*
710	07c01	LysR-type β-lactamase transcriptional regulator
711	26d11	LysR-type β-lactamase transcriptional regulator
712	26g10	transcriptional regulator; LysR family
713	20b04	transcriptional regulator TrpI; LysR family
714	21c12	transcriptional regulator TrpI; LysR family
715	21f05	sigma-54 dependent transcript. activator 4_Rme
716	22a01	transcriptional modulator MgpS
717	22a05	transcriptional regulator ChvI
717a	*26f05*	*overlaps clone 22a05*
718	22d06	Lrp-like transcript. regulatory4sM pNGR234*a*
719	22g04	leucine-responsive regulatory prot.
720	24d06	phosphoTase enzyme II, A PtsN, nitrogen regulation
721	25c01	cell division response regulator DivK
722	27c07	response regulator PleD
723	25d04	transcript. regulatory4tD pNGR234*a* ; AsnC family
724	28b08	LacI-GalR family of regulators, e.g. PckR
Other categories		
Adaptation to atypical conditions and protection		
725	02b01	nodulation competitiveness prot. NfeD
726	16b08	choline DHase (osmoregulation) BetA
727	26e09	choline DHase (osmoregulation) BetA
728	19c10	betaine aldehyde DHase (osmoregulation) BetB
729	06h05	choline sulfatase (osmoregulation) BetC
729a	*08c12*	*overlaps clone 06h05*
730	09g04	choline sulfatase (osmoregulation) BetC
731	13g02	betaine-aldehyde DHase (osmoregulation)
732	04f08	acid tolerance ActA prot.
733	08c02	copper resistance prot. precursor (detoxification)
734	10f07	catalase (detoxification)
735	29c08	epoxide hydrolase (detoxification)
736	29e02	peroxidase / catalase (detoxification)
736a	*22g02*	*overlaps clone 29e02*
737	19f11	cytochrome P450 (detoxification)
738	24b03	nickel resistance prot. NreB
739	24d08	ice nucleation prot. (cold adaptation)
740	11f07	ice nucleation prot. (cold adaptation)
741	10a08	poly-β-hydroxybutyrate polymerase PhbC
742	24f07	poly-β-hydroxybutyrate polymerase PhbC
743	14g10	poly-β-hydroxybutyrate polymerase PhbC
744	12a10	survival prot. SurE
745	28e06	biotin-regulated prot. BioS

Table 3 (continued)

Clone No.	Name	Homolog description	
746	15b05	6'-aminoglycoside (kanamycin 6')-N-acetylTase AacA	
747	19d11	arsenate reductase (arsenical pump modifier) ArsC	
748	05a05	5'-hydroxystreptomycin biosynthesis prot. StrU	
749	28e07	haloperoxidase	
Transposon-related functions			
750	25f04	ATP-binding prot. y4bM/kI/tA pNGR234*a*	NGRIS-3
751	26c12	ATP-binding prot. y4bL/kJ/tB pNGR234*a*	NGRIS-3
752	22b05	y4bA/pH pNGR234*a*	NGRIS-4
753	06e03	y4bA/pH pNGR234*a*	NGRIS-4
753a	*26e03*	*overlaps clone 06e03*	
754	11c03	y4bA/pH pNGR234*a*	NGRIS-4
755	29f05	y4bA/pH pNGR234*a*	NGRIS-4
756	12a08	y4bB/pI pNGR234*a*	NGRIS-4
757	07a07	y4bC/pJ pNGR234*a*	NGRIS-4
757a	*20b11*	*overlaps clone 07a07*	
758	11a02	y4bD/pK pNGR234*a*	NGRIS-4
759	23f01	y4ba/pH pNGR234*a* (NGRRS-1a left)	NGRIS-4
760	02e12	transposase y4jA/nE/sE pNGR234*a*	NGRIS-5
761	11d01	ISRm2011-2 transposase (IS630-Tc1 family)	
762	06d03	ISRm2011-2 transposase (IS630-Tc1 family)	
763	17c03	ISRm2011-2 transposase (IS630-Tc1 family)	
763a	*28d04*	*overlaps clone 17c03*	
764	03f02	transposase IS1380	
765	25h02	transposase IS1380	
765a	*17f06*	*overlaps clone 25h02*	
766	05h02	transposase IS1594	
767	17h11	transposase IS200	
768	26g05	ATP-binding prot. y4iQ/nD/sD pNGR234*a*	NGRIS-5
769	17e12	IS1248b orf1; sim. to frag. fs4 pNGR234*a*	NGRIS-9
770	13h02	IS869 orf1; sim. to frag. fs4 pNGR234*a*	NGRIS-9
771	07h02	transposase y4sN pNGR234*a*	NGRIS-9
772	28h05	IS427 orf4; sim. to y4sN pNGR234*a*	
773	18e04	transposase IS870.1	
774	19e10	RFRS9 25 kDa prot.	
774a	*04e10*	*overlaps clone 19e10*	
775	15d07	transposase y4bF pNGR234*a*	
776	18f10	transposase y4qJ pNGR234*a*	
776a	*06d11*	*overlaps clone 18f10*	
777	18c09	transposase y4qJ pNGR234*a*	
777a	*22a11*	*overlaps clone 18c09*	
778	17g01	IS110 family transposase y4uE pNGR234*a*	
779	25e06	IS110 family transposase y4uE pNGR234*a*	
780	28g06	IS110 family transposase y4uE pNGR234*a*	
781	11d04	IS110 transposase/integrase (C-term)	
782	02e10	H- repeat associated prot.	
783	27d02	H- repeat associated prot.	
784	06e10	IS-related y4hQ	
785	26h04	IS-related y4hQ	
786	12b01	IS-related y4hP	
787	12e04	IS-related y4qI	
787a	*11h07*	*overlaps clone 12e04*	
788	30b11	IS-related y4qI	
788a	*05h12*	*overlaps clone 30b11*	
789	04d03	IS-related y4gE	
790	20b02	IS-related y4rI	
790a	*25d08*	*overlaps clone 20b02*	
Phage-related functions			
791	12h09	symbiosis island integrase (phage P4 family)	
791a	*12h08*	*overlaps clone 12h09*	
792	17b04	bacteriophage P22 DNA packaging prot. GP2	
792a	*06h08*	*overlaps clone 17b04*	
793	28c01	bacteriophage P22 DNA packaging prot. GP2	
794	15a01	phage T7 internal virion prot. D	
Miscellaneous			
795	27b07	siroheme synthetase-like prot. CysG	
796	09h11	indoleacetamide hydrolase (auxin biosynthesis)	
797	16e08	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase HpcH	
798	18c10	NifS-like prot.	
799	12h07	serine/threonine prot. phosphatase	
800	29f06	y4vD pNGR234*a*, peroxiredoxin 2 family	
801	22b03	y4wM pNGR234*a*, possible binding-prot	
802	14c03	MelA, melanin synthesis; 4HPPD family	
803	14a03	aldehyde DHase	
804	07d01	aldehyde DHase	
805	02f01	aldehyde DHase	
806	05g03	aldehyde DHase	
807	03e10	aldehyde DHase	
808	07c02	aldehyde DHase	
809	12h02	betaine / aldehyde DHase	
810	09h07	betaine / aldehyde DHase	
811	29b08	oxidoRDase , sim. to various polyketide synthase	
812	14h02	molybdenum-containing aldehyde oxidoRDase	
813	01e11	oxidoRDase (short-chain type DHase/ RDase)	
814	06g07	oxidoRDase (short-chain type DHase/ RDase)	
815	18d03	y4eL pNGR234*a*, short-chain type DHase/ RDase	
816	06b08	short-chain DHase homolog	
817	24c07	oxidoRDase	
818	16c05	oxidoRDase	
819	12g04	oxidoRDase	
820	07f05	NADH-dependent flavin oxidoRDase	
821	04e11	NADH-dependent flavin oxidoRDase	
822	18a09	2-hydroxyacid DHase	
822a	*21c10*	*overlaps clone 18a09*	
823	11b06	chlorophenol-4-monooxygenase component 1	
824	09b01	phenylacetic acid degradation prot.	
825	12c05	phenylacetic acid degradation prot.	
826	24g08	phenylacetic acid degradation prot.	
827	09b09	export prot.	
828	16h05	potential multicopper oxidase	
829	24h11	2-hydroxyhepta-2,4-diene-1,7-dioate isomerase	
830	26d12	2-hydroxyhepta-2,4-diene-1,7-dioate isomerase	
831	09a06	ferredoxin RDase (naphthalene conversion)	
832	11f11	Carboxymuconolactone decarboxylase	
833	07b07	biotin / pyruvate carboxylase	
834	22h07	GTP-binding prot	
835	22b09	L-sorbose DHase (GMC oxidoRDase family)	
836	10e01	L-sorbose DHase (GMC oxidoRDase family)	
837	02d05	L-sorbose DHase FAD dependent	
838	24c08	carbon monoxide DHase medium sub.	
839	22f10	D-arabino 3-hexulose 6-P formaldehyde lyase	
840	27e05	NADH-dependent DHase homolog	
841	28b07	molybdenum-containing quinoline 2-oxidoRDase	
842	20e01	DHase sub. precursor	
843	20g03	pterin-4a-carbinolamine DTase	
844	21d04	contains hemolysin-type calcium-binding domain	
845	16e10	sulfate-starvation induced prot	
Unknown proteins of: (primary accession number)			
Escherichia coli			
846	15g10	P77388	
847	02a07	P32683	
848	02a09	P33362	
849	03e01	P45528	

Table 3 (continued)

Clone No.	Name	Homolog description
850	03h01	P76631
851	05g07	P76481
852	06h04	P75774
853	07f09	P37007
854	07h06	Q46890
855	08b02	P21498
856	08b06	BAA14942
857	08d12	P39333
858	13e12	P76641
858a	*09c08*	*overlaps clone 13e12*
859	10a03	P45568
860	10d02	P37675
861	10d04	P42901
862	10f12	P76481
863	12c11	P45475
864	12g02	AAC75037
865	15e04	P52049
866	15f04	P77748
867	16f04	BAA31826
868	17c04	P23522
869	18g09	AAA83544
870	18h06	AAC74284
871	18h11	P77368
872	19h07	P77165
873	20a06	P37619
874	22d02	AAC74824
875	22e07	AAC44004
876	22f08	P33362
877	25f09	P76397
878	28e04	P76397
879	25f10	P77470
880	26e05	AAC74522
881	27a05	AAC75727
882	27e02	P39829
883	09c12	P37053
884	28f07	AAC75038
885	17d02	P77391
886	24f10	P76235
887	28f08	P76235
888	28a06	P39338
889	30e08	P08367
890	27c11	P77165
891	02b12	P75844
892	24h06	P77334
893	22h02	P46854
893a	*21d08*	*overlaps clone 22h02*
Mycobacterium tuberculosis		
894	02h04	O05841
895	02h08	O06320
895a	*08h06*	*overlaps clone 02h08*
896	03d06	O05865
897	04f10	O53858
898	04h01	O06804
899	07e03	P71838
900	19b02	Q10846
901	19f01	P96814
902	20c01	O50466
903	24h02	O07220
904	25a05	O06235
905	25b06	P71984
906	25h05	O53203
907	27h05	Q10849
908	28a02	P96267
909	08e12	Q50709
910	22g11	Q11157
911	10f08	O69646
912	11a08	O07756
913	15c10	O06378
914	16f03	P95223
915	18e10	P96914
916	21g05	P72043
917	25f02	P95223
918	29h03	O53720
Bacillus subtilis		
919	16d09	O34932
919a	*02a06*	*overlaps clone 16d09*
920	05a01	P94437
920a	*05g06*	*overlaps clone 05a01*
921	06c04	P94937
922	06e01	AAB72069
923	04h03	O32272
924	07f08	AAB35255
925	08a08	BAA06611
926	17e06	BAA23396
927	19c11	P54724
928	11d08	P54178
929	11b08	P39640
929a	*08g04*	*overlaps clone 11b08*
930	11f08	P96683
931	09c03	P42966
932	17f04	O34398
933	24b07	P35155
934	18c06	O05220
935	24d11	Q07835
936	13a01	O07618
937	25a06	P37508
938	21a05	Q45584
938a	*12b10*	*overlaps clone 21a05*
Synechocystis sp.		
939	04a01	BAA17151
940	06c03	BAA17443
941	04h11	BAA18318
942	09g07	BAA18319
943	13c07	BAA18330
944	08e11	BAA16904
945	08f05	BAA17017
946	10e06	BAA17950
947	10f05	BAA16766
948	15h10	BAA18186
949	25e03	AAB41278
950	26c01	P72872
950a	*11h09*	*overlaps clone 26c01*
951	29a07	BAA10710
952	01c06	BAA10835
Haemophilus influenzae		
953	01c01	P44250
954	01e07	P31777
955	06b02	Q57151
956	13a08	P44093
957	25h09	P44886
958	01h02	P44540
959	22c12	P44543
960	19a07	Q57184

Table 3 (continued)

Clone No.	Name	Homolog description
Agrobacterium sp.		
961	03a07	AAB91569
962	05f01	AAB67296
963	28b11	AAB67297
964	06b11	AAB51512
965	19e06	AAC17194
966	09h12	AAB67297
967	17e09	P70791
968	16h09	P70795
969	01g12	P70795
970	22b02	P70795
Rhizobia		
971	17f01	P55362
972	30f11	P55362
973	16c11	P55388
974	30e04	P55424
975	27b10	P55480
976	10b09	P55552
977	12g11	P55552
978	29e06	P55552
979	11d10	P55590
980	07c10	P55694
981	20h08	P55706
982	02g08	P25893
983	11f12	P49305
984	14b08	AAB63673
985	16b12	AAA74241
986	16f09	AAA74241
987	16c07	AAB4153
988	09g03	Q52991
989	07h12	P25893
990	22h03	AAA88525
991	25c11	Q52967
992	25g03	AAB81416
993	20a04	CAA11961
994	30e03	CAB01954
995	20g09	AAC64871
996	04f06	AAB17515
997	22h04	AAB17515
998	19e07	AAB17514
999	05f08	AAA96138
1000	24e08	O69244
Other organisms		
1001	19a11	AAC16153
1002	28a01	AAC16139
1003	22d01	P30790
1004	03g04	Q06373
1005	07d10	Q58322
1006	05a02	Q57883
1006a	*29e09*	*overlaps clone 05a02*
1006b	*17d09*	*overlaps clone 05a02*
1007	06f06	Q46063
1008	07b09	AAB50572
1009	02h03	AAB50572
1009a	*12e05*	*overlaps clone 02h03*
1010	04e08	AAC46053
1011	28h08	AAC46056
1012	19f10	AAA96787
1013	10g08	BAA29686
1014	02g03	BAA29099
1014a	*13b10*	*overlaps clone 02g03*
1015	11g07	AAC82835
1016	17e05	P46378
1017	17h12	AAB51777
1018	18a11	CAA55879
1019	18h10	AAB66497
1020	19f09	AAB85316
1021	19f12	O52867
1021a	*23b03*	*overlaps clone 19f12*
1022	21a11	AAB38705
1023	21d12	AAB09035
1024	12d11	AAD03878
1025	21a04	AAC44077
1026	22f06	AAD03845
1027	16d06	AAD03912
1028	22g06	P70734
1029	22h06	AAC46243
1030	05d03	AAC06984
1030a	*06d06*	*overlaps clone 05d03*
1031	24b01	AAC07457
1032	30a07	AAC06721
1033	01a12	P38102
1034	24d02	P55176
1035	28e05	P29938
1036	09c10	AAC44553
1037	26a12	AAB89525
1038	27f04	AAC34291
1039	29c07	C36925
1040	29d11	Q49092
1041	11d06	Q15595
1041a	*07a02*	*overlaps clone 11d06*
1041b	*23g10*	*overlaps clone 11d06*
1042	06h07	P40896
1043	24d09	P34227
1043a	*15f12*	*overlaps clone 24d09*
1043b	*18f08*	*overlaps clone 15f12*

Abbreviations: No, number; prot., protein; sim., similar; sub, subunit; transcript., transcriptional; transp., transport; ATase, aminotransferase; CoA, coenzyme A; DHase, dehydrogenase; DTase, dehydratase; RDase, reductase; Sase, synthase; Tase, transferase; TGase, transglycosylase.

Conclusions

Random sequencing of ANU265 followed by homology searches of public databases resulted in the identification of 1,130 putative protein-coding sequences, of which 922 (41%) could be classified into functional groups. Comparison of these data with those derived from the complete sequence of the *B. subtilis* genome showed a similar distribution of putative coding sequences, except perhaps for functions related to transposable elements (Table 2). In fact, the genome of ANU265 carries more putative transposases and other IS-related functions (5.5% of all identified genes, and 2.2% of all shotgun sequences) than that of *B. subtilis*. Nevertheless, in proportion to their size, the chromosome and megaplasmid of NGR234 carry fewer IS sequences than pNGR234*a*. Furthermore, hybridization data indicate that the density of known transposable elements is higher in pNGR234*b* than on the chromosome (order of IS accretion is: pNGR234*a* > pNGR234*b* > chromosome) [11]. This suggests that IS

elements preferentially accumulate on plasmids, possibly because they are less likely to disrupt essential functions. In contrast, the many RIME elements present in NGR234 are clearly more abundant on the chromosome and megaplasmid than on pNGR234*a*. Together, the distinct G+C contents and structural features of the symbiotic plasmid, megaplasmid and chromosome suggest that different evolutionary constraints and histories contributed to shape these three replicons.

'Skimming' the genome of *Rhizobium* sp. NGR234 has given new insights into the evolution of its replicons and the integration of symbiotic functions in the genome of a soil bacterium. It also reinforced the assumption, which originated from host-range extension experiments [12,47], that pNGR234*a* carries most of the symbiotic genes. Although few *nod*, *nif* and *fix* homologs were found amongst the random clones, it is likely that additional chromosome- and megaplasmid-encoded functions contribute to successful symbioses between NGR234 and its many host plants. In this respect, transcriptional analyses using shotgun sequences as hybridization templates [11] will help identify such new symbiotic loci.

Materials and methods

Microbiological techniques

Rhizobium strain ANU265 [19], a strain of *Rhizobium* sp. NGR234 [7] cured of pNGR234*a*, was grown in Rhizobium minimal medium supplemented with succinate (RMM) [47]. *Escherichia coli* was grown on SOC or in TY [48]. Subclones in M13mp18 vectors [49] were grown in *E. coli* strain DH5αF'IQ [50].

Preparation of the random genomic library and M13 templates

Genomic DNA of *Rhizobium* strain ANU265 was prepared as in Perret and Broughton [51]. ANU265 genomic DNA (15 μg) was sheared by sonication and incubated for 10 min at 30°C with 30 units of mung bean nuclease. The resulting digest was extracted with phenol/chloroform (1:1) and precipitated with ethanol. Fragments ranging in size from 900 to 1,500 bp were purified from agarose gels and ligated into *Sma*I-digested M13mp18 vector DNA. Ligation mixtures were electroporated into *E. coli* strain DH5αF'IQ [48,52], and transformants were plated on 5-bromo-4-chloro-indoyl-β-D-galactoside (X-Gal) and isopropyl-β-thiogalactopyranoside (IPTG)-containing petri dishes [48]. Fresh 1 ml cultures of *E. coli* DH5αF'IQ were infected with phages from randomly selected white plaques, and grown for 6 h at 37°C in TY medium. Phages were precipitated from 600 μl of the culture supernatant by adding 150 μl 2.5 M NaCl/20% polyethylene glycol (PEG-8,000) (20 min at 25°C). Afterwards, they were centrifuged for 20 min at 3,000*g* at 25°C, and resuspended in 20 μl Triton-TE extraction buffer (0.5% Triton X-100; 10 mM Tris-HCl, 1 mM EDTA pH 8.0). Following 10 min incubation at 80°C and ethanol precipitation, single-stranded phage DNA was recovered in 50 μl H_2O.

Sequence analysis

Dye-terminator cycle sequencing of individual M13 subclones, gel electrophoresis and sequence editing was performed as described by Freiberg *et al.* [53]. Shotgun sequences were checked for redundancy using the XGAP program [54] and for significant homologies with BLASTX-BLASTN software [55] using nonredundant databases at NCBI [25].

Acknowledgements

We thank S. Brenner, C. Freiberg, S. Taudien and D. Gerber for their help with many aspects of this work. Financial assistance was provided by the Fonds National Suisse de la Recherche Scientifique (Grant No. 31-45921.95) and the Université de Genève.

References

1. Broughton WJ, Perret X: **Genealogy of legume-*Rhizobium* symbioses.** *Curr Opin Plant Biol* 1999, **2:**305-311.
2. Perret X, Staehelin C, Broughton WJ: **Molecular basis of symbiotic promiscuity.** *Microbiol Mol Biol Rev* 2000, **64:**180-201.
3. Fischer HM: **Genetic regulation of nitrogen fixation in rhizobia.** *Microbiol Rev* 1994, **58:**352-386.
4. Martínez-Romero E, Caballero-Mellado J: ***Rhizobium* phylogenies and bacterial genetic diversity.** *Critical Rev Plant Sci* 1996, **15:**113-140.
5. Kündig C, Hennecke H, Göttfert M: **Correlated physical and genetic map of the *Bradyrhizobium japonicum* 110 genome.** *J Bacteriol* 1993, **175:**613-622.
6. Sullivan JT, Ronson CW: **Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a *phe*-tRNA gene.** *Proc Natl Acad Sci USA* 1998, **95:**5145-5149.
7. Trinick MJ: **Relationships amongst the fast-growing rhizobia of *Lablab purpureus*, *Leucaena leucocephala*, *Mimosa* spp., *Acacia farnesiana*, and *Sesbania grandiflora* and their affinities with other rhizobial groups.** *J Appl Bacteriol* 1980, **49:**39-53.
8. Pueppke SG, Broughton WJ: ***Rhizobium* sp. strain NGR234 and *R. fredii* USDA257 share exceptionally broad, nested host ranges.** *Mol Plant-Microbe Interact* 1999, **12:**293-318.
9. Freiberg C, Fellay R, Bairoch A, Broughton WJ, Rosenthal A, Perret X: **Molecular basis of symbiosis between *Rhizobium* and legumes.** *Nature* 1997, **387:**394-401.
10. Flores M, Mavingui P, Girard L, Perret X, Broughton WJ, Martinez-Romero E, Davila G, Palacios R: **Three replicons of *Rhizobium* sp. strain NGR234 harbor symbiotic gene sequences.** *J Bacteriol* 1998, **180:**6052-6053.
11. Perret X, Viprey V, Broughton WJ: **Physical and genetic analysis of the broad host-range *Rhizobium* sp. NGR234.** In *Prokaryotic Nitrogen Fixation*. Edited by Triplett EW. Wymondham: Horizon Scientific Press, 2000, 679-692.
12. Broughton WJ, Heycke N, Meyer zAH, Pankhurst CE: **Plasmid linked *nif* and *nod* genes in fast-growing rhizobia that nodulate *Glycine max*, *Psophocarpus tetragonolobus*, and *Vigna unguiculata*.** *Proc Natl Acad Sci USA* 1984, **81:**3093-3097.
13. Perret X, Freiberg C, Rosenthal A, Broughton WJ, Fellay R: **High-resolution transcriptional analysis of the symbiotic plasmid of *Rhizobium* sp. NGR234.** *Mol Microbiol* 1999, **32:**415-425.
14. Perret X, Broughton WJ, Brenner S: **Canonical ordered cosmid library of the symbiotic plasmid of *Rhizobium* species NGR234.** *Proc Natl Acad Sci USA* 1991, **88:**1923-1927.
15. ***Ralstonia solanacearum* - Complete genome sequencing** [http://www.genoscope.cns.fr/externe/English/Projets/Projet_Y/Y.html]
16. ***Sinorhizobium meliloti* strain 1021 sequencing projects** [http://sequence.toulouse.inra.fr]
17. ***Xanthomonas citri* Genome Project** [http://genoma4.iq.usp.br/xanthomonas]

18. Brenner S, Elgar G, Sandford R, Macrae A, Venkatesh B, Aparicio S: **Characterization of the pufferfish (*Fugu*) genome as a compact model vertebrate genome.** *Nature* 1993, **366:**265-268.
19. Avalos J, Corrochano LM, Brenner S: **Genomic organization of the fungus *Phycomyces*.** *Gene* 1996, **174:**43-50.
20. Kim CW, Markiewicz P, Lee JJ, Schierle CF, Miller JH: **Studies of the hyperthermophile *Thermotoga maritima* by random sequencing of cDNA and genomic libraries. Identification and sequencing of *trpEG(D)* operon.** *J Mol Biol* 1993, **231:**960-981.
21. Peterson SN, Schramm N, Hu PC, Bott KF, Hutchison CA: **A random sequencing approach for placing markers on the physical map of *Mycoplasma genitalium*.** *Nucleic Acids Res* 1991, **19:**6027-6031.
22. Morrison NA, Hau CY, Trinick MJ, Shine J, Rolfe BG: **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* 1983, **153:**527-531.
23. Broughton WJ, Dilworth MJ, Passmore IK: **Base ratio determination using unpurified DNA.** *Anal Biochem* 1972, **46:**164-172.
24. Perret X, Viprey V, Freiberg C, Broughton WJ: **Structure and evolution of NGRRS-1, a complex, repeated element in the genome of *Rhizobium* sp. strain NGR234.** *J Bacteriol* 1997, **179:**7488-7496.
25. **BlastX and BlastN analyses on non-redundant nucleic and aminoacids databases** [http://www.ncbi.nlm.nih.gov/BLAST/]
26. Østeras M, Stanley J, Finan TM: **Identification of *Rhizobium*-specific intergenic mosaic elements within an essential two-component regulatory system of *Rhizobium* species.** *J Bacteriol* 1995, **177:**5485-5494.
27. Lupski JR, Weinstock GM: **Short, interspersed repetitive DNA sequences in prokaryotic genomes.** *J Bacteriol* 1992, **174:**4525-4529.
28. **Swissprot protein database** [http://expasy.ch]
29. Kunst F, Ogasawara N, Moszer I, Albertini AM, Alloni G, Azevedo V, Bertero MG, Bessieres P, Bolotin A, Borchert S, *et al.*: **The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.** *Nature* 1997, **390:**249-256.
30. Blattner FR, Plunkett G 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, *et al.*: **The complete genome sequence of *Escherichia coli* K-12.** *Science* 1997, **277:**1453-1474.
31. Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, Bult CJ, Kerlavage AR, Sutton G, Kelley JM, *et al.*: **The minimal gene complement of *Mycoplasma genitalium*.** *Science* 1995, **270:**397-403.
32. Münchbach M, Nocker A, Narberhaus F: **Multiple small heat shock proteins in rhizobia.** *J Bacteriol* 1999, **181:**83-90.
33. Streit WR, Joseph CM, Phillips DA: **Biotin and other water-soluble vitamins are key growth factors for alfalfa root colonization by *Rhizobium meliloti* 1021.** *Mol Plant-Microbe Interact* 1996, **9:**330-338.
34. Streit WR, Phillips DA: **A biotin-regulated locus *bioS*, in a possible survival operon of *Rhizobium meliloti*.** *Mol Plant-Microbe Interact* 1997, **10:**933-937.
35. Miranda-Rios J, Morera C, Taboada H, Davalos A, Encarnacion S, Mora J, Soberon M: **Expression of thiamin biosynthetic genes (*thiCOGE*) and production of symbiotic terminal oxidase *cbb3* in *Rhizobium etli*.** *J Bacteriol* 1997, **179:**6887-6893.
36. Cheng HP, Walker GC: **Succinoglycan production by *Rhizobium meliloti* is regulated through the ExoS-ChvI two-component regulatory system.** *J Bacteriol* 1998, **180:**20-26.
37. Grob P, Michel P, Hennecke H, Göttfert M: **A novel response-regulator is able to suppress the nodulation defect of a *Bradyrhizobium japonicum nodW* mutant.** *Mol Gen Genet* 1993, **241:**531-541.
38. Murphy PJ, Saint CP: **Rhizopines in the legume-*Rhizobium* symbiosis.** In *Molecular Signals in Plant-Microbe Communications.* Edited by Verma DPS. Boca Raton, Florida: CRC Press, 1992, 377-390.
39. Rossbach S, Rasul G, Schneider M, Eardly B, de Bruijn FJ: **Structural and functional conservation of the rhizopine catabolism (*moc*) locus is limited to selected *Rhizobium meliloti* strains and unrelated to their geographical origin.** *Mol Plant-Microbe Interact* 1995, **8:**549-559.
40. Pocard JA, Vincent N, Boncompagnie E, Smith LT, Poggi MC, Le Rudulier D: **Molecular characterization of the *bet* genes encoding glycine betaine synthesis in *Sinorhizobium meliloti* 102F34.** *Microbiol* 1997, **143:**1369-1379.
41. Willis LB, Walker GC: **The *phbC* (poly-beta-hydroxybutyrate synthase) gene of *Rhizobium* (*Sinorhizobium*) *meliloti* and characterization of *phbC* mutants.** *Can J Microbiol* 1998, **44:**554-564.
42. York GM, Walker GC: **The *Rhizobium meliloti exoK* gene and *prsD/prsE/exsH* genes are components of independent degradative pathways which contribute to production of low-molecular-weight succinoglycan.** *Mol Microbiol* 1997, **25:**117-134.
43. Finnie C, Zorreguieta A, Hartley NM, Downie JA: **Characterization of *Rhizobium leguminosarum* exopolysaccharide glycanases that are secreted via a type I exporter and have a novel heptapeptide repeat motif.** *J Bacteriol* 1998, **180:**1691-1699.
44. Sutton JM, Peart J, Dean G, Downie JA: **Analysis of the C-terminal secretion signal of the *Rhizobium leguminosarum* nodulation protein NodO; a potential system for the secretion of heterologous proteins during nodule invasion.** *Mol Plant-Microbe Interact* 1996, **9:**671-680.
45. Finnie C, Hartley NM, Findlay KC, A. DJ: **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* 1997, **25:**135-146.
46. Viprey V, Del Greco A, Golinowski W, Broughton WJ, Perret X: **Symbiotic implications of type III protein secretion machinery in *Rhizobium*.** *Mol Microbiol* 1998, **28:**1381-1389.
47. Broughton WJ, Wong CH, Lewin A, Samrey U, Myint H, Meyer zAH, Dowling DN, Simon R: **Identification of *Rhizobium* plasmid sequences involved in recognition of *Psophocarpus*, *Vigna*, and other legumes.** *J Cell Biol* 1986, **102:**1173-1182.
48. Sambrook J, Fritsch EF, Maniatis T: *Molecular Cloning: A Laboratory Manual*, second edn. Cold Spring Harbor, NY: Cold Spring Harbor University Press, 1989.
49. Yanisch-Perron C, Ira J, Messing J: **Improved M13 phage cloning vectors and host strains: Nucleotide sequences of M13mp18 and pUC19 vectors.** *Gene* 1985, **33:**103-119.
50. Hanahan D: **Studies on transformation of *Escherichia coli* with plasmids.** *Mol Biol* 1983, **166:**557-580.
51. Perret X, Broughton WJ: **Rapid identification of *Rhizobium* strains by targeted PCR fingerprinting.** *Plant Soil* 1998, **204:**21-34.
52. Dower WJ, Miller JF, Ragsdale CW: **High efficiency transformation of *E. coli* by high voltage electroporation.** *Nucleic Acids Res* 1988, **16:**6127-6145.
53. Freiberg C, Perret X, Broughton WJ, Rosenthal A: **Sequencing the 500-kb GC-rich symbiotic replicon of *Rhizobium* sp. NGR234 using dye terminators and a thermostable "sequenase": a beginning.** *Genome Res* 1996, **6:**590-600.
54. Dear S, Staden R: **A sequence assembly and editing program for efficient management of large projects.** *Nucleic Acids Res* 1991, **19:**3907-3911.
55. Altschul SF, Gish W, Miller W, Myers EW, Lipman D: **Basic local alignment search tool.** *J Mol Biol* 1990, **215:**403-410.