Phylogenetic clusters of rhizobia revealed by genome structures

ZHENG Junfang[1], LIU Guirong[1,2], ZHU Wanfu[1], ZHOU Yuguang[3] & LIU Shulin[1,2]

1. Department of Microbiology, School of Basic Medical Sciences, Peking University, Beijing 100083, China;
2. Department of Microbiology and Infectious Diseases, University of Calgary, Calgary, ABT2N 4N1, Canada;
3. Institute of Microbiology, Chinese Academy of Sciences, Beijing 100080, China
Correspondence should be addressed to Liu Shulin (email: slliu@ucalgary.ca)

Received June 26, 2003

Abstract Rhizobia, bacteria that fix atmospheric nitrogen, are important agricultural resources. In order to establish the evolutionary relationships among rhizobia isolated from different geographic regions and different plant hosts for systematic studies, we evaluated the use of physical structure of the rhizobial genomes as a phylogenetic marker to categorize these bacteria. In this work, we analyzed the features of genome structures of 64 rhizobial strains. These rhizobial strains were divided into 21 phylogenetic clusters according to the features of genome structures evaluated by the endonuclease I-*Ceu*I. These clusters were supported by 16S rRNA comparisons and genomic sequences of four rhizobial strains, but they are largely different from those based on the current taxonomic scheme (except 16S rRNA).

Keywords: rhizobia, phylogeny, genome structure.

DOI: 10.1360/03yc0132

Rhizobia are agriculturally and environmentally important bacteria. Their symbiosis with leguminous plants is responsible for most of the atmospheric nitrogen fixed on land. Classification of these bacteria based on their natural relationships will promote their application. This research employs a new phylogenetic method, i.e., revelation and comparison of genome structure, to categorize rhizobia. Phylogeny is the study of the evolutionary relationships among organisms[1]. Currently phylogenetic relationships among rhizobia are mostly inferred from comparisons of 16S rRNA gene sequences. However, horizontal gene transfer involving 16S rRNA sequences among bacteria in different species or of greater evolutionary distances may confuse the analysis, making the phylogenetic relationships thus deduced not accurate or reliable[2,3]. We then turned to the entire genomes, attempting to find other conservative features that can be revealed by novel methods, which might be supplementary to, and confirmative for, the 16S rRNA methodology.

The analysis of bacterial complete genome sequence is obviously the most ideal method for studying molecular phylogeny for the accuracy and completeness of the information it can provide. However, exploration of the general rules of bacterial genomic divergence and evolution requires analysis and comparison of large numbers of bacteria, which makes it unrealistic to employ whole genome sequencing as the primary method for its cost. In this report, we describe the I-*Ceu*I method, which utilizes the global information of the genome and reveals the overall physical structure of the genome, for phylogenetic studies of the rhizobia. I-*Ceu*I is an endonuclease encoded by a

group I intron in the large subunit rRNA gene of *Chlamydomonas eugametos*[4,5]. It cleaves DNA in a 26 bp sequence in the gene coding for 23S rRNA, *rrl*, which often exists together with other rRNA genes as an operon (*rrn* operon). rRNA sequences are highly conserved in evolution across all life forms, so we hypothesized that *rrn* operons might be conserved also in copy number and genomic location. If so, I-*Ceu*I can reveal bacterial phylogenetic relationship by comparing the copy number and genomic location of their *rrn* operons which are the important parameters of their genome structures.

1 Materials and methods

1.1 Bacterial strains

The rhizobial strains used in this study, listed in table 1, were isolated from different localities over different periods of time or obtained from other investigators. They were stocked at Culture Collection of Beijing Agricultural University (CCBAU) in 20% (*v*/*v*) glycerol at –80℃ and were cultured in YMA medium at 28℃[6]. Included in this study were 26 type strains representing the described rhizobial species, 4 strains whose genomic sequences are completed, 13 reference strains and 25 other isolates as detailed in table 1.

1.2 Enzymes and chemicals

I-*Ceu*I was purchased from New England Bio-Labs. Proteinase K was from Roche. Most other chemicals were from the Sigma Chemical Co.

1.3 I-*Ceu*I methods

Intact genomic DNA was prepared in agarose

Table 1 Bacterial strains

Strain	Host	Origin	Reference	GenBank access no.
Agrobacterium tumefaciens C_{58} [a]	*Prunus pseudocerasus*	USA	Wood et al. 2001	AE008688
A. tumefaciens AS 1.1415 [b]		USA		
A. tumefaciens AS 1.1416 [c]		USA		
A. tumefaciens AS 1.1488 [c]		France		
A. tumefaciens AS 1.1602 [c]		Germany		
A. tumefaciens AS 1.1603 [c]		Germany		
A. tumefaciens IAM 12048	unknown	Holland		AJ389904
Allorhizobium undicola LMG 11875 [c]	*Neptunia natans*	Belgium	de Lajudie et al. 1998	Y17047
Bradyrhizobium elkanii USDA 76 [b]	*Glycine max*	USA	Kuykendall et al. 1992	U35000
Bradyrhizobium japonicum USDA 110 [a]	*Glycine max*	USA	Kuykendall et al. 1992;	BA000040
B. japonicum AS 1.826 [c]	*Glycine max*	Shenyang	Kaneko 2002	
B. japonicum AS 1.828 [c]	*Glycine max*	Shenyang		
Mesorhizobium huakuii A 106	*Astragalus sinicus*	Hubei	Chen et al. 1991	
M. huakuii CCBAU 2609 [b]	*Astragalus sinicus*	Nanjing	Chen et al. 1991	D12797
M. huakuii PL-52	*Astragalus sinicus*	Hubei	Chen et al. 1991	
Mesorhizobium loti MAFF 303099 [a]	*Lotus*	Japan	Kaneko et al. 2000	BA000012
Mesorhizobium mediterraneum USDA 3392 [b]	*Cicer arietinum*	Spain	Nour et al. 1995	L38825
Mesorhizobium plurifarium USDA 4413	*Acacia senegal*	Senegal	De Lajudie et al. 1994	
Mesorhizobium tianshanense CCBAU 3306 [b]	*Glycyrrhiza pallidiflora*	Xinjiang	Chen et al. 1995	
Rhizobium etli CFN 42 [b]	*Phaseolus vulgaris*	Mexico	Segovia 1993	U28916
Rhizobium galegae HAMBI 1185	*Galega orientalis*	UK	Lindstrom et al. 1989	
R. galegae HAMBI 503	*Galega orientalis*	USA	Lindstrom et al. 1989	
R. galegae HAMBI 540 [b]	*Galega orientalis*	Finland	Lindstrom et al. 1989	
Rhizobium gallicum FL 27	*Phaseolus vulgaris*	Mexico	Laguerre et al. 1994	
R. gallicum USDA 2918 [b]	*Phaseolus vulgaris*	France	Laguerre et al. 1994; Amarger et al. 1997	U86343
Rhizobium giardinii USDA 2914 [b]	*Phaseolus vulgaris*	France	Laguerre et al. 1994; Amarger et al. 1997	U86344
Rhizobium hainanense CCBAU 57003	*Desmodium gyroides*	Hainan		

(*To be continued on the next page*)

(*Continued*)

Strain	Host	Origin	Reference	GenBank access no.
R. hainanense CCBAU 57015[b)]	*Desmodium sinuatum*	Hainan	Chen et al. 1997	U71078
Rhizobium huautlense S 02[b)]	*Sesbania herbacea*	Mexico	Wang et al., 1998	AF025852
Rhizobium indigoferae CCBAU 71042[b)]	*Indigofera amblyantha*	Shaanxi	Wei et al. 2002	AF364068
Rhizobium leguminosarum AS 1.167[c)] (*bv. trifolii*)		Beijing		
R. leguminosarum AS 1.168[c)] (*bv. trifolii*)		Beijing		
R. leguminosarum AS 1.170[c)] (*bv. trifolii*)		Beijing		
R. leguminosarum 162K68 (*bv. trifolii*)	*Trifolium* sp.	USA		
R. leguminosarum AS 1.144[c)] (*bv. viceae*)		Beijing		
R. leguminosarum AS 1.145[c)] (*bv. viceae*)		Beijing		
R. leguminosarum AS 1.87[c)] (*bv. viceae*)		Beijing		
R. leguminosarum USDA 2370[b)] (*bv. viceae*)	*Pisum sativum*	USA	Jordan 1984	U29386
Rhizobium mongolense USDA 1844[b)]	*Medicago ruthenica*	Inner Mongolia	van Berkum 1998	U89817
Rhizobium sp. AS 1.79[c)]	*astragula*	Beijing		
Rhizobium sp. AS 1.80[c)]	*astragula*	Beijing		
Rhizobium sp. AS 1.81[c)]	*astragula*	Beijing		
Rhizobium sp. AS 1.82[c)]	*astragula*	Beijing		
Rhizobium sp. AS 1.83[c)]	*astragula*	Beijing		
Rhizobium sp. AS 1.171[c)]	*phaseoli*	Wuhan		
Rhizobium sp. AS 1.536[c)]	*phaseoli*	Wuhan		
Rhizobium sp. AS 1.1201[c)]	*phaseoli*	Wuhan		
Rhizobium tropici Type A CFN 299[b)]	*Phaseolus vulgaris*	Brazil	Martinez-Romero et al. 1991	X67233
R. tropici Type B CIAT 899[b)]	*Phaseolus vulgaris*	Colombia	Martinez-Romero et al. 1991	X67234
Rhizobium yanglingense CCBAU 71012	*Coronilla varia*	Gansu	Tan et al. 2001	
R. yanglingense CCBAU 71113	*Coronilla varia*	Shaanxi	Tan et al. 2001	
R. yanglingense CCBAU 71623[b)]	*Gueldenstaedtia multiflora*	Gansu	Tan et al. 2001	AF003375
Sinorhizobium arboris HAMBI 1552[b)]	*Prosopis chilensis*	Sudan	Nick et al. 1999	Z78204
Sinorhizobium fredii USDA 205[b)]	*Glycine soja*	Henan	Scholla 1984	X67231
Sinorhizobium kostiense HAMBI 1489[b)]	*Acacia senegal*	Sudan	Nick et al. 1999	Z78203
Sinorhizobium kummerowiae CCBAU 71714[b)]	*Kummerowia stipulacea*	Shaanxi	Wei et al. 2002	AF364067
Sinorhizobium medicae USDA 1037[b)]	*Medicago truncatula*	France	Rome et al. 1996	L39882
Sinorhizobium meliloti 1021[a)]	*Alfalfa*		Galibert et al. 2001	AL591688
S. meliloti AS 1.159[c)]		Beijing		
S. meliloti AS 1.160[c)]		Beijing		
S. meliloti AS 1.161[c)]		Beijing		
S. meliloti AS 1.163[c)]		Beijing		
S. meliloti 102F28	*Medicago sativa*	USA		
S. meliloti USDA 1002[b)]	*Medicago sativa*	USA	Jordan 1984	X67222
Sinorhizobium saheli LMG 7837[b)]	*Sesbania pachycarpa*	Senegal	de Lajudie et al. 1994	X68390
Sinorhizobium terangae LMG 7834[b)]	*Acacia laeta*	Senegal	de Lajudie et al. 1994	X68387
Sinorhzobium xinjiangense CCBAU 107	*Glycine max*	Xinjiang	Chen et al. 1988	
S. xinjiangense CCBAU 110[b)]	*Glycine max*	Xinjiang	Chen et al. 1988	AF250354

a) Strains whose genomic sequences are completed; b) strains stocked in Institute of Microbiology, the Chinese Academy of Sciences; c) type strain; AS, Academy Sinica; IAM, Institute of Applied Microbiology, The University of Tokyo, Tokyo, Japan; LMG, Collection of the Laboratorium voor Microbiologie en Microbiele Genetics, Rijksuniversiteit, B-9000,Gent, Belgium; USDA, US Department of Agriculture, Beltsville, MD; CCBAU, Culture Collection of Beijing Agricultural University; CFN, Centro de Investigacion sobre Fijacion de Nitrogeno, Universidad Nacional Autonoma de Mexico, Cuernavaca, Mexico; HAMBI, Culture Collection of the Department of Microbiology, Faculty of Agriculture and Forestry, University of Helsinki, Helsinki, Finland; CIAT, *Rhizobium* Collection, Centro International de Agricultura Tropical, Cali, Columbia.

blocks, cleaved with I-*Ceu*I and separated by PFGE as described previously[7−10]. PFGE was performed with Bio-Rad CHEF DRII. Images were collected by Quantity One Software and converted to Canvas files by Photoshop for comparative analyses.

1.4 Nucleotide sequence analysis

The sequences of 16S rRNA analyzed in this study were obtained from public databases and the GenBank accession numbers are given in table 1. They are the 16S rRNA sequences of 25 type strains and 4 sequenced strains. The sequences were aligned using Clustal X (1.81)[11] and the phylogenetic tree was constructed using the neighbour-joining method[12], with 1000 bootstrap replications. The evolutionary distances were estimated using Jukes-Cantor (PHYLIP 3.752c Version[13,14]).

2 Results

2.1 The features of rhizobial genome structures and their phylogenetic clusters

The 64 rhizobial strains used in this study were divided into 21 phylogenetic clusters based on features of their genome structures revealed by I-*Ceu*I cleavage patterns, with some of them further divided into subclusters based on minor differences (fig. 1 and table 2). For example, the strains of genus *Agrobacterium* were divided into subclusters XIId, XIIe and XIIf due to slight differences (table 2). Among the phylogenetic clusters resolved, sixteen, including I, II, III, V, VII, VII, IX, XI, XV, XVI, XVIII, XIX, XX, XXIV, XXV and XXVI, consisted of strains belonging to a single taxonomic genus or species. Three of the phylogenetic clusters (VI, XXI and XXII) consisted of strains belonging to 2 genera, *Rhizobium* and *Sinorhizobium*. Two phylogenetic clusters consisted of strains belonging to 3 genera, with cluster XII consisting of *Rhizobium giardinii* USDA2914^T, *Agrobacterium tumefaciens* and type strains of 8 species of genus *Sinorhizobium*, and cluster XXIII consisting of strains belonging to *Rhizobium*, *Sinorhizobium* and *Mesorhizobium*.

2.2 Comparison between phylogenetic clusters and taxonomic groupings (taxa)

(i) Consistency between results derived from the two clustering methods

i) Similar features in genome structure among strains within the same taxa

Type strains within the same taxa of some of the rhizobial bacteria used in this study have similar features in genome structure. For example, there are two kinds of similar but distinct features of genome structures for type strains within genus *Rhizobium* (except for *R. giardinii*); the species within genus *Bradyrhizobium* have the same features of genome structures; the genome structures of most species within genus *Sinorhizobium* are indistinguishable (table 2); and the species within genus *Mesorhizobium* have similar features of genome structures with slightly different genomic distributions of *rrn* operons.

ii) Dissimilar features in genome structure among strains belonging to different taxa

The features of genome structures of strains belonging to the different taxa are mostly different. For example, genome structures of type strains of *Rhizobium*, *Bradyrhizobium*, *Sinorhizobium* and *Mesorhizobium* are similar within a genus but different among the genera.

(ii) Inconsistency between results derived from the two clustering methods

i) Different features in genome structure among strains within the same taxa

Some rhizobial strains belonging to the same taxonomic genus had apparently different features in genome structure. As shown in fig. 2, the genome structures of *S. kummerowiae* CCBAU 71714 and *S. meliloti* 1021 are apparently different from those of other type strains within the taxonomic genus *Sinorhizobium*.

The genome structures of rhizobial strains within the same taxonomic species may also be so different as to be categorized into different phylogenetic clusters. Such cases include *M. huakuii* (subcluster Ic, cluster V and cluster XXIIId), *R. hainanense* (cluster VI and cluster VIIa), *S. meliloti* (cluster IX, cluster XIIg, cluster XVI, cluster XXI, cluster XXIIa and

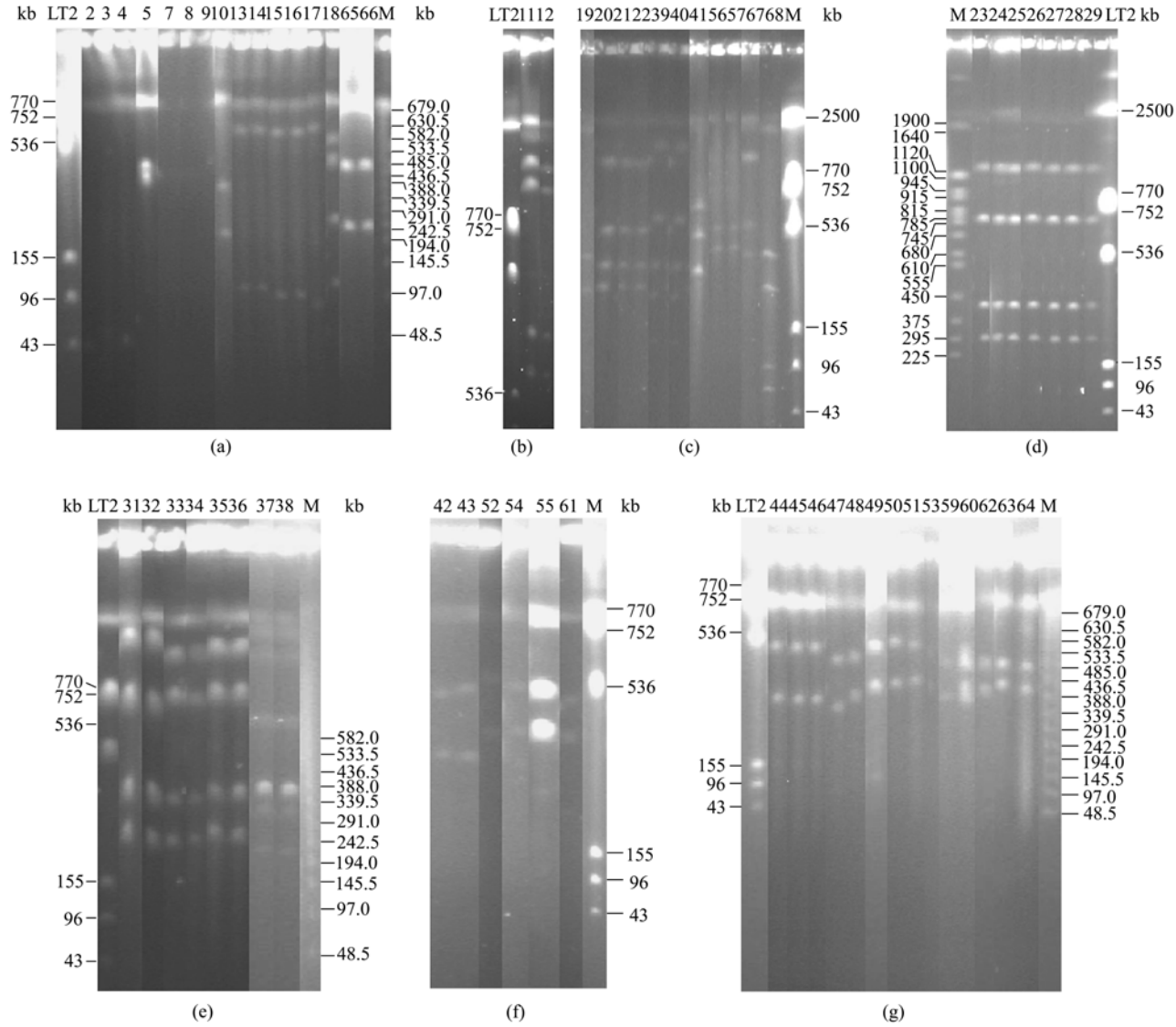

Fig. 1. PFGE gel of genomic DNA of 64 rhizobial strains after I-*Ceu*I cleavage. PFGE conditions: 30—120 s, 5.4 V/cm, 16 h; 50—60 s, 5.4 V/cm, 16 h; 80—120 s, 5.4 V/cm, 16 h. M, Yeast chromosomes as molecular size markers; *Salmonella typhimurium* LT2 as another molecular size marker.

cluster XXIIIa), *R. galegae* (cluster VIIa and cluster XXIV), *R. gallicum* (cluster XIXa and cluster XXIIIc), *R. leguminosarum* (cluster XVIIIb, XXb, XXc, XXd and cluster XXI), *R. tropici* (cluster VIIc and cluster VIId), *R. yanglingense* (cluster XIXb, XIXc and cluster XXIIa), and *S. xinjiangense* (cluster VI and cluster XIIa). This result suggests that the genome structures may have higher resolution than other taxonomic methods for rhizobia.

ii) Similar genome structures among strains belonging to different taxa

An important finding in this study is that some strains belonging to different taxonomic species or genus have similar genome structures. For example, strains of *Agrobacterium*, *R. giardinii* and type strains of *Sinorhizobium* have similar genome structures and so have been categorized into the same phylogenetic clusters. Some strains of *Rhizobium* and some strains

Table 2 Phylogenetic clusters and I-*Ceu*I cleavage patterns of the rhizobial strains

No.	Phylogenetic clusters	Strain	I-*Ceu*I cleavage pattern (Frag size in kb)
1	I	*Mesorhizobium loti* MAFF 303099[a)]	7028, 7.5
2	Ia	*Mesorhizobium mediterraneum* USDA 3392[b)]	>1500, 43
3	Ic	*Mesorhizobium huakuii* CCBAU 2609[b)]	>1500, 60
4	Id	*Mesorhizobium plurifarium* USDA 4413	>1500, 50
5	II	*Sinorhizobium kummerowiae* CCBAU 71042[b)]	>1500, 470, 460
6	III	*Bradyrhizobium japonicum* USDA 110[a)]	9 105
7	III	*Bradyrhizobium elkanii* USDA 76[b)]	>1500
8	III	*Bradyrhizobium japonicum* AS 1.826	>1500
9	III	*B. japonicum* AS 1.828	>1500
10	V	*Mesorhizobium huakuii* PL-52	>1500, 380, 200
11	VI	*Rhizobium hainanense* CCBAU 57003	>1500, 1150, 1050, 360, 225
12	VI	*Sinorhizobium xinjiangense* CCBAU 107	>1500, 1150, 1050, 360, 225
13	VIIa	*Rhizobium galegae* HAMBI 540[b)]	>1500, 700, 110
14	VIIa	*Rhizobium hainanense* CCBAU 57015[b)]	>1500, 700, 110
15	VIIc	*Rhizobium tropici*: typeB CIAT 899[b)]	>1500, 700, 90
16	VIIc	*Rhizobium mongolense* USDA 1844[b)]	>1500, 700, 90
17	VIId	*Rhizobium tropici*: typeA CFN 299[b)]	>1500, 680, 80
18	VIII	*Allorhizobium undicola* LMG 11875[b)]	>1500, 600, 475, 250, 110
19	IX	*Sinorhizobium meliloti* AS 1.161	>1500, 400, 240
20	XI	*Rhizobium* sp. AS 1.80	>1500, 900, 680, 370, 280
21	XI	*Rhizobium* sp. AS 1.82	>1500, 900, 680, 370, 280
22	XI	*Rhizobium* sp. AS 1.83	>1500, 900, 680, 370, 280
23	XIIa	*Rhizobium giardinii* USDA 2914[b)]	>1500, 1050, 770, 370, 260
24	XIIa	*Sinorhizobium saheli* LMG 7837[b)]	>1500, 1050, 770, 370, 260
25	XIIa	*Sinorhizobium terangae* LMG 7834[b)]	>1500, 1050, 770, 370, 260
26	XIIa	*Sinorhizobium medicae* USDA 1037[b)]	>1500, 1050, 770, 370, 260
27	XIIa	*Sinorhizobium fredii* USDA 205[b)]	>1500, 1050, 770, 370, 260
28	XIIa	*Sinorhizobium xinjiangense* CCBAU 110[b)]	>1500, 1050, 770, 370, 260
29	XIIa	*Sinorhizobium Kostiense* HAMBI 1489[b)]	>1500, 1050, 770, 370, 260
30	XIId	*Agrobacterium tumefaciens* C_{58}[a)]	2 453, 1046, 768, 388, 262
31	XIId	*A .tumefaciens* IAM 12048	>1500, 1050, 770, 388, 260
32	XIId	*A .tumefaciens* AS 1.1603	>1500, 1050, 770, 388, 260
33	XIIe	*A .tumefaciens* AS 1.1415	>1500, 1040, 780, 370, 260
35	XIIf	*A .tumefaciens* AS 1.1488	>1500, 1050, 780, 370, 260
36	XIIf	*A .tumefaciens* AS 1.1602	>1500, 1050, 785, 370, 260
37	XIIg	*Sinorhizobium meliloti* USDA 1002[b)]	>1500, 1050, 770, 470, 370, 260
38	XIIh	*Sinorhizobium arboris* HAMBI 1552[b)]	>1500, 1050, 770, 470, 370, 260
39	XV	*Rhizobium* sp. AS 1.79	>1500, 1000, 700, 370, 260
40	XV	*Rhizobium* sp. AS 1.81	>1500, 1000, 700, 370, 260
41	XVI	*Sinorhizobium meliloti* AS 1.160	>1500, 680, 550, 390
42	XVIIIa	*Mesorhizobium tianshanense* CCBAU 3306[b)]	>1500, 520, 410
43	XVIIIb	*Rhizobium etli* CFN 42[b)]	>1500, 530, 410
44	XVIIIb	*Rhizobium leguminosarum* AS 1.144 (biovar viceae)	>1500, 530, 410
45	XVIIIb	*Rhizobium leguminosarum* AS 1.167 (biovar trifolii)	>1500, 530, 410
46	XVIIIb	*R. leguminosarum* AS 1.168 (biovar trifolii)	>1500, 530, 410

(*To be continued on the next page*)

(*Continued*)

No.	Phylogenetic clusters	Strain	I-*Ceu*I cleavage pattern (Frag size in kb)
47	XIXa	*Rhizobium gallicum* FL 27	>1500, 530, 350
48	XIXb	*Rhizobium yanglingense* CCBAU 71623[b)]	>1500, 510, 410
49	XIXc	*R. yanglingense* CCBAU 71012	>1500, 530, 430
50	XXa	*Rhizobium* sp. AS 1.536	>1500, 610, 480
51	XXb	*Rhizobium leguminosarum* 162K68	>1500, 610, 500
52	XXb	*R. leguminosarum* USDA 2370[b)]	>1500, 610, 500
53	XXc	*R. leguminosarum* AS 1.170 (biovar trifolii)	>1500, 600, 510
54	XXe	*Rhizobium indigoferae* CCBAU 71042[b)]	>1500, 615, 490
55	XXI	*Sinorhizobium meliloti* 102F28	>1500, 580, 510
56	XXI	*Rhizobium leguminosarum* AS 1.145 (biovar viceae)	>1500, 550, 500
57	XXI	*R. leguminosarum* AS 1.87 (biovar viceae)	>1500, 550, 500
58	XXIIa	*Sinorhizobium meliloti* 1021[a)]	2 722, 530, 403
59	XXIIa	*S. meliloti* AS 1.163	>1500, 530, 400
60	XXIIa	*Rhizobium yanglingense* CCBAU 71113	>1500, 530, 400
61	XXIIb	*Rhizobium huautlense* S 02[b)]	>1500, 530, 420
62	XXIIIa	*Sinorhizobium meliloti* AS 1.159	>1500, 530, 410
63	XXIIIc	*Rhizobium gallicum* USDA 2918[b)]	>1500, 520, 430
64	XXIIId	*Mesorhizobium huakuii* A106	>1500, 520, 420
65	XXIV	*Rhizobium galegae* HAMBI 1185	>1500, 480, 210
66	XXIV	*R. galegae* HAMBI 503	>1500, 480, 210
67	XXV	*Rhizobium* sp. AS 1.171	>1500, 900, 560, 400
68	XXVI	*Rhizobium* sp. AS 1.1201	>1500, 1 200, 400, 200, 90, 60

a) Strains whose genomic sequences are completed; b) type strain.

of *Sinorhizobium* also have similar genome structures (table 2). For instance, *R. hainanense* CCBAU 57003 and *S. xinjiangense* CCBAU 107 were clustered together in VI; *S. meliloti* 102F28, *R. leguminosarum* AS 1.145, and *R. leguminosarum* AS 1.87 in XXI, *R. yanglingense* CCBAU 71113, and *S. meliloti* 1021 and *S. meliloti* AS 1.163 in XXIIa.

2.3 Comparison of phylogenetic clusters inferred from features of genome structures and those derived from 16S rRNA complete sequences

The results inferred from the feature of genome structure and those derived from 16S rRNA sequences are consistent in most cases.

The phylogenetic tree based on 16S rRNA complete sequences of 25 rhizobial type strains has 6 major branches (fig. 2).

Branch 1 includes type strains of 9 species within genus *Rhizobium*. They are distributed in 5 phylogenetic clusters, i.e. VII, XVIIIb, XIXb, XX and XXIIIc, based on features of their genome structures. Notably, although for the rhizobia in this branch, taxonomic results are consistent with those of the 16S rRNA comparisons, these bacteria are divided into different clusters based on features of their genome structures.

Branch 2 includes type strains of 9 species within genus *Sinorhizobium*. There are very short subbranches and distances among them, suggesting that phylogenetic relationships of these bacteria are very close. To the rhizobia in this branch, taxonomic results also correspond well with 16S rRNA conclusion, and they are divided into different clusters (i.e. cluster II, cluster XII and cluster XXIIa) based on features of their genome structures.

Branch 3 includes only 1 strain, *R. giardinii* USDA2914. It has a unique genome structure, which is quite different from those of the other *Rhizobium* strains, and was assigned to cluster XII as the only

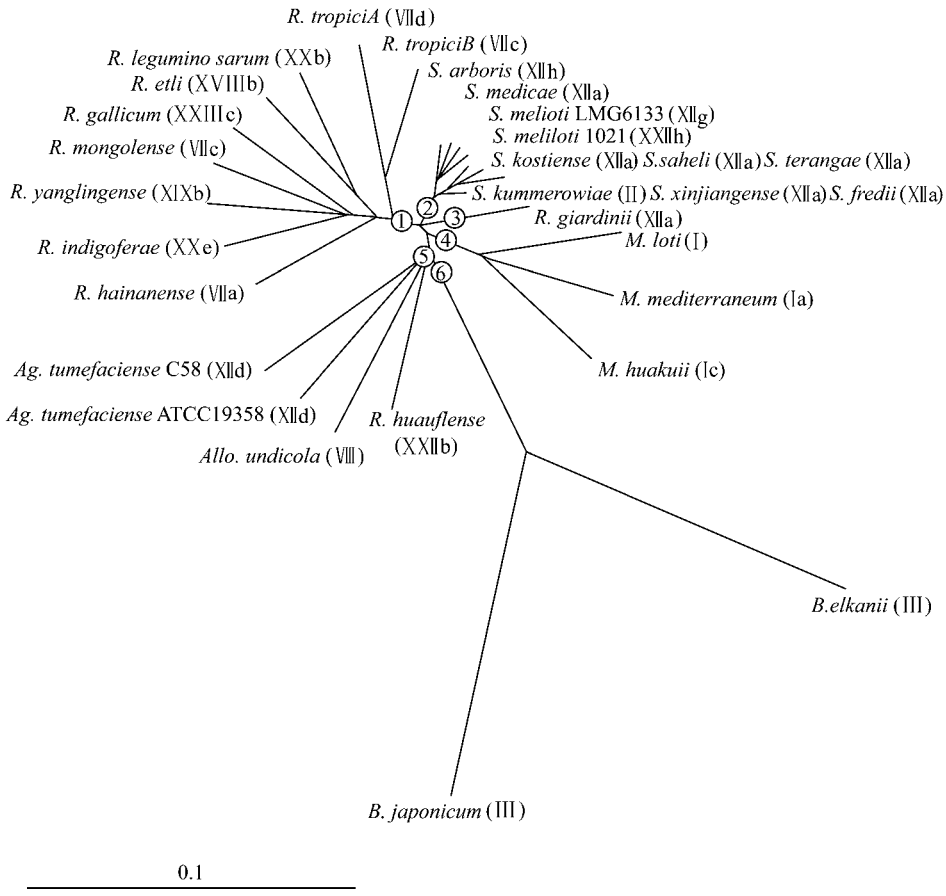

Fig. 2. Phylogenetic tree based on 16S rRNA sequences of 25 rhizobial type strains. GenBank accession numbers are listed in table 1. The scale bar represents the 0.1 substitution per site. ①—⑥ are the branches to which the strains belong.

strain. This genome structure, on the other hand, is very similar to those of *Sinorhizobium* strains, suggesting that *R. giardinii* USDA2914 is actually closely related to *Sinorhizobium*, rather than *Rhizobium*. Interestingly, 16S rRNA sequence analysis also suggests that *R. giardinii* USDA2914 is phylogenetically very close to *Sinorhizobium* (fig. 2).

Branch 4 includes type strains of genus *Mesorhizobium*. To the rhizobia in this branch, taxonomic results, 16S rRNA conclusion and phylogenetic clusters based on feature of genome structure are consistent. This branch includes only 1 phylogenetic cluster, i.e. cluster I, and cluster I only appears in this branch.

Branch 5 includes 2 type strains of genus *Bradyrhizobium* whose relationships are close based on 16S rRNA. They have indistinguishable genome structures and so were put together as phylogenetic cluster III. To the rhizobia in this branch, taxonomic results, 16S rRNA conclusion and phylogenetic clusters based on feature of genome structure are also consistent.

Branch 6 includes type strains of 3 genera (*Agrobacterium*, *Allorhizobium* and *Rhizobium*). Their genome structures are more diverse than those of the other branches and were categorized into the following clusters: XIId (*Agrobacterium*), VIII (*Allorhizobium*) and XXIIb (*Rhizobium*). The genome structure of bacteria in phylogenetic cluster XII also appears in other 2

closely adjacent branches (6 XIIa, 1 XIIg and 1 XIIh in branch 2; 1 XIIa in branch 3), which is inconsistent with clusters based on 16S rRNA sequences. Clusters VIII and XXIIb only appear in this branch and are consistent with clusters based on 16S rRNA sequences. Here, *Rhizobium huautlense* S02^T belongs to genus *Rhizobium* based on current taxonomic scheme, but 16S rRNA analysis proves that phylogenetically it is not very close to other strains of *Rhizobium*. Based on its genome structure, it also definitely stands out from other strains of *Rhizobium*, further supporting the hypothesis which suggests that features of genome structures reflect bacterial phylogenetic relationship.

3 Discussion

In this study, we employed a new clustering method which reflects bacterial phylogenetic relationship by features of their genome structures. This method has the following advantages: (1) PFGE apparatus can reveal genome structures for 40 bacterial strains in 2—3 days, so it is suitable for systematically comparing genomes of large numbers of bacteria and for clustering analysis; (2) this method does not need complex molecular cloning and expensive sequencing instrument; (3) the high specificity of I-*Ceu*I cleavage makes the results very consistent and comparable among different laboratories. In our laboratory, the clustering method based on genome structures had been applied to *Pasteurella*, *Klebsiella* and *Neisseria* and showed potential application in recognizing and determining their phylogenetic relationships[15,16]. To our knowledge, no similar work has been reported for rhizobia.

Current classification of rhizobia adopts polyphasic taxonomic scheme, i.e., firstly preliminarily clustering rhizobial isolates by numerical taxonomic methods based on phenotypic features and nucleotide fingerprinting methods based on genetic features, then selecting representative strains of these clusters to perform DNA-DNA hybridization with strains within the same cluster and type strains of related taxa, and measuring its G+C mol% and 16S rRNA sequence to infer the phylogenetic relationships of new taxon with closely related bacteria. Among these methods, the latter three reflect rhizobial phylogenetic relationships to a certain degree. However, DNA-DNA hybridization procedures are very tedious and G+C mol% only tells that a certain bacterial strain does not belong to a specific genus or species but can not tell that a strain belongs to a specific genus or species. In addition, 16S rRNA in many cases may not unambiguously reflect the phylogenetic relationships of rhizobia due to a variety of factors such as lateral gene transfer. The method we reported here can circumvent many of these problems by revealing evolutionarily conservative aspects of the bacteria through exploring the global information of the entire genome with efficient and reliable techniques. The genomes of many rhizobia include chromosomes (basic genome) and megaplasmids (accessory genome)[17]. In rhizobia, *rrn* operons (including 16S rRNA gene *rrs*, 23S rRNA gene *rrl* and 5S rRNA gene *rrf*) are located only on the chromosomes, not on the megaplasmids. As previous work has demonstrated that the copy number and genomic locations of *rrl* genes reflect bacterial phylogeny in enteric bacteria[7,15,16], we wanted to know whether I-*Ceu*I cleavage[18] would do the same in rhizobia.

This study revealed two interesting phenomena. One is that genome structures of strains within the same taxonomic genus or species may be different and the other is that genome structures of strains belonging to different taxonomic genera can be indistinguishable. Both situations are obviously conflicting with the current taxonomy of rhizobia but both can be interpreted on the phylogenetic basis as demonstrated in this study. For example, the finding that genome structures of strains within the same taxonomic genus are different is actually consistent with 16S rRNA clustering results, giving strong support to the assumption that genome structure reflects phylogeny. It is possible that different genome structures of closely related bacteria may have resulted from genomic rearrangement such as inversions or translocations when they are adapting to different environments, so that genomic locations of their *rrn* operons change as has been reported in *Sal-*

monella typhi[8,19], although other interpretations cannot be completely excluded yet at present[20]. The fact that strains of *Rhizobium* and *Sinorhizobium* that fall into the same phylogenetic cluster also demonstrates the phylogenetic nature of genome structure, which has been supported by 16S rRNA comparisons. In addition, strains belonging to *Rhizobium* and *Sinorhizobium* are similar in both phenotypic characteristics and genetic features[21−23], suggesting that they may have diverged from a common ancestor not very long ago.

This research shows that groupings based on genome structures in most of the cases corresponded well with the 16S rRNA comparison results. Some inconsistency with the 16S rRNA data was encountered. For example, genome structure of *Agrobacterium* is similar to that of strains of *Sinorhizobium*. However, on the phylogenetic tree based on 16S rRNA sequence, they do not belong to the same branch. This nevertheless is agreeable with the results of whole genome sequence analysis, which, in *A. tumefaciens* and *S. meliloti*, shows that these bacteria share a large number of orthologous genes (67% of *A. tumefaciens* genes), extensive nucleotide colinearity and conserved gene order[24], further demonstrating the phylogenetic value of genome structure as a method of categorizing the bacteria. The four rhizobial strains whose genome sequences are completed (*S. meliloti* 1021, *M. loti* 303099, *B. japonicum* 110 and *A. tumefaciens* C58) have different genome structures[22, 24 − 26], and the clustering method based on the feature of genome structures in this study also separated them very clearly.

Because this method utilizes global genome information in the analysis, it may reliably resolve the phylogenetic relationships among the rhizobia. In addition, this method is very efficient, allowing analysis of large numbers of bacterial strains within relatively short periods of time.

Acknowledgements The authors thank Pref. Chen Wenxin, Dr. Sui Xinhua and Dr. Chen Wenfeng of China Agricultural University for their bacterial strains and advice throughout this research. This work was supported by a 985 Foundation grant from Peking University Health Sciences Center and operating grants from the Medical Research Council of Canada (MRC, Grant # GOP-38106), the Natural Sciences and Engineering Research Council of Canada (NSERC, Grant # 216912-00), and the Canadian Institutes for Health Research (CIHR, Grant # MOP-47817).

References

1. Brock, T. D., Medigan, M. T., Martinko, J. M. et al., Biology of Microorganisms, 7th ed., NJ: Prenticw-Hall International, Inc., 1994, 695.
2. Gaunt, M. W., Turner, S. L., Rigottier-Gois, L. et al., Phylogenies of *atp*D and *rec*A support the small subunit rRNA-based classification of rhizobia, Int. J. Syst. Evol. Microbiol., 2001, 51(Pt 6): 2037—2048.
3. van Berkum, P., Terefework, Z., Paulin, L. et al., Discordant phylogenies within the *rrn* loci of rhizobia, J. Bacteriol., 2003, 185(10): 2988—2998.[DOI]
4. Gauthier, A., Turmel, M., Lemieux, C., A group I intron in the chloroplast large subunit rRNA gene of *Chlamydomonas eugametos* encodes a double-strand endonuclease that cleaves the homing site of this intron, Curr. Genet., 1991, 19(1): 43—47.
5. Marshall, P., Lemieux, C., Cleavage pattern of the homing endonuclease encoded by the fifth intron in the chloroplast large subunit rRNA-encoding gene of *Chlamydomonas eugametos*, Gene, 1991, 104(2): 241—245. [DOI]
6. Vincent, J. M., A Manual for the Practical Study of the Root-nodule Bacteria, IBP Handbook 15, Oxford: Blackwell Scientific Publications, 1970, 3—4.
7. Liu, S. L., Hessel, A., Sanderson, K. E., Genomic mapping with I-*Ceu* I, an intron-encoded endonuclease specific for genes for ribosomal RNA, in *Salmonella* spp., *Escherichia coli*, and other bacteria, Proc. Natl. Acad. Sci. USA, 1993, 90(14): 6874—6878.
8. Liu, S. L., Sanderson, K. E., Highly plastic chromosomal organization in *Salmonella typhi*, Proc. Natl. Acad. Sci. USA, 1996, 93(19): 10303—10308. [DOI]
9. Liu, S. L., Sanderson, K. E., A physical map of the *Salmonella typhimurium* LT2 genome made by using *Xba* I analysis, J. Bacteriol., 1992, 174(5): 1662—1672.
10. Liu, S. L., Sanderson, K. E., I-*Ceu* I reveals conservation of the genome of independent strains of *Salmonella typhimurium*, J. Bacteriol., 1995, 177(11): 3355—3357.
11. Thompson, J. D., Gibson, T. J., Plewniak, F. et al., The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools, Nucleic Acids Res., 1997, 25(24): 4876—4882. [DOI]
12. Saitou, N., Nei, M., The neighbor-joining method: A new method for reconstructing phylogenetic trees, Mol. Biol. Evol., 1987, 4(4): 406—425.
13. Felsenstein, J., Confidence limits on phylogenies: An approach using the bootstrap, Evolution, 1985, 39: 783—791.
14. Felsenstein, J., Phylip-(Phylogeny Inference Package), Seattle, University of Washington, 1993.
15. Liu, S. L., Schryvers, A. B., Sanderson, K. E. et al., Bacterial phylogenetic clusters revealed by genome structure, J. Bacteriol., 1999, 181(21): 6747—6755.
16. Liu, S. L., Liu, G. R., Li, S. X. et al., Bacterial genome structure: A molecular marker to reveal phylogenetic clusters, J. Peking Univ. (Med), 2002, 34(5): 457—463.

17. Jumas-Bilak, E., Michaux-Charachon, S., Bourg, G. et al., Unconventional genomic organization in the alpha subgroup of the proteobacteria, J. Bacteriol., 1998, 180(10): 2749—2755.
18. Liu, S. L., Sanderson, K. E., Bacterial Genome Analysis: Methods in Microbiology: Bacterial Pathogenesis, vol. 27, N. Y.: Academic Press, 1998, 371—381.
19. Liu, S. L., Sanderson, K. E., Rearrangements in the genome of the bacterium *Salmonella typhi*, Proc. Natl. Acad. Sci. USA, 1995, 92(4): 1018—1022.
20. Galibert, F., Finan, T. M., Long, S. R. et al., The composite genome of the legume symbiont *Sinorhizobium meliloti*, Science, 2001, 293(5530): 668—672.
21. Chen, W. X., Yan, G. H., Li, J. L., Numerical taxonomy study of fast-growing soybean rhizobia and a proposal that *Rhizobium fredii* be assigned to *Sinorhizobium* gen. nov, Int. J. Syst. Bacteriol., 1988, 38(4): 392—397.
22. Jarvis, B. D., Downer, H. L., Young, J. P., Phylogeny of fast-growing soybean-nodulating rhizobia support synonymy of *Sinorhizobium* and *Rhizobium* and assignment to *Rhizobium fredii*, Int. J. Syst. Bacteriol., 1992, 42(1): 93—96.
23. de Lajudie, P., Willems, A., Pot, B. et al., Polyphasic taxonomy of rhizobia: Emendation of the genus *Sinorhizobium* and description of *Sinorhizobium meliloti* comb. nov., *S. saheli* sp. nov., and *Sinorhizobium teranga* sp. Nov, Int. J. Syst. Bacteriol., 1994, 44: 715—733.
24. Wood, D. W., Setubal, J. C., Kaul, R. et al., The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58, Science, 2001, 294(5550): 2317—2323. [DOI]
25. Kaneko, T., Nakamura, Y., Sato, S. et al., Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*, DNA Res., 2000, 7(6): 331—338.
26. Kaneko, T., Nakamura, Y., Sato, S. et al., Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110, DNA Res., 2002, 9(6): 189—197.