

# Computer-assisted selection of restriction enzymes for *rrs* genes PCR-RFLP discrimination of rhizobial species

Marc Neyra<sup>a,b\*</sup>, Bouchaib Khbaya<sup>b,c</sup>, Philippe de Lajudie<sup>a</sup>,  
Bernard Dreyfus<sup>a</sup>, Philippe Normand<sup>b</sup>

<sup>a</sup> Laboratoire de microbiologie, ORSTOM, Institut français de recherche scientifique pour le développement en coopération, BP 1386 Dakar, Senegal

<sup>b</sup> Laboratoire d'écologie microbienne, UMR CNRS 5557 Université Claude-Bernard-Lyon I, 43, boulevard du 11 novembre 1918, 69622 Villeurbanne cedex, France

<sup>c</sup> Laboratoire de microbiologie et biologie moléculaire, Université Mohammed V, Rabat, Morocco

**Abstract** – Restriction fragment length polymorphism analysis of amplified deoxyribonucleic acid (DNA) (polymerase chain reaction/restriction fragment length polymorphism [PCR-RFLP]) is an appropriate tool for microbial diversity characterisation, because of its universality, simplicity, rapidity and low cost. To determine which and how many restriction enzymes are optimal for the analysis, we simulated the digestion of 24 small subunit ribosomal DNA (rDNA) (*rrs*) available sequences of rhizobia and relatives by 25 tetrameric enzymes, and systematically compared the patterns generated, singly or in combination, by the enzymes. Discriminating combinations can be determined, enabling all the selected species to be distinguished. The calculation of the frequency at which each enzyme is found in discriminating combinations allows enzyme sorting. The most appropriate enzymes in combination are not necessarily those that induce higher numbers of distinct patterns singly, but those that show complementary discriminative properties when analysed together. The combination of six of the seven most discriminant enzymes results on average in the discrimination of 99.5 % of the sequences. Therefore, we propose that a minimum of six enzymes should be used to ensure a good discrimination of known and unknown species of Rhizobiaceae. © Inra/Elsevier, Paris

**PCR-RFLP of *rrs* genes / restriction enzyme selection / characterisation / Rhizobiaceae**

---

\* Correspondence and reprints

**Résumé – Sélection d'enzymes de restriction par ordinateur pour la discrimination des espèces de *Rhizobium* par PCR-RFLP de l'ADNr 16S.** L'analyse du polymorphisme de longueur des fragments de restriction d'ADN amplifié (PCR-RFLP) présente de nombreux avantages pour la caractérisation de la diversité microbienne : relative rapidité, faible coût, possibilité d'application à l'ensemble des taxons bactériens, bonne reproductibilité, possibilité d'analyse simultanée d'un grand nombre de souches, intérêt au niveau taxonomique. Afin d'optimiser son usage, nous avons cherché à déterminer combien d'enzymes et lesquels étaient le plus approprié, en effectuant une analyse systématique par ordinateur de la capacité de 25 enzymes de restriction différentes à générer par PCR-RFLP, seules ou en combinaison, des profils de digestion permettant la distinction d'espèces proches taxonomiquement. La sous-unité 16S de l'ARN ribosomal (*rrs*, pour ribosomal RNA small subunit) a été choisie pour simuler ces PCR-RFLP : elle correspond à la région la plus fréquemment étudiée et permet en général une bonne discrimination au niveau de l'espèce ; de plus, de nombreuses séquences *rrs* sont disponibles dans les banques de données. L'étude a été focalisée sur les rhizobia qui représentent un bon modèle de genres différents, entremêlés avec d'autres genres très proches taxonomiquement, et parce que leur importance en agriculture induit la nécessité d'outils performants pour la caractérisation de grands nombres de nouveaux isolats. La digestion de 24 séquences *rrs* correspondant aux souches-types de différentes espèces a été simulée par ordinateur. Les profils de digestion obtenus pour les différentes séquences ont été comparés deux par deux afin de déterminer le nombre de profils distincts générés par chaque enzyme individuellement ou par les combinaisons systématiques de 2, 3, 4, 5, 6 ou 7 enzymes. Différentes combinaisons de 3, 4 ou 5 enzymes permettent la distinction entre les espèces. La fréquence à laquelle les différentes enzymes sont trouvées dans les combinaisons discriminantes ne correspond pas complètement à leur pouvoir de discrimination individuel, mais à la complémentarité qu'elles présentent lorsqu'elles sont analysées ensemble. La combinaison des profils obtenus avec six enzymes, choisies parmi les sept plus fréquentes dans les combinaisons discriminantes, permet en moyenne de distinguer plus de 99,5 % des séquences étudiées, quelle que soit la similarité entre ces séquences. L'utilisation d'au moins six de ces sept enzymes est suggérée afin d'assurer une bonne distinction des espèces de rhizobia, connues et inconnues. © Inra/Elsevier, Paris

PCR-RFLP / sélection d'enzymes de restriction / caractérisation / Rhizobiaceae

## 1. INTRODUCTION

Restriction fragment length polymorphism analysis of amplified deoxyribonucleic acid (DNA) (PCR-RFLP) has been widely used in recent years for characterising microbial diversity. PCR-RFLP analysis offers the advantage of relative rapidity (a large number of strains can be processed at the same time), low cost, universality (possible amplification of all bacterial taxa with universal primers) and good reproducibility. Several studies have demonstrated the validity of this technique in differentiating isolates belonging to different soil bacterial genera: *Nitrobacter* (Navarro et al., 1992), *Frankia* (Jamann et al., 1993), *Agrobacterium* (Ponsonnet and Nesme, 1994; Bouzar et al., 1995), *Rhizobium* (Laguerre et al., 1994a, 1996), *Pseudomonas* (Laguerre et al., 1994b; Manceau and Horvais, 1997), *Xanthomonas* and *Stenotrophomonas* (Nesme et al., 1995), and in estimating genetic relationships at species and higher levels. It has also proved to be useful for evaluation of microbial diversity in other natural habitats (Moyer et al., 1994; Haddad et al., 1995).

A recurrent question concerning PCR-RFLP is how to determine which and how many restriction enzymes must be used for the analysis. The selection of the most appropriate enzymes has frequently been proposed by empirical analysis of experimental results (and often by taking into account commercial availability and price). This kind of approach is appropriate to distinguish between a small number of known species. However, if the number of examined species increases, it rapidly becomes a tedious and difficult task. In contrast to other PCR-based DNA-typing methods such as RAPD (randomly amplified polymorphic DNA; Williams et al., 1990), AFLP (amplified fragment length polymorphism; Zabeau and Vos, 1993), use of primers derived from repetitive extragenic palindromic (REP) and enterobacterial repetitive intergenic consensus (ERIC) sequences (Versalovic et al., 1991) or transfer ribonucleic acid (tRNA) gene fragments (McClelland et al., 1992), PCR-RFLP offers the possibility, often underused, of checking experimental results with DNA sequencing data. Moyer et al. (1996) proposed the use of this advantage to estimate the efficacy of ten commonly used tetrameric restriction enzymes for studies of microbial diversity in nature, by computer-simulated restriction enzyme digestion of known sequences and subsequent analysis of size-frequency distribution of the resulting fragments. In this study we developed a different computer-assisted approach to estimate the ability of 25 different commercially available tetrameric restriction enzymes to generate, singly or in combination, discriminating restriction patterns allowing the distinction of closely related species. The final aim was to propose a screening approach for the selection of the enzymes, and to predict the number of enzymes required to ensure differentiation of known and unknown species.

We focused on the 16S rRNA region (SSU rRNA, for small subunit of ribosomal RNA or *rrs*) because it is the most frequently used bacterial gene for PCR-RFLP, and because it allows, in general, good discrimination at the species level. Furthermore, the large number of available *rrs* sequences in data bases makes it possible to simulate restriction patterns for a wide range of microorganisms.

Rhizobia were chosen as target microorganisms because they represent a good model of different genera interleaved with various other closely related genera, such as *Agrobacterium* and *Phyllobacterium* (Jarvis et al., 1988; Willems and Collins, 1993; Yanagi and Yamasato, 1993; Young and Haukka, 1996), and because their importance in agricultural and environmental studies induces the need for powerful tools for efficient screening of large numbers of new isolates.

## 2. MATERIALS AND METHODS

### 2.1. Sequences selection and analysis

A total of 24 sequences derived from the type strains (wherever possible) of the named species of rhizobia and some related bacteria were selected from the GenBank data base. Seventeen *rrs* sequences were obtained for the rhizobia sensu stricto: *Azorhizobium caulinodans* (GenBank accession no. D11342), *Bradyrhizobium elkanii* (U35000), *B. japonicum* (D13430), *Mesorhizobium ciceri* (U07934), *M. huakuii* (D13431), *M. loti* (X67229), *M. mediterraneum* (L38825), *Rhizobium etli* (U28916), *R. galegae* (D11343), *R. leguminosarum*

(X67227), *R. tropici* IIA (X67234), *R. tropici* IIB (X67233), *Sinorhizobium fredii* (D14516), *S. medicae* (L39882), *S. meliloti* (X67222), *S. saheli* (X68390) and *S. teranga* (X68388). Seven *rrs* sequences of species closely related to fast-growing rhizobia were also analysed: *Agrobacterium rhizogenes* (D14501), *A. rubi* (X67228), *A. tumefaciens* (D14500), *A. vitis* (U45329), *Blastobacter aggregatus* (X73041), *Mycoplana dimorpha* (D12786) and *Phyllobacterium rubiacearum* (D12790).

Sequences were aligned using the Clustal program (Higgins and Sharp, 1988). Uncertainties were resolved by comparing with the sequences of the other species, and, when available, with sequences of strains of the same species. PCR amplification with primers FGPS5-255 (5'-ATGGAAGCTTGATCCTGGCT-3') and FGPS1509'-153 (5'-AAGGAGGGGATCCAGCCGCA-3') (Normand, 1995) was simulated by ending all the *rrs* sequences with primer sequences. These two primers were selected because they experimentally allow successful amplification of almost full length *rrs* genes of all rhizobial strains tested (data not shown).

## 2.2. Restriction patterns simulation

For all the selected sequences, the size of fragments obtained after digestion with each of the 25 following tetrameric restriction endonucleases was determined: *AccII*, *AciI*, *AluI*, *BfaI*, *BsaJI*, *BsII*, *Cac81*, *DdeI*, *DsaV*, *FnuAHI*, *HaeIII*, *HhaI*, *HinfI*, *HpaII*, *MaeII*, *MaeIII*, *MboI*, *MseI*, *MwoI*, *NlaIII*, *NlaIV*, *RsaI*, *Sau96I*, *TaqI*, *Tsp509I* (recognition sites and isoschizomers are given in table I).

The patterns obtained for the different species were compared pairwise to determine the distinct restriction patterns generated by each enzyme. To simulate experimental conditions, due to the resolution threshold of standard 3 % w/v agarose gel electrophoresis, only fragments larger than 90 bp were taken into account in the analysis. Furthermore, fragments differentiated by less than 5 % were considered as similar.

## 2.3. Enzyme combinations analysis

The cumulative patterns generated by multiple combinations of different sets of enzymes were determined for each of the species. For each combination, the cumulative patterns of the different species were compared pairwise in order to evaluate the number of distinct cumulative patterns. The number of discriminating combinations (i.e. combinations that differentiate all the species) was scored for each set of combinations; the frequency of occurrence of each enzyme in the discriminating combinations was calculated and expressed as the ratio of occurrence of the enzyme to the total number of discriminating combinations for a given set. Because of the great number of combinations to be tested, we developed a simple analysis using HyperTalk language on Macintosh computers.

## 3. RESULTS

All the 25 enzymes tested were theoretically able to cut at least once each of the *rrs* sequences selected. They induced between two (*HinfI* on

**Table 1.** Mean number of measurable fragments, number of distinct patterns and discriminatory index for each restriction enzyme, estimated after theoretical digestion of rhizobia and closely related species *r73* rRNA genes.

Name	Enzymes	Rhizobia sensu stricto				All sequences		
		Recognition site	Isoschizomers	Mean no. of fragments <sup>a</sup>	No. of patterns	Discrim. index <sup>b</sup>	No. of patterns	Discrim. index <sup>b</sup>
<i>Fnu4HI</i>		GC/NGC	<i>BsoFI</i>	7.6	9	0.59	15	0.59
<i>BsiI</i>		C/CN>NNNNNGG		3.6	12	0.50	16	0.49
<i>Cac8I</i>		GCN/NGC		6.9	12	0.48	16	0.49
<i>AclI</i>		C/CGC		7.4	10	0.37	14	0.37
<i>HirtI</i>		G/ANTC		3.8	9	0.29	11	0.28
<i>Sau96I</i>		G/GNCC	<i>AsuI, Cfr13I</i>	6.6	10	0.23	14	0.23
<i>HpaII</i>		C/CGG	<i>MspI</i>	5.3	12	0.22	17	0.22
<i>MaeII</i>		A/CGT		4.5	8	0.16	10	0.16
<i>MaeIII</i>		/GTNAC		5.3	10	0.16	14	0.16
<i>AluI</i>		AG/CT		5.8	8	0.16	12	0.16
<i>MboI</i>		/GATC	<i>NdeII, Sma3AI</i>	3.7	8	0.15	9	0.14
<i>HhaI</i>		GCG/C	<i>CfoI</i>	6.5	6	0.14	8	0.14
<i>HaeIII</i>		GG/CC	<i>PaeI</i>	5.2	8	0.14	11	0.14
<i>RsaI</i>		GT/AC		3.6	5	0.14	5	0.13
<i>MseI</i>		T/TAA		3.8	6	0.13	8	0.13
<i>DsaV</i>		CC/NGG		5.9	9	0.13	11	0.13
<i>AccII</i>		CG/CG		6.8	8	0.13	9	0.12
<i>DdeI</i>		C/TNAG	<i>BstUI, Thal</i>	5.3	7	0.11	7	0.11
<i>NlaIV</i>		GGN/NCC		2.5	7	0.11	10	0.11
<i>MwoI</i>		GCN>NNN/NGC		7.6	8	0.11	14	0.11
<i>BfaI</i>		C/TAG	<i>MaeI</i>	5.6	8	0.11	8	0.11
<i>NlaIII</i>		CATG/		5.7	3	0.11	4	0.11
<i>BsaJI</i>		C/CNNGG		4.1	5	0.11	8	0.11
<i>TaqI</i>		T/CGA		3.5	6	0.11	7	0.11
<i>Tsp509I</i>		/AATT		3.9	4	0.11	5	0.11

<sup>a</sup> Measurable fragments that can be resolved in 3% w/v agarose (> 90 pb); <sup>b</sup> discriminatory index: frequency at which an enzyme is found in the discriminating combinations, estimated here for combinations of five enzymes.

*A. tumefaciens* and *R. etli*; *MnII* on *B. elkanii* and *B. japonicum*) and 16 restriction fragments (*Bsa*JI on *M. dimorpha*). The average number of measurable fragments (> 90 bp) for the rhizobia sensu stricto showed a range of 2.5 to 7.6 per enzyme, with an overall average of 5.2 (table I). These theoretical patterns were confirmed in vitro for some of the enzymes and species (data not shown). One pattern only (obtained by digestion with *Rsa*I of the amplified *rrs* rRNA of *S. saheli* type strain ORS 609) was not as expected and yielded three bands instead of the four predicted by computer simulation, but the comparison of aligned sequences showed an error in the published sequence, inducing a false additional restriction site. It was corrected for further analysis.

From three to 12 distinct restriction patterns were detected with the 25 enzymes for the 17 *rrs* sequences of the rhizobia sensu stricto, and from four to 17 for the global set of 24 sequences (table I). The number of patterns was not related to the number of fragments. For example, a mean number of measurable fragments of 3.6 was sufficient to obtain 12 distinct patterns with *Bs*II, whereas only three patterns were found for a mean number of fragments of 5.7 with *Nla*III. None of the enzymes was able to distinguish alone among the set of sequences. Decreasing the limit of detection of the fragments (i.e. from 90 to 80 bp) did not significantly modify this analysis (data not shown). The different patterns obtained for each enzyme and each of the 24 sequences are listed in table II.

The discriminating power of multiple combinations of the 25 enzymes was estimated for the rhizobia sensu stricto and for the global set of sequences. The total number of possible combinations of  $n$  enzymes chosen from a total of  $N$  available enzymes is equal to:  $N!/(N-n)!n!$ . Therefore, for 25 available enzymes the numbers of combinations of two, three, four and five enzymes are 300, 2 300, 12 650 and 53 130, respectively. The average number of distinct patterns per combination for the 17 sequences of rhizobia alone was 11.2, 12.9, 13.9 and 14.7 for the combinations of two, three, four and five enzymes, respectively, and 16.2, 19.0, 20.5 and 21.6 for the global set of 24 sequences. The efficiency of the different enzyme combinations in distinguishing between the sequences (expressed as the average percent of sequences detected) is presented in figure 1.

Only one combination of two enzymes (*Bs*II and *Fnu*AHI) distinguished between all the sequences of rhizobia. The same combination was also sufficient to distinguish between the global set of sequences. The number of discriminating combinations of three, four and five enzymes was nearly the same for rhizobia alone and for the global set of sequences (table III). The discriminatory index (i.e. the frequency at which an enzyme is found in a discriminating combination) was estimated for each enzyme and for each combination. The result for the combinations of five enzymes is shown in table I. The discriminatory indices were practically identical for the two sets of sequences, and ranged from 0.11 to 0.59. As for single enzyme patterns, the discriminatory index of enzymes was not related to the number of fragments. On the other hand, the enzymes producing the best discriminatory indices in multiple combinations were also the ones showing as a whole the larger number of distinct patterns when used singly (table I). For the ten enzymes that were the best when used in combination to distinguish between sequences, the mean number of individual distinct patterns was 10, whereas for the ten less discriminatory enzymes the mean was 6.5 only. However, the discriminatory index was not entirely correlated with

**Table II.** Restriction patterns obtained after theoretical digestion of rhizobia and related species *rrs* genes by 25 different enzymes.

Species	AccII	AluI	BsaJI	Cac81*	DsaV	HaeIII	HhaI	HinfI*	MaeII	MboI	MscI	MtuOI	NlaIV	RsaI	Sma961*	TaqI	Tsp509I
	AccI*	BfaI	BstII*	DdeI	Fnu4HI*	HhaI	HpaII*	MaeII*	MaeIII	MscI	MscI	MtuOI	NlaIII	RsaI	Sma961*	TaqI	Tsp509I
<i>R. leguminosarum</i>	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a
<i>R. etli</i>	a	a	a	b	a	b	b	b	a	b	a	a	a	b	b	a	a
<i>R. galegae</i>	b	a	b	c	b	a	a	a	b	c	b	a	a	c	a	c	b
<i>R. tropici</i> IIA	c	c	c	d	d	c	c	d	c	d	c	b	a	c	d	c	c
<i>R. tropici</i> IIB	a	a	d	d	d	a	c	a	d	a	a	d	c	a	a	e	a
<i>M. ciceri</i>	d	d	e	a	e	a	d	e	e	e	d	e	c	e	d	f	d
<i>M. huakuii</i>	d	d	e	a	e	f	a	e	f	e	b	b	c	e	d	g	d
<i>M. loti</i>	d	e	a	e	e	a	d	f	e	e	d	e	c	e	d	f	d
<i>M. mediterraneum</i>	d	d	e	a	e	e	e	e	e	e	b	e	c	e	d	g	d
<i>S. fredii</i>	e	f	e	a	b	a	e	d	g	f	b	d	c	e	d	h	e
<i>S. medicae</i>	e	f	a	h	c	a	d	a	h	f	b	d	c	e	d	i	e
<i>S. meliloti</i>	e	f	a	h	i	a	d	d	a	f	b	d	c	e	a	i	e
<i>S. saheli</i>	f	g	a	i	c	a	d	i	a	h	g	b	d	c	e	a	a
<i>S. teranga</i>	f	g	a	i	c	a	d	i	a	h	g	b	d	c	e	a	a
<i>B. japonicum</i>	b	h	f	j	e	g	f	g	g	a	h	b	d	c	e	a	a
<i>B. elkanii</i>	g	i	g	k	f	g	h	k	h	i	a	e	g	a	g	e	j
<i>A. caulinodans</i>	h	j	h	l	g	i	l	c	j	b	f	h	a	e	a	k	f
<i>A. rhizogenes</i>	a	e	a	a	d	b	a	d	c	k	a	d	c	a	a	e	a
<i>A. rubi</i>	g	k	i	b	m	c	j	i	h	a	e	g	i	c	h	a	e
<i>A. tumefaciens</i>	g	k	j	b	n	c	i	k	j	e	g	i	c	h	a	e	b
<i>A. vitis</i>	i	l	i	b	c	i	l	c	e	a	n	i	c	e	b	m	e
<i>B. aggregatus</i>	d	m	k	b	g	o	c	k	m	k	h	a	o	a	l	i	e
<i>M. atomorpha</i>	d	n	l	f	h	o	o	a	e	n	b	e	j	p	a	m	e
<i>P. rubiacearum</i>	e	f	l	e	a	p	a	d	o	a	i	k	q	j	n	f	d

Each lowercase letter indicates a different pattern detected with each enzyme among the 25 sequences; \* indicates the seven most discriminative enzymes.

the number of distinct patterns: for example, *Fnu4HI*, the best enzyme in combination, allowed the detection of only nine distinct patterns among rhizobia species when used singly, whereas *HpaII* had a discriminatory index almost three times lower, in spite of its ability to distinguish alone 12 patterns.

The seven enzymes that gave the best results (*Fnu4HI*, *BsII*, *Cac81*, *AccI*, *HinfI*, *Sau96I* and *HpaII*) were selected for further analysis. Their ability to distinguish at the species level was studied by comparing the patterns generated with these enzymes when used either singly or in combination.

All the possible combinations of two, three, four, five, six and seven enzymes were tested on the two previously studied sets of sequences (rhizobia sensu stricto and the global set of 24 sequences): the percentage of differing sequences detected ranged from 60 when one enzyme was used to 100 when seven enzymes were used in combination (figure 1).

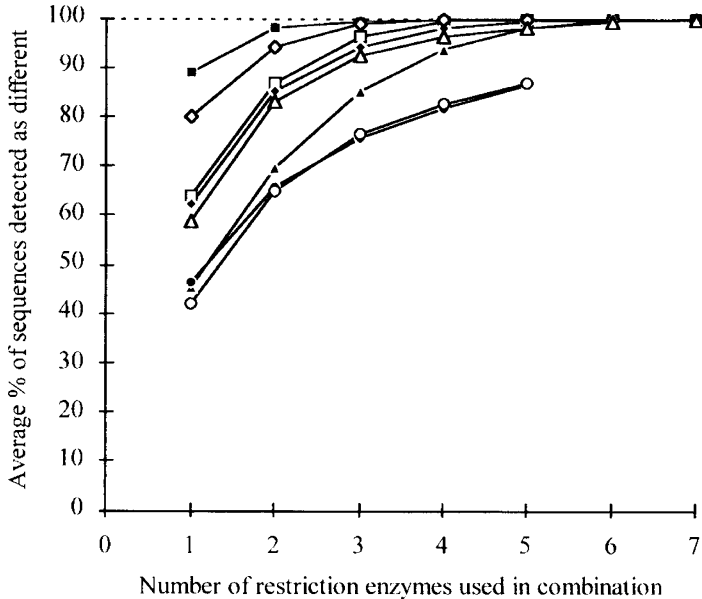
The different combinations of the seven selected enzymes were also applied independently to the sequences corresponding to the three genera of fast-growing rhizobia (*Rhizobium*, *Mesorhizobium* and *Sinorhizobium*). The results for the *Mesorhizobium* set were very similar to those obtained with the global sets of sequences (figure 1). The average percentages of differing sequences calculated for the *Rhizobium* species were clearly higher, from 80 % with one enzyme up to 98.9 % for the combination of only three enzymes. In contrast, more enzymes were necessary to distinguish between the *Sinorhizobium* sequences.

#### 4. DISCUSSION

Rhizobia classification is constantly and rapidly changing. Since the separation of the fast- and slow-growing rhizobial species into two different genera, *Rhizobium* and *Bradyrhizobium* (Jordan, 1982), three new genera have been recognised among root- and stem-nodulating bacteria of legumes: *Azorhizobium* (Dreyfus et al., 1988), *Sinorhizobium* (de Lajudie et al., 1994) and *Mesorhizobium* (Lindström et al., 1995). Seventeen different species are currently recognised for the five genera (Young and Haukka, 1996), but this number increases continuously as new species are discovered and described using the classic tools of DNA/DNA hybridisation and *rrs* sequencing. Furthermore, various bacteria have *rrs* sequences that are very similar to those of fast-growing rhizobia. *Agrobacterium* and *Phyllobacterium* are two genera of bacteria of the Rhizobiaceae, generally inducing cortical hypertrophies (galls or nodules) on plants (Jordan, 1994). *Blastobacter aggregatus* (Hirsh and Müller, 1985) and *Mycoplana dimorpha* (Gray and Thornton, 1928) are two species isolated from lake water and soil, respectively. Because both species can be isolated from the same habitats (*Agrobacterium* especially), it is important to be able to differentiate them from rhizobia.

Our computer-assisted simulation confirmed that PCR-RFLP of the *rrs* gene is an appropriate tool to distinguish between species of rhizobia and related bacteria, as previously demonstrated by Laguerre et al. (1994a), who also showed the validity of this method in estimating the genetic relationship between species. However, as a consequence of the increasing number of recognised species, it appears that discrimination between all the named species of rhizobia becomes increasingly difficult to ensure with a single set of few enzymes. For example, the minimum of four enzymes (*CfoI*, *HinfI*, *MspI* and





**Figure 1.** Efficiency of different restriction enzyme combinations in distinguishing between rhizobia and some related bacteria. Combinations of 25 enzymes: ● rhizobia sensu stricto; ○ global set of sequences. Combinations of seven enzymes: ◆ rhizobia sensu stricto; △ global set of sequences; ◇ *Rhizobium* spp.; ▲ *Sinorhizobium* spp.; □ *Mesorhizobium* spp.

**Table III.** Number of combinations of one, two, three, four or five restriction enzymes allowing complete distinction between 17 *rrs* rRNA sequences of rhizobia sensu stricto and between 24 sequences of rhizobia and related species.

No. of enzymes in combination	1	2	3	4	5
Total no. of possible combinations	25	300	2 300	12 650	53 130
Rhizobia sensu stricto	0	1	41	631	5 392
All sequences	0	1	41	626	5 328

*NdeII*) found by Laguerre et al. (1994a) to be sufficient to discriminate between all the species recognised at that time can no longer separate *S. meliloti* from the newly described species *S. medicae* (Rome et al., 1996). To overcome this problem, one may be tempted to screen more enzymes in order to find a new discriminating combination. When looking at 25 different tetrameric enzymes, various combinations of three, four or five enzymes can be selected to allow complete discrimination of all the species studied (table III). It is even possible to find two enzymes (*BsII* and *Fnu4HI*) that can, in combination, distinguish between all currently known rhizobial species. However, such a choice remains too empirical, and the appropriate combinations selected are likely to become obsolete with the increasing number of recognised species.

Therefore, for a more accurate choice it is important to sort the enzymes that are the most effective in combination. Looking at the frequency at which the enzymes are found in a discriminating combination allows them to be sorted according to their real efficacy in combination. The classification we obtained with the discriminatory indices was not directly related to the number of distinct patterns generated by each enzyme. This may be due to a complementary discriminatory power of the best enzymes when the results of each of them are considered together, and to a possible superfluous effect for the less effective enzymes. Moyer et al. (1996) also considered the same possibility of a 'synergistic' effect between enzymes to explain the better efficiency of one of the restriction enzyme groups they studied. The average percentage of sequences that can be distinguished by combination of the different enzymes enables prediction of the number of enzymes required to ensure the detection of a large part of the existing species. The average percentage of sequences detected by the seven selected enzymes used singly is quite different according to the set of sequences examined (*figure 1*). This can be explained by the level of similarity of the sequences within each set. The five sequences of *Sinorhizobium* studied are very close, and so are more difficult to distinguish. On the other hand, the percentage of *Rhizobium* sequences detected is high, and can be explained by the fact that species which are not closely related are still gathered in this genus: *Rhizobium galegae* is not a typical member of the genus *Rhizobium* and its *rrs* sequence is close to the sequence of *Agrobacterium* species (Young and Haukka, 1996); even more important is the presence of an insertion of 72 nucleotides in the *rrs* of *Rhizobium tropici* IIA (Willems and Collins, 1993), which results in the fact that restriction patterns for this sequence are always different from the others. Also interesting is that, whatever the differences for one enzyme, the average percentage of sequences detected always exceeds 99.5 % when combinations of six of the seven selected enzymes are used (*figure 1*). This suggests that six selected enzymes will be sufficient to ensure good separation between rhizobia species.

Unfortunately, the best enzymes (*Fnu4HI*, *BsII*, *Cac81* and *AcI*) are also among the most expensive. This can explain why we found only *Fnu4HI* in previously published works on PCR-RFLP. The most frequently used enzymes are *AluI*, *CfoI* (= *HhaI*), *HaeIII*, *MspI* (= *HpaII*), *NdeII* (= *MboI*), *RsaI* and *TaqI* (Navarro et al., 1992; Jamann et al., 1993; Laguerre et al., 1994a; Moyer et al., 1994; Ponsonnet and Nesme, 1994; Nesme et al., 1995; Brunk et al., 1996; Laguerre et al., 1996; Moyer et al., 1996; Brunel et al., 1997; Manceau and Horvais, 1997), all of which are cheap and widely available from commercial suppliers. However, those enzymes already do not discriminate *Mesorhizobium loti* from *M. ciceri*, and it is obvious that it will be increasingly difficult to distinguish new species with these enzymes.

Although computer simulation of enzymatic digestion of *rrs* may underestimate technical problems such as the partial digestion of PCR products with some enzymes, or the existence in the same bacteria of different *rrs* copies that are not necessarily identical, theoretical testing of the discriminatory power of the different restriction enzymes appears to be a good way of optimising PCR-RFLP analysis. It allows the choice of the best enzymes to discriminate, in combination, the different species of rhizobia. It shows that these best enzymes are not necessarily those that are the most discriminating when used singly,

but that there is a complementary discriminative ability when the results of digestion with these enzymes are analysed in combination. It indicates also that the use of six enzymes chosen from the more discriminating ensures the distinction of isolates at the species level. This approach could be used on other bacterial taxa.

## ACKNOWLEDGEMENTS

We thank X. Nesme and G. Grundman for helpful discussions. This work was supported by the Bureau des Ressources Génétiques, Paris, France.

## REFERENCES

- Bouzar H., Chilton W.S., Nesme X., Dessaux Y., Vaudequin V., Petit A., Jones J.B., Hodge N.C., A new *Agrobacterium* strain isolated from aerial tumors on *Ficus benjamina* L., *Appl. Environ. Microb.* 61 (1995) 65–73.
- Brunel B., Givaudan A., Lanois A., Akhurst R.J., Boemare N., Fast and accurate identification of *Xenorhabdus* and *Photorhabdus* species by restriction analysis of PCR-amplified 16S rRNA genes, *Appl. Environ. Microb.* 63 (1997) 574–580.
- de Lajudie P., Willems A., Pot B., Dewettinck D., Maestrojuan G., Neyra M., Collins M.D., Dreyfus B.L., Kersters K., Gillis M., Polyphasic taxonomy of rhizobia: emendation of the genus *Sinorhizobium* and description of *Sinorhizobium meliloti* comb. nov., *Sinorhizobium saheli* sp. nov., and *Sinorhizobium teranga* sp. nov., *Int. J. Syst. Bacteriol.* 44 (1994) 715–733.
- Dreyfus B., Garcia J.L., Gillis M., Characterization of *Azorhizobium caulinodans* gen. nov., sp. nov., a stem-nodulating nitrogen-fixing bacterium isolated from *Sesbania rostrata*, *Int. J. Syst. Bacteriol.* 38 (1988) 89–98.
- Haddad A., Camacho F., Durand P., Cary S.C., Phylogenetic characterization of the epibiotic bacteria associated with the hydrothermal vent polychaete *Alvinella pompejana*, *Appl. Environ. Microb.* 61 (1995) 1679–1687.
- Higgins D.G., Sharp P.M., Clustal: a package for performing multiple alignment on a microcomputer, *Gene* 73 (1988) 237–244.
- Jamann S., Fernandez M.P., Normand P., Typing method for N<sub>2</sub>-fixing bacteria based on PCR-RFLP – Application to the characterisation of *Frankia* strains, *Mol. Ecol.* 2 (1993) 17–26.
- Jarvis B.D., Gillis M., De Ley J., Intra- and intergeneric similarities between the ribosomal ribonucleic acid cistrons of *Rhizobium* and *Bradyrhizobium* species and some related bacteria, *Int. J. Syst. Bacteriol.* 36 (1986) 129–138.
- Jarvis B.D., Downer H.L., Young J.P.W., Phylogeny of fast-growing soybean rhizobia supports synonymy of *Sinorhizobium* and *Rhizobium* and assignment to *Rhizobium fredii*, *Int. J. Syst. Bacteriol.* 42 (1992) 93–96.
- Jordan D.C., Transfer of *Rhizobium japonicum* Buchanan 1980 to *Bradyrhizobium* gen. nov., a genus of slow-growing, root nodule bacteria from leguminous plants, *Int. J. Syst. Bacteriol.* 32 (1982) 136–139.
- Jordan D.C., Rhizobiaceae, in: Krieg N.R. (Ed.), *Bergey's Manual of Systematic Bacteriology*, vol. 1, Williams and Wilkins, Baltimore, 1984, pp. 234–256.

Laguerre G., Allard M.R., Revoy F., Amarger N., Rapid identification of rhizobia by restriction fragment length polymorphism analysis of PCR-amplified 16S rRNA genes, *Appl. Environ. Microb.* 60 (1994a) 56–63.

Laguerre G., Rigottier-Gois L., Lemanceau P., Fluorescent *Pseudomonas* species categorized by using polymerase chain reaction (PCR)/restriction fragment analysis of 16S rDNA, *Mol. Ecol.* 3 (1994b) 479–487.

Laguerre G., Mavingui P., Allard M.R., Charnay M.P., Louvrier P., Mazurier S.I., Rigottier-Gois L., Amarger N., Typing of rhizobia by PCR DNA fingerprinting and PCR-restriction fragment length polymorphism analysis of chromosomal and symbiotic gene regions: application to *Rhizobium leguminosarum* and its different biovars, *Appl. Environ. Microb.* 62 (1996) 2029–2036.

Lindström K., van Berkum P., Gillis M., Martinez E., Novikova N., Jarvis B., Report from the roundtable on rhizobium taxonomy, in: Tikhonovich I.A., Provorov N.A., Romanov V.I., Newton W.E. (Eds.), *Nitrogen Fixation: Fundamentals and Applications*, Kluwer, Dordrecht, 1995, pp. 807–810.

Manceau C., Horvais A., Assessment of genetic diversity among strains of *Pseudomonas syringae* by PCR-restriction fragment length polymorphism analysis of rRNA operons with special emphasis on *P. syringae* pv. tomato, *Appl. Environ. Microb.* 63 (1997) 498–505.

Moyer C.L., Dobbs F.C., Karl D.M., Estimation of diversity and community structure through restriction fragment length polymorphism distribution analysis of bacterial 16S rRNA genes from a microbial mat at an active, hydrothermal vent system, Loihi Seamount, Hawaii, *Appl. Environ. Microb.* 60 (1994) 871–879.

Moyer C.L., Tiedje J.M., Dobbs F.C., Karl D.M., A computer-simulated restriction fragment length polymorphism analysis of bacterial small-subunit rRNA genes: efficacy of selected tetrameric restriction enzymes for studies of microbial diversity in nature, *Appl. Environ. Microb.* 62 (1996) 2501–2507.

Navarro E., Simonet P., Normand P., Bardin R., Characterization of natural populations of *Nitrobacter* spp. using PCR/RFLP analysis of the ribosomal intergenic spacer, *Arch. Microbiol.* 157 (1992) 107–115.

Nesme X., Vanechoutte M., Orso S., Hoste B., Swings J., Diversity and genetic relatedness within genera *Xanthomonas* and *Stenotrophomonas* using restriction endonuclease site differences of PCR-amplified 16S rRNA gene, *Syst. Appl. Microbiol.* 18 (1995) 127–135.

Normand P., Utilisation des séquences 16S pour le positionnement phylétique d'un organisme inconnu, *Oceanis* 21 (1995) 31–56.

Ponsonnet C., Nesme X., Identification of *Agrobacterium* strains by PCR-RFLP analysis of pTi and chromosomal regions, *Arch. Microbiol.* 161 (1994) 300–309.

Rome S., Fernandez M.P., Brunel B., Normand P., Cleyet-Marel J.C., *Sinorhizobium medicae* sp. nov., isolated from annual *Medicago* spp., *Int. J. Syst. Bacteriol.* 46 (1996) 972–980.

Versalovic J., Koeuth T., Lupski J.R., Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes, *Nucleic Acids Res.* 19 (1991) 6823–6831.

Willems A., Collins M.D., Phylogenetic analysis of rhizobia and agrobacteria based on 16S rRNA gene sequence, *Int. J. Syst. Bacteriol.* 43 (1993) 305–313.

Williams J.G.K., Kubelic A.R., Livak K.J., Rafalski J.A., Tingey S.V., DNA polymorphisms amplified by arbitrary primers are useful as genetic markers, *Nucleic Acids Res.* 18 (1990) 6531–6535.

Yanagi M., Yamasato K., Phylogenetic analysis of the family *Rhizobiaceae* and related bacteria by sequencing of 16S rRNA gene using PCR and DNA sequencer, *FEMS Microbiol. Lett.* 107 (1993) 115–120.

Young J.P.W., Haukka K.E., Diversity and phylogeny of rhizobia, *New Phytol.* 133 (1996) 87–94.

Zabeau M., Vos P., Selective restriction fragment amplification: a general method for DNA fingerprinting, European Patent Office, publication 0534858 A1, 1993.