RESEARCH ARTICLE

INTERNATIONAL MICROBIOLOGY

Identification of *Sinorhizobium* (*Ensifer*) *medicae* based on a specific genomic sequence unveiled by M13-PCR fingerprinting

Ana Catarina Dourado,[1] Paula I. L. Alves,[1,2] Tania Tenreiro,[3] Eugénio M. Ferreira,[4] Rogério Tenreiro,[3] Paula Fareleira,[4]* M. Teresa Barreto Crespo[1,2]

[1]Institute of Experimental and Technological Biology (IBET), Oeiras, Portugal. [2]Institute of Chemical and Biological Technology (ITQB), New University of Lisbon, Oeiras, Portugal. [3]Center for Biodiversity, Functional and Integrative Genomics (BioFIG), Faculty of Sciences, University of Lisbon, Lisbon, Portugal. [4]National Institute of Agricultural Research/National Institute of Biological Resources, I.P. (INIA/INRB, IP), Oeiras, Portugal

Summary. A collection of nodule isolates from *Medicago polymorpha* obtained from southern and central Portugal was evaluated by M13-PCR fingerprinting and hierarchical cluster analysis. Several genomic clusters were obtained which, by 16S rRNA gene sequencing of selected representatives, were shown to be associated with particular taxonomic groups of rhizobia and other soil bacteria. The method provided a clear separation between rhizobia and co-isolated non-symbiotic soil contaminants. Ten M13-PCR groups were assigned to *Sinorhizobium* (*Ensifer*) *medicae* and included all isolates responsible for the formation of nitrogen-fixing nodules upon re-inoculation of *M. polymorpha* test-plants. In addition, enterobacterial repetitive intergenic consensus (ERIC)-PCR fingerprinting indicated a high genomic heterogeneity within the major M13-PCR clusters of *S. medicae* isolates. Based on nucleotide sequence data of an M13-PCR amplicon of ca. 1500 bp, observed only in *S. medicae* isolates and spanning locus Smed_3707 to Smed_3709 from the pSMED01 plasmid sequence of *S. medicae* WSM419 genome's sequence, a pair of PCR primers was designed and used for direct PCR amplification of a 1399-bp sequence within this fragment. Additional in silico and in vitro experiments, as well as phylogenetic analysis, confirmed the specificity of this primer combination and therefore the reliability of this approach in the prompt identification of *S. medicae* isolates and their distinction from other soil bacteria. [**Int Microbiol** 2009; 12(4):215-225]

Keywords: *Sinorhizobium* (*Ensifer*) *medicae* · *Medicago polymorpha* L. · rhizobia · nitrogen-fixing nodules · M13-PCR fingerprinting

Introduction

Rhizobia are a diverse and widely distributed group of soil bacteria that establish symbiotic interactions with leguminous plants through the formation of nitrogen-fixing root or stem nodules. Although the isolation of rhizobia from nodules relies on simple experimental procedures, co-isolation of non-symbiotic saprophytic soil bacteria and endophytes can occur and the recognition of these contaminants is not always straightforward [7,21,33,40]. Conventional methods for the identification of rhizobial isolates rely on the evaluation of nodulation ability by plant re-infection. These assays, however, are time-consuming and the results are often misleading, which has strengthened the need for developing new screening tools. DNA-based molecular techniques do not exclude the need for characterization by plant tests within these trials but have already proven useful in providing a pre-

*Corresponding author: P. Fareleira
INIA/INRB, IP
Quinta do Marquês, Av. da República
2784-505 Oeiras, Portugal
Tel. +351-214403500. Fax +351-214416011
E-mail: paula.fareleira@gmail.com

liminary selection of isolates with specific genomic properties [5,22,30].

DNA-typing methods have been widely employed for the study of rhizobial populations, offering greater discrimination and reproducibility for strain characterization than the usual phenotyping approaches. This is the case for techniques based on the polymerase chain reaction (PCR) that use primers directed to highly repetitive DNA sequences, such as the nucleotide sequences designated ERIC (enterobacterial repetitive intergenic consensus) or REP (repetitive extragenic palindromic), thus generating strain-specific fingerprints [6,16,23]. The M13-PCR method represents a similar approach but relies on the use of a single primer specific to a particular sequence of the gene encoding the core protein of wild-type phage M13, which has been found as multiple and hypervariable microsatellite insertions in most known genomes [29,37]. The primer anneals to these repeated insertions and, if inversely oriented sequences occur close enough to each other, then the region between them, mostly consisting of unstructured, non-coding sequences with variable length, is amplified. This method has been applied to genomic fingerprinting of different microorganisms [2,13,17,18,36] but thus far has never been used for rhizobia genotyping.

The aim of this work was to test the usefulness of PCR fingerprinting in the screening and characterization of a large collection of root-nodule isolates. The host legume selected for the study, the burr medic *Medicago polymorpha* L., is an annual, highly polymorphic pasture species that is native to the Mediterranean basin and well established in many other regions of the world with Mediterranean-type climates [19,35]. Burr medic is quite strict in its symbiotic requirements, showing a stringent specificity for *Sinorhizobium* (now *Ensifer*) *medicae*, while the closely related species *S. meliloti* only forms rudimentary and ineffective nodules in this host [11,27]. The advantages of *M. polymorpha* over other *Medicago* members include its tolerance to drought [1] and ability to grow on mildly acidic soils [15]. Such conditions are characteristic of the southern regions of Portugal, where *M. polymorpha* grows spontaneously and represents the most common species of annual medics, being highly disseminated through a wide range of habitats. In this study, M13-PCR was applied in the genomic fingerprinting of a collection of burr medic nodule isolates, obtained from either native plants or soil samples collected at different locations in those Portuguese regions. The method was combined with ERIC-PCR fingerprinting for further characterization of the isolates and with 16S rRNA gene sequencing for taxonomic evaluation. An M13-PCR fragment potentially specific for *S. medicae* was identified and a primer pair designed which, as shown by in silico and in vitro experiments as well as phylo-

genetic analysis, together provide a simple and reliable PCR protocol for the prompt identification of this species among large collections of field isolates.

Materials and methods

Strain isolation and growth conditions. Soil samples and spontaneously occurring burr medic plants were collected, between March and May 2004 from 24 locations in two regions with different geological and climatic characteristics: southern (Alentejo) and central (Estremadura) Portugal. The lack of a known historical record of cultivated pastures and the presence of spontaneous burr medic plants were a pre-requisite for the choice of sampling spots. At each site, one or several native burr medic plants showing good growth status and a healthy condition were randomly chosen. Whole plants were uprooted, packed into plastic bags, and kept at 4°C. The largest, healthier nodules in each plant were removed and surface-disinfected. Bacteria were isolated by the crushed-nodule method [39]. Additionally, surface soils were sampled at each collection site, and seeds of commercial burr medic (*Medicago polymorpha* cv. Santiago) were used to trap rhizobia as follows: surface-sterilized, pre-germinated seeds were sowed directly into 0.5 g of undiluted soil, placed in 25-ml serum-type vials containing sterile river sand moistened with 4 ml of nitrogen-free plant-growth solution, and incubated for 6–8 weeks in a controlled-environment growth cabinet, according to the method of Vincent [39]. Each soil sample was used to inoculate four seedlings. After growth, the plant roots were examined for the presence of nodules. The largest nodules were selected and used for bacterial isolation as described above. A total of 132 isolates was obtained from the samples collected at the different sites (Table 1).

Each isolate was identified with a three-letter code identifying the location of origin; one letter (uppercase) specifying the place of sampling; and the number given to the isolation in that place. Isolates were routinely grown on tryptone-yeast (TY) medium [3] at 27°C, and stock cultures were kept in 20% glycerol at −80°C. Nitrogen-fixing activity associated with *M. polymorpha* was evaluated through the effectiveness index (*E*) of Ferreira and Marques [8], comparing the mean dry weight of shoots from inoculated plants grown in nitrogen-free medium with those of non-inoculated nitrogen-supplied plants. Isolates were classified as ineffective, intermediate, and highly effective corresponding to values determined for *E* of <25%, 25–74%, and ≥75%, respectively. The following reference or type strains were included in the study: *Sinorhizobium* (*Ensifer*) *medicae* LMG 18864 (although *Sinorhizobium* and *Ensifer* are now combined as the genus *Ensifer*, the former designation of *Sinorhizobium* will be maintained in this paper), *S. meliloti* DSMZ 30135[T], *S. meliloti* ATCC 51124, *S. fredii* DSMZ 5851[T], *S. saheli* DSMZ 11273[T], *S. terangae* DSMZ 11282[T], *Rhizobium leguminosarum* DSM 30132[T] (reclassified as *R. pisi* [25]), *Azorhizobium caulinodans* DSMZ 5975[T], *Bradyrhizobium japonicum* DSMZ 30131[T], *Mesorhizobium loti* DSMZ 2626[T], *Blastobacter denitrificans* DSMZ 1113[T] and *Burkholderia phymatum* DSMZ 17167[T]. In addition, *S. medicae* strain 55Mp, *S. meliloti*, *Agrobacterium* spp., and *Burkholderia* spp. field isolates were from internal laboratory collections of INIA/ INRB, IP (Oeiras, Portugal). Strain 55Mp, a highly effective isolate from *M. polymorpha* originating from the Serpa region (Alentejo, Portugal) (E.M. Ferreira, unpublished data), was included for comparison in genomic fingerprinting studies and nitrogen fixation assays.

DNA extraction and PCR fingerprinting. Isolates were cultured in TY medium for 48 h at 27°C, and cells were harvested by centrifugation at 10,000 rpm for 5 min at 4°C. Total genomic DNA was extracted using the AquaPure genomic DNA isolation kit (BioRad). Extracted DNA from each isolate was submitted to PCR amplification of M13 microsatellites using the single primer csM13 (5′-GAGGGTGGCGGTTCT-3′) [20]. M13-PCR reactions were performed using Diamond DNA polymerase (Bioline).

Table 1. Sampling sites of soils and of spontaneous annual burr medics, geological characteristics, and number of nodule isolates

Location	Site	Geological characteristics	No. of isolates
Alter do Chão	AltA	Schist, grauwacke/sand, sandstone, clay	4
Avis	AviA	Eruptive rocks	2
Caldas da Rainha	CalA	Sand	9
Crato	CraA, B	Granite	4
Marvão	MarA	Schist, grauwacke	5
Mértola	MerA, B, C, D, E, F, G	Slate clay, grauwacke, sandstone	47
Mora	MorA	Schist, grauwacke/sand, sandstone, clay	11
Moura	MouA	Limestone with diorite and schist	3
Moura	MouB	Limestone	4
Nisa	NisA, B	Granite	7
Portalegre	PtgA	Schist, grauwacke	2
Portel	PorA	Schist, grauwacke	18
Serpa	SerA, D	Granite	13
Serpa	SerB	Diorite, garboard-strake/limestone/granite	3

Amplifications were carried out as follows: an initial denaturation step at 95°C for 5 min, 40 cycles consisting of 1 min at 94°C, 1 min at 50°C, and 2 min at 72°C, followed by a final extension step at 72°C for 7 min. ERIC sequences were amplified according to Versalovic et al. [38] using primers ERIC2 (5′-AAGTAAGTGACTGGGGTGAGCG-3′) and ERIC1r (3′-CACTTAGGGGGTCCTCGAATGTA-5′), and Immolase DNA polymerase (Bioline). Amplified fragments were separated by standard electrophoresis using 1.2% (w/v) agarose gels.

Analysis of genomic fingerprints. BioNumerics software (version 5.0, Applied Maths) was used for the alignment of fingerprinting profiles and the determination of fragment size, as well as to calculate the similarity matrices of whole densitometric curves of DNA band patterns by using the pair-wise Pearson's product-moment correlation coefficient (r value) [14]. The similarity matrices were subjected to cluster analysis according to the unweighted pair group with arithmetical average (UPGMA) algorithm. The reproducibility level of each fingerprinting method was assessed by the minimum correlation value observed in a random sample of 10% duplicate fingerprints. Genetic diversity was evaluated by calculating the Simpson's index (D) [32], which varies between 0, when all isolates belong to the same genomic group, and 1, when each isolate corresponds to a different genomic group.

PCR-amplification and sequence analysis of the 16S rRNA gene. The 16S rRNA gene for sequencing was PCR-amplified using the universal primers 27f (5′-AGAGTTTGATYMTGGCTCAG-3′) and 1492r (5′-GGYTACCTTGTTACGACTT-3′). After separation by electrophoresis in 1.2% (w/v) agarose gels, the PCR products were purified with the QIAquick PCR purification kit (Qiagen). They were then sequenced using standard protocols and either the universal primer 907r (5′-CCGTCAATTCMTTTRAGTTT-3′) or the primers pair (27f and 1492r) described above. The sequences were edited and assembled with the Vector NTI

Advance 10.3 software (Invitrogen) and then compared with the available public nucleotide databases using NCBI Mega BLAST algorithm to identify significant sequence alignments. The maximum homology criterion was applied to assign a putative identification at the species level to the isolates.

The identities of *S. medicae* isolates were also checked by PCR amplification of a 419-bp fragment of the 16S rRNA gene using specific primers for either *S. medicae* or *S. meliloti*, according to Garau et al. [11]. To further support this identification, a phylogenetic analysis was conducted using MEGA version 4 [34]. Reference sequences for the rhizobial species were selected according to TOBA information [12]. Experimental and reference sequences were aligned using the ClustalW algorithm. The aligned sequences were truncated to the minimum length of experimental sequences, corresponding to a 600-bp region spanning the 16S rRNA gene variable regions V3–V5. Based on this region, a phylogenetic tree was constructed using the neighbor-joining algorithm and the maximum composite likelihood model. The bootstrap method was used to test the inferred phylogeny. The obtained 16S rRNA gene sequences, from the order Rhizobiales, were deposited in GenBank (Table 2).

Nucleotide sequence determination of an M13-PCR fragment and PCR amplification of a *S. medicae* target amplicon. M13-PCR was carried out as described above, except that Taq DNA polymerase was used and the final extension step of the PCR protocol was extended to 10 min. The amplified products were separated by overnight 1.2 % (w/v) agarose gel electrophoresis. Selected M13-PCR fragments were excised from the agarose gel, purified using NucleoSpin Extract II (Macherey-Nagel), and concentrated by ethanol precipitation. The purified and concentrated DNA fragments were cloned using pGEM-T Easy Vector System I (Promega) according to the manufacturer's recommendations. The fragments were sequenced using the universal primers SP6 (5′-CATTTAGGTGACACTATAG-3′) and T7 (5′-TAATACGACTCACTATAGGG-3′), and the sequences edited and analyzed using the same tools as described above.

Table 2. GenBank accession numbers of 16S rRNA gene partial sequences from Rhizobiales obtained in this work

Isolate	Accession number	Isolate	Accession number
MerE-6	GU143692	SerB-2	GU143701
MerF-3	GU143693	MerB-2	GU143702
NisA-4	GU143694	MerF-1	GU143703
AltA-2	GU143695	MerG-3	GU143704
MorA-12	GU143696	MerE-11	GU143705
MorA-5	GU143697	NisB-1	GU143706
CalA-11	GU143698	MerA-7	GU143707
AviA-3	GU143699	MarA-4	GU143708
MouA-1	GU143700	MarA-6	GU143709

The primer pair SMED01-f (5′-GCGCGTAGTTCTGAAAGACC-3′) and SMED01-r (5′-GGTTCTGGACGATCGTGTTT-3′) was designed using the software Primer3 [http://frodo.wi.mit.edu] [28] and Serial Cloner 1.3 (Serial Basics) and specifically targeted a partial sequence (1399 bp) of a ca. 1500-bp M13-PCR genomic fragment in *S. medicae*. The NCBI Mega BLAST algorithm was used to compare primer sequences with sequences in public nucleotide databases in order to determine significant identities. These primers were also tested through in silico PCR experiments with the program In Silico Simulation of Molecular Biology Experiments [http://insilico.ehu.es] [4], using the available genomes of genera of the family Rhizobiaceae as well as those of type species of type genera of all the other families of the Rhizobiales.

Genomic BLAST was performed in the NCBI database against all available bacterial genomes and using the 1399-bp sequence as the query. A phylogenetic tree was constructed with neighbor-joining, with 0.75 as the maximum allowed fraction of mismatched bases in the aligned region between any pair of sequences. Additional neighbor-joining trees were also constructed using locus Smed_3707, Smed_3708, Smed_3709, and Smed_3710 from *S. medicae* WSM419 pSMED01 plasmid as the query against all the bacterial sequences available in KEGG database [http://www.genome.jp/kegg] [24].

In vitro PCR was performed using 1 U Taq DNA polymerase, 1 pM of SMED01-f and SMED01-r primer, and the following PCR protocol: initial denaturation at 94°C for 5 min, 34 cycles at 94°C for 45 s, 62°C for 30 s and 72°C for 1.5 min, and a final 7-min extension cycle at 72°C. Amplified fragments were separated by standard electrophoresis using 1.2% (w/v) agarose gels.

Results

M13-PCR fingerprinting of nodule isolates.

A total of 132 isolates were recovered from root nodules of burr medic and submitted to genotyping by M13-PCR. Genomic amplification with primer csM13 generated a variable number of PCR products, typically consisting of 2–13 bands distributed within the 200- to 3200-bp region. Replicate M13-PCR analysis of randomly chosen isolates produced PCR profiles with an average similarity of 90%, indicating a good level of reproducibility for the method.

Hierarchic analysis of M13-PCR banding patterns allowed categorization of the isolates into 28 clusters at 65% minimum similarity (Fig. 1). Although a significant number of isolates gave rise to individual profiles, most clustered together in larger groups of up to 36 individuals. Each of the major clusters included isolates of different geographic origins and obtained from either spontaneous field plants or soil-inoculated laboratory plants.

Twenty-eight isolates were randomly chosen among the collection and further characterized by determining the partial nucleotide sequence of their 16S rRNA genes (Fig. 1). BLAST analysis of these sequences resulted in the assignment of 12 of the isolates to the genus *Sinorhizobium*, based on at least 99% identity with the 16S rRNA gene sequences of *S. medicae* and *S. meliloti*. In fact, a high level of sequence similarity was previously reported for the 16S rRNA genes of *S. meliloti* LMG 6133[T] and *S. medicae* A 321[T] [27]. Alignment of the 16S rRNA gene sequences of *S. medicae* strain WSM419 (GenBank accession no. CP000738) and *S. meliloti* LMG 6133[T] (GenBank accession no. X67222) identified four nucleotide substitutions. These sequences were further aligned with the partial sequences of the *Sinorhizobium* isolates, which in the differing positions between the two species were visually checked.

Additionally, each *Sinorhizobium* isolate was PCR-analyzed using the set of primers proposed by Garau et al. [11] for the specific amplification of a 419-bp region in the 16S rRNA gene of either *S. medicae* or *S. meliloti*. This combined strategy allowed the identification of all 12 isolates assigned to the *Sinorhizobium* genus as *S. medicae*. Six other 16S rRNA gene sequences from nodule isolates were found to be closely related to those of members of the genus *Rhizobium*, with no significant relationship at the species level (data not shown). Sequences from other isolates were related to known 16S rRNA gene sequences from members of the γ-subgroup of Proteobacteria (*Pseudomonas*) and gram-positive bacteria (*Bacillus*).

A phylogenetic tree was constructed that included reference sequences from Rhizobiales. It confirmed the clustering of the twelve *S. medicae* isolates in a clade with the reference strain *S. medicae* WSM419 (data not shown). Five of the *Rhizobium* isolates were closely associated in a clade with *R. leguminosarum*, and the remainder distantly related with *R. giardinii*. It was observed that, within the sample of isolates selected for identification, those included in the same M13-PCR cluster were also assigned to the same genus/species, each taxonomic group having several different M13-PCR profiles (Fig. 1). Thus, through similarity analysis of the M13
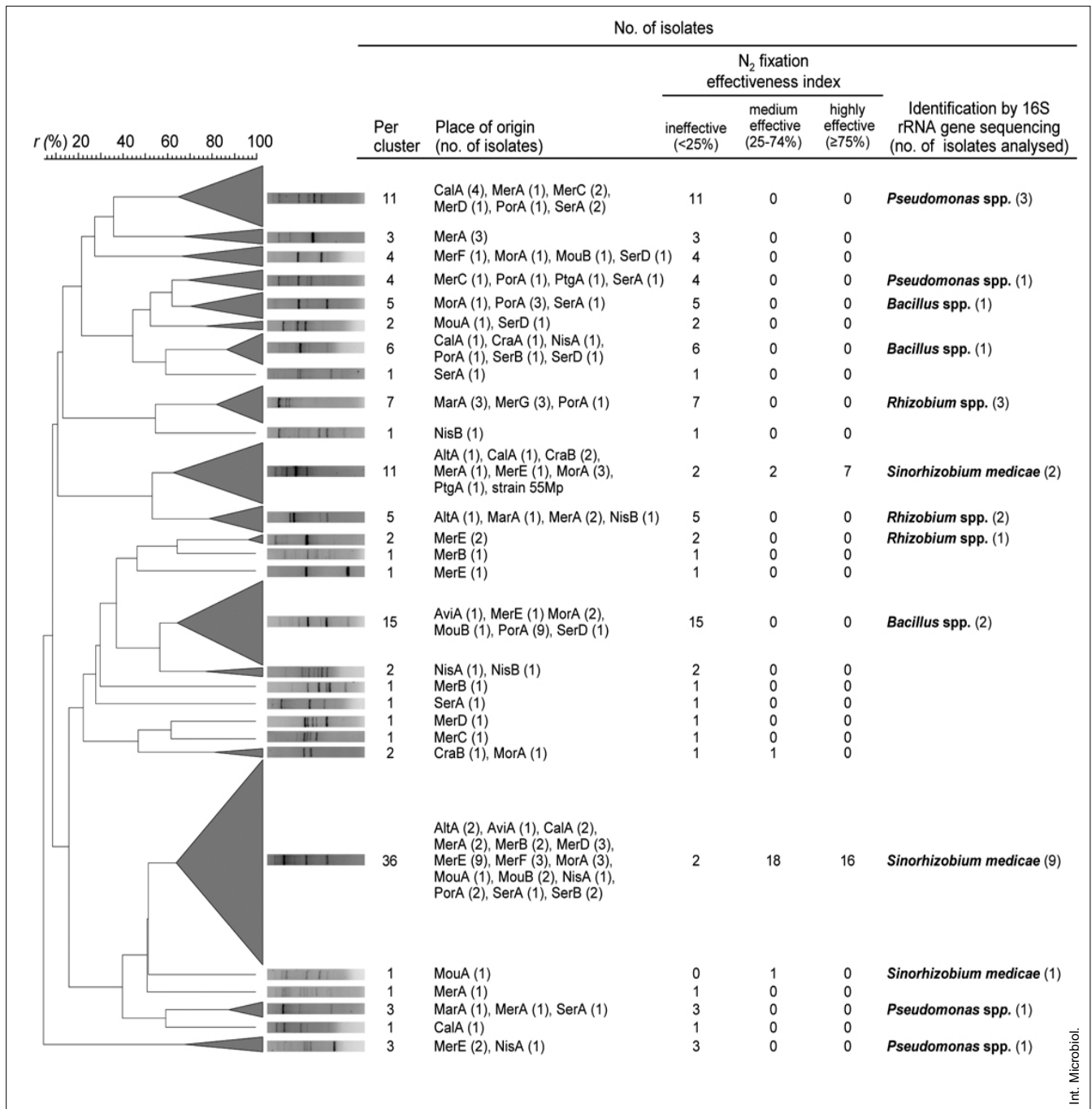
| | | No. of isolates | | | |
| | | N₂ fixation effectiveness index | | | Identification by 16S rRNA gene sequencing (no. of isolates analysed) |
Per cluster	Place of origin (no. of isolates)	ineffective (<25%)	medium effective (25-74%)	highly effective (≥75%)	
11	CalA (4), MerA (1), MerC (2), MerD (1), PorA (1), SerA (2)	11	0	0	*Pseudomonas* spp. (3)
3	MerA (3)	3	0	0	
4	MerF (1), MorA (1), MouB (1), SerD (1)	4	0	0	
4	MerC (1), PorA (1), PtgA (1), SerA (1)	4	0	0	*Pseudomonas* spp. (1)
5	MorA (1), PorA (3), SerA (1)	5	0	0	*Bacillus* spp. (1)
2	MouA (1), SerD (1)	2	0	0	
6	CalA (1), CraA (1), NisA (1), PorA (1), SerB (1), SerD (1)	6	0	0	*Bacillus* spp. (1)
1	SerA (1)	1	0	0	
7	MarA (3), MerG (3), PorA (1)	7	0	0	*Rhizobium* spp. (3)
1	NisB (1)	1	0	0	
11	AltA (1), CalA (1), CraB (2), MerA (1), MerE (1), MorA (3), PtgA (1), strain 55Mp	2	2	7	*Sinorhizobium medicae* (2)
5	AltA (1), MarA (1), MerA (2), NisB (1)	5	0	0	*Rhizobium* spp. (2)
2	MerE (2)	2	0	0	*Rhizobium* spp. (1)
1	MerB (1)	1	0	0	
1	MerE (1)	1	0	0	
15	AviA (1), MerE (1) MorA (2), MouB (1), PorA (9), SerD (1)	15	0	0	*Bacillus* spp. (2)
2	NisA (1), NisB (1)	2	0	0	
1	MerB (1)	1	0	0	
1	SerA (1)	1	0	0	
1	MerD (1)	1	0	0	
1	MerC (1)	1	0	0	
2	CraB (1), MorA (1)	1	1	0	
36	AltA (2), AviA (1), CalA (2), MerA (2), MerB (2), MerD (3), MerE (9), MerF (3), MorA (3), MouA (1), MouB (2), NisA (1), PorA (2), SerA (1), SerB (2)	2	18	16	*Sinorhizobium medicae* (9)
1	MouA (1)	0	1	0	*Sinorhizobium medicae* (1)
1	MerA (1)	1	0	0	
3	MarA (1), MerA (1), SerA (1)	3	0	0	*Pseudomonas* spp. (1)
1	CalA (1)	1	0	0	
3	MerE (2), NisA (1)	3	0	0	*Pseudomonas* spp. (1)

Int. Microbiol.

Fig. 1. Cluster analysis of M13-PCR genomic fingerprints obtained for 132 root nodule isolates from *Medicago polymorpha*, native to southern Portuguese locations. The dendogram was constructed using the Pearson correlation coefficient and the UPGMA clustering method. A representative fingerprint is shown on the right hand side of each cluster at 65% minimum similarity. The effectiveness index for N₂ fixation associated with *M. polymorpha* was determined by the mean dry weight of shoots from inoculated plants grown in nitrogen-free medium, shown as the percentage of the value obtained with non-inoculated nitrogen-supplied plants. Genus/species assignment was based on 16S rRNA gene sequencing data.

PCR profiles, the isolates were clustered according to their taxonomic affiliation. Based on these results, 48 nodule isolates in ten M13-PCR groups at 80% minimum similarity were assigned to the species *S. medicae*; these groups were designated M13-1 through M13-10 (Fig. 2A). Analysis of the M13-PCR pattern of *S. medicae* strain LMG 18864 led to its inclusion in group M13-8. This group consisted of six nodule isolates and was related at 74.3 % similarity to group
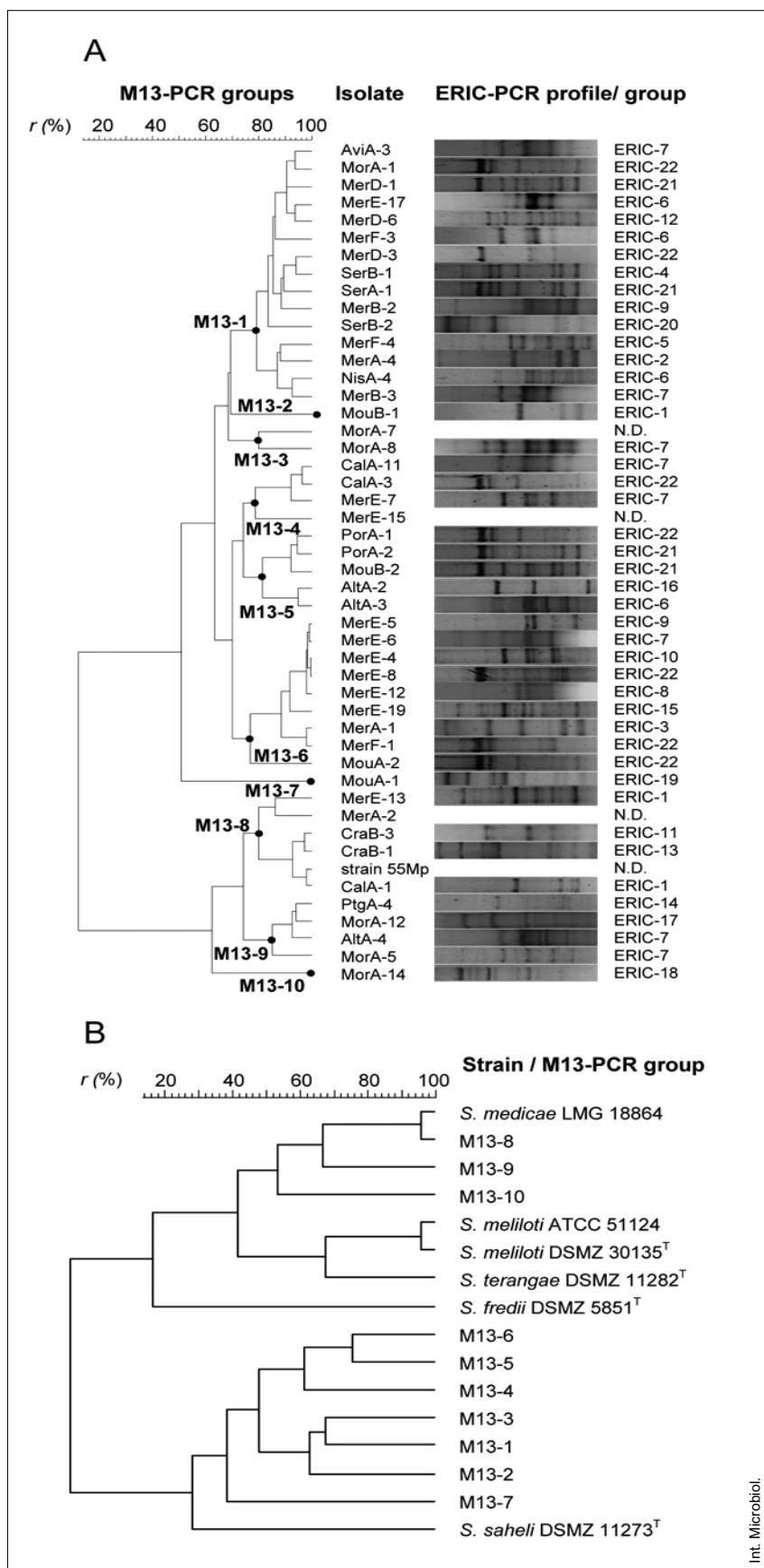
Fig. 2. (**A**) Dendogram (UPGMA with the Pearson correlation coefficient) of M13-PCR fingerprints of *Sinorhizobium medicae* isolates, showing the ten clusters defined at 80% minimum similarity. For each isolate, the corresponding ERIC-PCR profile and ERIC-PCR group are indicated at the right. ND, not determined. (**B**) Simplified dendogram showing the clustering of one representative isolate of each M13-PCR cluster and reference/type strains of six *Sinorhizobium* species.

M13-9, but it was quite distinct from the remaining groups, in which the majority of the isolates were included. Analysis of reference/type strains belonging to other species of the genus *Sinorhizobium* revealed diverse M13-PCR profiles lacking significant correlation with any group of isolates (Fig. 2B).

The results from nitrogen fixation assays in association with *M. polymorpha* generally agreed with the hierarchy derived from the M13-PCR analysis (Fig. 1). With a few exceptions, most isolates presenting M13-PCR patterns assigned to *S. medicae* were consistently responsible for the formation of intermediate or highly effective nitrogen-fixing root nodules in *M. polymorpha* test plants upon re-inoculation. Conversely, the majority of the remaining isolates were either unable to nodulate or ineffective for nitrogen fixation with *M. polymorpha*. The nitrogen-fixing activities of *S. medicae* isolates were variable, but no correlation was found between the levels of nitrogen fixation and particular M13-PCR genotypes.

Distribution and diversity of sinorhizobial M13-PCR groups. The distribution of *S. medicae* isolates among different M13-PCR groups was neither related to geographic location nor to the geological characteristics of their sites of origin. The relatively small number of M13-PCR groups found within the set of *S. medicae* isolates (10 profiles for 48 isolates) suggested a low discriminating capability of the technique. The Simpson's diversity index of the M13-PCR genotypes within this population was 0.841. To increase the resolution of the analysis, 44 isolates presenting M13-PCR profiles assigned to *S. medicae* were further characterized by ERIC-PCR. Hierarchical analysis of the obtained ERIC-PCR profiles clustered the isolates into 22 clusters at 80% minimum similarity. The Simpson's diversity index of the ERIC-PCR profiles was 0.934, showing the technique's higher discriminating power compared to M13-PCR analysis.

Based upon data from the ERIC-PCR analysis, we concluded that, within most of the *S. medicae* M13-PCR groups, the isolates were genetically heterogeneous. For example, within group M13-6, comprising nine isolates mostly originating from the same geographic region (Mértola), seven different ERIC-PCR profiles were observed ($D = 0.917$); similarly, 10 out of 15 isolates within profile M13-1 differed in their ERIC-PCR band patterns ($D = 0.943$). Conversely, isolates distributed into different M13-PCR groups shared similar ERIC-PCR profiles. This was true for four isolates (MorA-1, CalA-3, PorA-1, and MouA-2) originating from different geographic regions and presenting distinct M13-PCR profiles, but with highly similar (96.5%) ERIC-PCR fingerprints.

Identification of a M13-PCR fragment specific for *S. medicae*. An M13-PCR amplification product of ca. 1500 bp was observed in all *S. medicae* isolates except two (MerA-2 and MerE-13), but was absent in members of other species or genera, suggesting that it was specific for *S. medicae* (Fig. 3A). The DNA fragment retrieved from a randomly selected isolate (MerE-6) was sequenced (GenBank accession no. GU143691), and database searches revealed 100% similarity with a 1468-bp segment spanning *loci* Smed_3707 to Smed_3709 in the pSMED01 megaplasmid of *S. medicae* WSM419 (GenBank accession no. CP000739) (Fig. 3B). These loci belong to a gene cluster that putatively encodes a glycerol-3-phosphate ABC transport system including UgpC protein [Smed_3707], an ATP-binding subunit; UgpE [Smed_3708] and UgpA [Smed_3709], inner membrane components; and UgpB [Smed_3710], a periplasmic/extracellular solute-binding protein.

Analysis of the published sequence of pSMED01 confirmed the presence of two nucleotide sequences in the flanking regions of this fragment with, respectively, 80% (149404–149418 bp) and 67% (150857–150871 bp) identity to the csM13 primer sequence. Based on the nucleotide sequence of this M13 amplicon, a primer pair (SMED01-f and SMED01-r) was designed to amplify a 1399-bp fragment. The SMED01-f primer also showed highly significant sequence alignments (20/20) with plasmid and/or chromosomal sequences of *S. meliloti*, *R. leguminosarum* bv. *trifolii* and bv. *viciae*, *Roseobacter denitrificans*, and *Thermotoga petrophila*, but the SMED01-r primer seems to be highly specific for *S. medicae*. The specificity of the SMED01 primer combination was confirmed after a genomic BLAST of the 1399-bp sequence of isolate MerE-6 against all 1106 bacterial genomes available in the NCBI, as revealed by its resolved position with *S. medicae* WSM419 in the phylogenetic tree depicted in Fig. 3C. Although a phylogenetic analysis of each *ugp* gene in the KEGG database evidenced the wide distribution of all loci amongst gram-negative bacteria, a well-resolved clade including *S. medicae* and *S. meliloti* was constant in all cases (data not shown).

In silico PCR experiments further reinforced the potential applicability of the proposed primer pair, since no PCR product was obtained with any of the available genomes of genera comprising the family Rhizobiaceae, as well as those of type species of type genera comprising all the other families of the Rhizobiales, except for *S. medicae* WSM419, for which a single 1399-bp product was predicted (Fig. 3D). In vitro PCR was then performed using a representative assortment of previously identified isolates, including the 48 *S. medicae* isolates obtained in this work, *Rhizobium* spp., *Agrobacterium* spp., *Burkholderia* spp., *Pseudomonas* spp., and
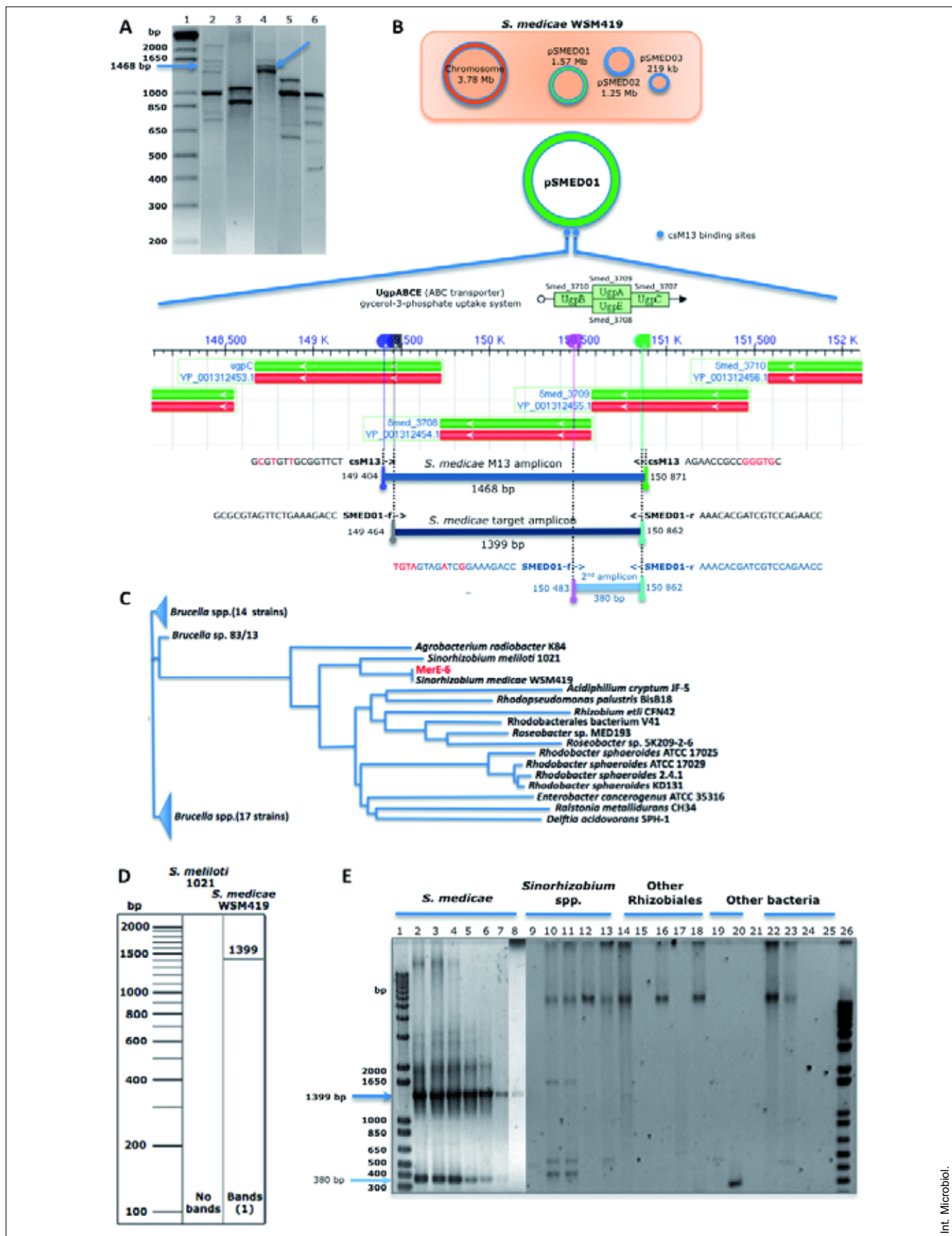
Bacillus spp., as well as type and reference strains of several rhizobial and non-rhizobial species (Fig. 3E). A PCR product of the expected size was obtained with all *S. medicae* isolates originating from different locations in Portugal (including isolates MerE-13 and MerA-2 typed as *S. medicae* but lacking the 1468-bp fragment amplified with primer csM13) and the reference strain LMG 18864. In contrast, no amplification product was obtained with *S. meliloti* or any other species used for analysis, thus indicating that this primer pair was suitable for the direct detection of *S. medicae* strains. Unexpectedly, a fainter 380-bp amplicon was also present in *S. medicae* isolates (Fig. 3E). However, a second priming site with six mismatches was found in locus Smed_3708 for primer SMED01-f (Fig. 3B), which could explain this secondary amplicon.

Discussion

The usefulness of M13-PCR fingerprinting for the identification and discrimination of microbial isolates has already been well demonstrated. For instance, the method has been successfully applied in the genotypic analysis of yeast and bacterial clinical isolates, allowing not only their assignment to particular species but also the differentiation between strains within species [18,31], thus providing a valuable tool in epidemiological studies [9,10]. In one of those studies, the screening of 88 clinical isolates of *Pseudomonas aeruginosa* by PCR-fingerprinting using csM13 primer has allowed the establishment of 50 different genogroups, each one clustering isolates with band patterns of ≥80% similarity [10]. The typing and traceability of cyanobacteria within a freshwater reservoir has been also achieved using M13-PCR fingerprinting combined with ERIC-PCR [36]. In this work, the M13-PCR fingerprinting approach provided a firm separation between sinorhizobial isolates on the one hand and other rhizobial species and non-nodulating soil bacteria on the other,

thus clustering the target microorganisms in clearly distinct genomic groups.

We identified ten distinct M13-PCR profile-groups among 48 *S. medicae* isolates, although no relationship could be established between the distribution of *S. medicae* isolates among these M13-PCR genomic groups and either the geographic location, the climate, or the geological characteristics of their places of origin. Moreover, in spite of the observed dispersion, genotypic variations at the strain level could be detected only by the combined use of a higher discriminatory method, which restricted the applicability of this approach in diversity analysis. An interesting result from the M13-PCR clustering analysis was the observation that one of the groups (M13-8), which included the reference strain of *S. medicae*, was more closely related to type or reference strains of other species of *Sinorhizobium* than to the remaining groups of *S. medicae* isolates. This suggested that, although sharing characteristic properties of the species, such as the ability to establish nitrogen-fixing symbiosis with *M. polymorpha* and the appropriate 16S rRNA gene nucleotide sequence, a significant proportion of native *S. medicae* from Portuguese soils might have genomic features different from those of other members of the same species and from other species of the same genus. These differences might be associated with a specific adaptation response to a peculiar and restrictive habitat—as is the case with other, unrelated bacteria [41]—thus making the role of saprophytic stress resistance, competitiveness for nodulation, and host-range preferences interesting topics that certainly deserve further investigation. Our results also raise the possibility that sinorhizobial reference or type strains might be genomically unrepresentative of field isolates assigned to the same species, strengthening the necessity of carefully using the information retrieved from comparative studies.

A major outcome of the M13-PCR fingerprinting approach was the finding that primer csM13 amplified a 1468-bp DNA fragment with 100% sequence identity to a

---

← **Fig. 3.** (**A**) M13-PCR profiles of sinorhizobia. Lane 1: 1-kb Plus DNA Ladder (Invitrogen); lanes 2–6: *S. medicae* LMG 18864, *S. meliloti* DSMZ 30135[T], MerE-6, MerE-13, MerA-2. (**B**) Structural organization of the 148- to 152-kb region of pSMED01 plasmid of *S. medicae* WSM419 encompassing the UgpABCE sn-glycerol-3-phosphate uptake system. The 1468-bp M13 amplicon, the 1399-bp target, and the 380-bp secondary amplicon are also depicted. Plasmid sequences at annealing sites of primers csM13, SMED01-f and SMED01-r are indicated, with nucleotide mismatches in red. (**C**) Phylogenetic tree of the 1399-bp *S. medicae* target amplicon obtained by genome BLAST against all 1106 available bacterial genomes in the NCBI database. (**D**) Graphical representation of in silico PCR using SMED01-f and SMED-r primers with available sequenced genomes of Rhizobiales. (**E**) PCR amplification of genomic DNA from root nodule isolates and type or reference strains of several soil bacteria using primers SMED01-f and SMED01-r. The *S. medicae* 1399-bp target amplicon and the 380-bp second amplicon are indicated by the arrows. Lanes 1 and 26, 1-kb Plus DNA Ladder (Invitrogen); lanes 2–8, *S. medicae* (strain LMG 18864, and isolates MerE-4, MerE-5, MerE-6, MerE-12, MerE-13, MerA-2, respectively); lanes 9-11, *S. meliloti* (strain DSMZ 30135T, and two field isolates, respectively); lane 12, *S. fredii* DSMZ 5851T; lane 13, *S. terangae* DSMZ 11282T; lane 14, *Rhizobium pisi* DSMZ 30132T; lane 15, *Rhizobium* spp. (isolate MerG-3); lane 16, *Azorhizobium caulinodans* DSMZ 5875T; lane 17, *Bradyrhizobium japonicum* DSMZ 30131T; lane 18, *Mesorhizobium loti* DSMZ 2626T; lane 19, *Burkholderia phymatum* DSMZ 17167T; lane 20, *Burkholderia* spp. (field isolate); lane 21, *Blastobacter denitrificans* DSMZ 1113T; lanes 22-23, *Pseudomonas* spp. (isolates SerA-7 and CalA-5, respectively); lanes 24-25, *Bacillus* spp. (isolates SerD-4 and CalA-7, respectively).

region in the pSMED01 megaplasmid of *S. medicae* WSM419, which is part of an ABC transporter operon. This sequence and the PCR primer combination SMED01-f and SMED01-r developed in this work were shown, by in silico and in vitro analysis, to be sufficiently specific to allow the reliable distinction of *S. medicae* from other rhizobial and non-nodulating soil bacteria.

Notwithstanding the considerable range of geological types and geographic distribution covered with the group of isolates tested in this work, all isolates originated from southern and central Portugal (except the reference strain LMG 18867), which might question the universality of the proposed method. However, all the work performed in silico in genes and genome databases, as well as the negative results obtained with all tested strains of non-symbiotic soil bacteria, including the common nodule contaminant *Agrobacterium* spp. [7,21], contradict this possibility and as such validates the method's use in routine analysis.

Several molecular techniques based on sequence particularities of the 16S rRNA gene have been previously proposed for the identification of *S. medicae* and its distinction from the related species *S. meliloti*. These approaches rely on the use of species-specific oligonucleotide probes complementary to 16S rRNA gene sequences [26], restriction fragment length polymorphism (RFLP) analysis of amplified 16S rDNA sequences [42], or PCR amplification using 16S-rDNA-specific primers [11]. The PCR protocol described in this work represents a novel approach for the direct identification of *S. medicae* strains among field isolates, providing an additional tool for the already implemented 16S-rRNA-based methods.

References

1. Andrew WD, Hely FW (1960) Frequency of annual species of *Medicago* on the major soil groups of the Macquarie region of New South Wales. Aust J Agr Res 11:705-714

2. Bartie KL, Williams DW, Wilson MJ, Potts AJC, Lewis MAO (2001) PCR fingerprinting of *Candida albicans* associated with chronic hyperplastic candidosis and other oral conditions. J Clin Microbiol 39:4066-4075

3. Beringer JE (1974) R factor transfer in *Rhizobium leguminosarum*. J Gen Microbiol 84:188-198

4. Bikandi J, San Millán R, Rementeria A, Garaizer J (2004) *In silico* analysis of complete bacterial genomes: PCR, AFLP-PCR, and endonuclease restriction. Bioinformatics 20:798-799

5. Bromfield ESP, Barran LR, Wheatcroft R (1995) Relative genetic structure of a population of *Rhizobium meliloti* isolated directly from soil and from nodules of alfalfa (*Medicago sativa*) and sweet clover (*Melilotus alba*). Mol Ecol 4:183-188

6. De Bruijn FJ (1992) Use of repetitive (repetitive extragenic palindromic and enterobacterial repetitive intergenic consensus) sequences and the polymerase chain reaction to fingerprint the genomes of *Rhizobium meliloti* isolates and other soil bacteria. Appl Environ Microbiol 58:2180-2187

7. De Lajudie P, Willems A, Nick G, et al. (1999) *Agrobacterium* bv. 1 strains isolated from nodules of tropical legumes. Syst Appl Microbiol 22:119-132

8. Ferreira EM, Marques JF (1992) Selection of Portuguese *Rhizobium leguminosarum* bv. *trifolii* strains for production of legume inoculants. I. Screening for effectiveness in laboratory conditions. Plant Soil 147:151-158

9. Fonseca AP, Correia P, Extremina CI, Sousa JC, Tenreiro R, Barros H (2008) Molecular epidemiology of *Pseudomonas aeruginosa* clinical isolates from Portuguese central hospital. Folia Microbiol 53:540-546

10. Fonseca AP, Correia P, Sousa JC, Tenreiro R (2007) Association patterns of *Pseudomonas aeruginosa* clinical isolates as revealed by virulence traits, antibiotic resistance, serotype and genotype. FEMS Immunol Med Microbiol 51:505-516

11. Garau G, Reeve WG, Brau L, et al. (2005) The symbiotic requirements of different *Medicago* spp. suggest the evolution of *Sinorhizobium meliloti* and *S. medicae* with hosts differentially adapted to soil pH. Plant Soil 176:263-277

12. Garrity GM, Lilburn TG, Cole JR, Harrison SH, Euzeby J, Tindall BJ (2007) Taxonomic outline of the Bacteria and Archaea, Release 7.7. Michigan State University Board of Trustees.

13. Grif K, Karch H, Schneider C, et al. (1998) Comparative study of five different techniques for epidemiological typing of *Escherichia coli* O157. Diagn Microbiol Infect Dis 32:165-176

14. Häne BG, Jäger K, Drexler H (1993) The Pearson product-moment correlation coefficient is better suited for identification of DNA fingerprint profiles than band matching algorithms. Electrophoresis 14:967-972

15. Howieson JG, Ewing MA (1989) Annual species of *Medicago* differ greatly in their ability to nodulate on acid soils. Aust J Agr Res 40:842-850

16. Judd AK, Schneider M, Sadowsky MJ, de Bruijn FJ (1993) Use of repetitive sequences and the polymerase chain reaction to classify genetically related *Bradyrhizobium japonicum* serocluster 123 strains. Appl Environ Microbiol 59:1702-1708

17. Kure CF, Abeln ECA, Holst-Jensen A, Skaar I (2002) Differentiation of *Penicillium commune* and *Penicillium palitans* isolates from cheese and indoor environments of cheese factories using M13 fingerprinting. Food Microbiol 19:151-157

18. Latouche GN, Daniel H-M, Lee OC, Mitchell TG, Sorrell TC, Meyer W (1997) Comparison of use of phenotypic and genotypic characteristics for identification of species of the anamorph genus *Candida* and related teleomorph yeast species. J Clin Microbiol 35:3171-3180

19. Lesins KA., Lesins I (1979) Genus *Medicago* (*Leguminosae*): A taxogenetic study. Kluwer Academic Pub., Dordrecht, Netherlands

20. Meyer W, Lieckfeldt E, Kuhls K, Freedman EZ, Borner T, Mitchell TG (1993) DNA and PCR-fingerprinting in fungi. EXS (Basel) 67:311-320

21. Mhamdi R, Laguerre G, Aouani ME, Mars M, Amarger N (2002) Different species and symbiotic genotypes of field rhizobia can nodulate *Phaseolus vulgaris* in Tunisian soils. FEMS Microbiol Ecol 41:77-84

22. Niemann S, Dammann-Kalinowski T, Nagel A, Pühler A, Selbitschka W (1999) Genetic basis of enterobacterial repetitive intergenic consensus (ERIC)-PCR fingerprint pattern in *Sinorhizobium meliloti* and identification of *S. meliloti* employing PCR primers derived from an ERIC-PCR fragment. Arch Microbiol 172:22-30

23. Niemann S, Pühler A, Tichy HV, Simon R, Selbitschka W (1997) Evaluation of the resolving power of three different DNA fingerprinting

methods to discriminate among isolates of a natural *Rhizobium meliloti* population. J Appl Microbiol 82:477-484

24. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M (1999) KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acid Res 27:29-34

25. Ramírez-Bahena MH, García-Fraile P, Peix A, et al. (2008) Revision of the taxonomic status of the species *Rhizobium leguminosarum* (Frank 1879) Frank 1889[AL], *Rhizobium phaseoli* Dangeard 1926[AL] and *Rhizobium trifolii* Dangeard 1926[AL]. *R. trifolii* is a later synonym of *R. leguminosarum*. Reclassification of the strain *R. leguminosarum* DSM 30132 (=NCIMB 11478) as *Rhizobium pisi* sp. nov. Int J Syst Evol Microbiol 58:2484-2490

26. Rome S, Cleyet-Marel JC, Materon LA, Normand P, Brunel B (1997) Rapid identification of *Medicago* nodulating strains by using two oligo-nucleotide probes complementary to 16S rDNA sequences. Can J Microbiol 43:854-861

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

28. Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics methods and protocols: Methods in molecular biology. Humana Press, Totowa, NJ, pp 365-386

29. Ryskov AP, Jincharadze AG, Prosnyak MI, Ivanov PL, Limborska SA (1988) M13 phage DNA as a universal marker for DNA fingerprinting of animals, plants and microrganisms. FEBS Lett 233:388-392

30. Sánchez-Contreras M, Lloret J, Martín M, Villacieros M, Bonilla I, Rivilla R (2000) PCR use of highly conserved DNA regions for identification of *Sinorhizobium meliloti*. Appl Environ Microbiol 66: 3621-3623

31. Schönian G, Meusel O, Tietz HJ, et al. (1993) Identification of clinical strains of *Candida albicans* by DNA-fingerprinting with the polymerase chain reaction. Mycoses 36:171-179

32. Simpson A (1949) Measurement of diversity. Nature 163:688

33. Sturz AV, Christie BR, Matheson BG, Nowak J (1997) Biodiversity of endophytic bacteria which colonize red clover nodules, roots, stems and foliage and their influence on growth. Biol Fertil Soils 25:13-19

34. Tamura K, Dudley J, Nei M & Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596-1599

35. Tivoli B, Baranger A, Sivasithamparam K, Barbetti MJ (2006) Annual *Medicago*: from a model crop challenged by a spectrum of necrotrophic pathogens to a model plant to explore the nature of disease resistance. Ann Bot 98:1117-1128

36. Valério E, Chambel L, Paulino S, Faria N, Pereira P, Tenreiro R (2009) Molecular identification, typing and traceability of cyanobacteria from freshwater reservoirs. Microbiology 155:642-656

37. Vassart G, Georges M, Monsieur R, Brocas H, Lequarre AS, Christophe D (1987) A sequence in M13 phage detects hypervariable minisatellites in human and animal DNA. Science 235:683-684

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

39. Vincent JM (1970) A manual for the practical study of root-nodule bacteria. IBP Handbook no.15. Blackwell Scientific Publications, Oxford, UK

40. Zakhia F, Jeder H, Willems A, Gillis M, Dreyfus B, de Lajudie P (2006) Diverse bacteria associated with root nodules of spontaneous legumes in Tunisia and first report for *nifH*-like gene within the genera *Microbacterium* and *Starkeya*. Microb Ecol 51:375-393

41. Zé-Zé L, Chelo IM, Tenreiro R (2008) Genome organization in *Oenococcus oeni* strains studied by comparison of physical and genetic maps. Int Microbiol 11:247-244

42. Zribi K, Mhamdi R, Huguet T, Aouani ME (2004) Distribution and genetic diversity of rhizobia nodulating natural populations of *Medicago truncatula* in Tunisian soils. Soil Biol Biochem 36:903-908