International Journal of Systematic and Evolutionary Microbiology (2008), **58**, 200–214 **DOI** 10.1099/ijs.0.65392-0

Advantages of multilocus sequence analysis for taxonomic studies: a case study using 10 housekeeping genes in the genus *Ensifer* (including former *Sinorhizobium*)

Miet Martens, Peter Dawyndt, Renata Coopman, Monique Gillis, Paul De Vos and Anne Willems

Laboratorium voor Microbiologie (WE10), Universiteit Gent, K. L. Ledeganckstraat 35, B-9000 Gent, Belgium

Correspondence
Anne Willems
Anne.Willems@UGent.be

There is a need for easy, practical, reliable and robust techniques for the identification and classification of bacterial isolates to the species level as alternatives to 16S rRNA gene sequence analysis and DNA–DNA hybridization. Here, we demonstrate that multilocus sequence analysis (MLSA) of housekeeping genes is a valuable alternative technique. An MLSA study of 10 housekeeping genes (*atpD*, *dnaK*, *gap*, *glnA*, *gltA*, *gyrB*, *pnp*, *recA*, *rpoB* and *thrC*) was performed on 34 representatives of the genus *Ensifer*. Genetic analysis and comparison with 16S and 23S rRNA gene sequences demonstrated clear species boundaries and a higher discrimination potential for all housekeeping genes. Comparison of housekeeping gene sequence data with DNA–DNA reassociation data revealed good correlation at the intraspecies level, but indicated that housekeeping gene sequencing is superior to DNA–DNA hybridization for the assessment of genetic relatedness between *Ensifer* species. Our MLSA data, confirmed by DNA–DNA hybridizations, support the suggestion that *Ensifer xinjiangensis* is a later heterotypic synonym of *Ensifer fredii*.

INTRODUCTION

Nowadays, bacterial classification involves techniques to determine both phenotypic and genotypic characteristics. Of the genotypic methods, 16S rRNA gene sequencing and genomic DNA–DNA reassociation serve as 'gold standards' for bacterial species determination (Stackebrandt & Goebel, 1994). DNA–DNA hybridization involves a pairwise comparison of two entire genomes and reflects the overall sequence similarity between them. Currently, a species is defined as a set of strains with approximately 70 % or greater DNA–DNA relatedness and with 5 °C or less ΔT_m. Phenotypic characteristics should be in agreement with this definition (Stackebrandt *et al.*, 2002; Wayne *et al.*, 1987). However, DNA–DNA hybridization is a technically challenging, labour-intensive and time-consuming method. Also, it is not possible to establish a central database, mainly because the technique provides a non-cumulative, relative DNA relatedness value, but also because of technical non-uniformity and variability between different laboratories and methodologies (for a recent review of the different methods of DNA–DNA hybridization, see Rosselló-Mora, 2006). Moreover, the technique has the drawback that hybridization values of 50 % or less are less informative and therefore DNA–DNA hybridizations are not suitable for the estimation of genetic distances between distantly related species (Owen & Pitcher, 1983).

Together with DNA–DNA hybridization, sequence analysis of the 16S rRNA gene is also standard practice in bacterial taxonomy. In contrast to the former technique, 16S rRNA gene sequence analysis has demonstrated high resolving power for measuring the degree of relatedness between organisms above the species level (Stackebrandt & Goebel, 1994). It has been observed that organisms with total genomic relatedness above 70 % (assessed by DNA–DNA hybridization) share more than 97 % 16S rRNA gene sequence similarity (Stackebrandt & Goebel, 1994). In contrast to DNA–DNA hybridization, however, 16S rRNA gene sequence analysis often lacks resolving power at and below the species level; several studies have reported bacteria that represent different species with identical or

Abbreviations: ANI, average nucleotide identity; BT, bootstrap; ILD, incongruence-length difference; ML, maximum-likelihood; MLSA, multilocus sequence analysis; MP, maximum-parsimony; NJ, neighbour-joining.

The GenBank/EMBL/DDBJ accession numbers of newly reported sequences are provided in Table 1.

Details of primers and PCR cycling conditions, scatter plots of genetic similarity, various parameters for some of the sequences analysed and results of ILD tests are available as supplementary material with the online version of this paper.

nearly identical 16S rRNA gene sequences (Amann *et al.*, 1992; Fox *et al.*, 1992; Jaspers & Overmann, 2004; Sullivan *et al.*, 1996). Therefore, an absolute minimal 16S rRNA gene sequence similarity value for the delineation of species cannot be set (Goodfellow *et al.*, 1997). A further potential problem for identification purposes is 16S rRNA gene sequence heterogeneity due to the occurrence of multiple *rrn* operons within single genomes (Acinas *et al.*, 2004).

As more whole-genome sequences become available, various new opportunities to study the genetic relatedness of bacterial strains may be exploited. Coenye *et al.* (2005) described several novel approaches, e.g. comparison of gene order, gene content, nucleotide composition and codon usage, to assess bacterial relationships based on whole-genome sequences. Konstantinidis & Tiedje (2005) defined the average nucleotide identity (ANI) as the percentage of the total genomic sequence shared between two strains. The ANI was proven to be a robust and sensitive tool for measurement of the genetic relatedness between allied bacterial strains (from strain to genus level and possibly family level) (Konstantinidis & Tiedje, 2005; Konstantinidis *et al.*, 2006). Notwithstanding the fact that whole-genome sequencing projects are delivering new sequences at a rapidly increasing pace, the limited availability of whole-genome sequences of related strains and taxonomic reference strains currently restricts the use of whole-genome-based approaches for broad-spectrum identification and phylogenetic purposes. Therefore, reliable alternatives, which do not require full genome sequences, for the assessment of bacterial relationships are needed. For example, Cho & Tiedje (2001) developed a method based on random genome fragments and DNA microarray technology that can be applied to the identification of bacteria as well as the determination of the genetic distance between bacteria. This alternative DNA–DNA hybridization technique provides species- to strain-level resolution and avoids laborious cross-hybridizations.

Recently, the analysis of multiple protein-encoding housekeeping genes has become a widely applied tool for the investigation of taxonomic relationships (Adekambi & Drancourt, 2004; Christensen *et al.*, 2004; Holmes *et al.*, 2004; Naser *et al.*, 2005; Thompson *et al.*, 2005; Wertz *et al.*, 2003). The use of information from the comparison and combination of multiple genes can give a global and reliable overview of interorganismal relationships. The ad hoc committee for re-evaluation of the species definition regarded the sequencing of a minimum of five well-chosen housekeeping genes, universally distributed, present as single copies and located at distinct chromosomal loci, as a method of great promise for prokaryotic systematics (Stackebrandt *et al.*, 2002). In comparison with 16S rRNA genes, the higher degree of sequence divergence of housekeeping genes is superior for identification purposes, since the more-conserved rRNA gene sequences do not always allow species discrimination. Zeigler (2003) stated that a small number of carefully selected gene sequences could equal, or perhaps even surpass, the precision of DNA–DNA hybridization for quantification of genome relatedness. In contrast to DNA–DNA hybridization and 16S rRNA gene sequence analysis, multilocus sequence analysis (MLSA) is capable of yielding sequence clusters at a wide range of taxonomic levels, from intraspecific through the species level to clusters at higher levels (Gevers *et al.*, 2005). However, in order to validate the MLSA approach, the ad hoc committee for re-evaluation of the species definition called for comparative studies with organisms for which DNA–DNA reassociation data are available and the intraspecific diversity has been evaluated by DNA profiling methods (Stackebrandt *et al.*, 2002).

In a previous study, we evaluated five housekeeping genes for their use as taxonomic and phylogenetic markers in the genus *Ensifer* (Martens *et al.*, 2007). The genus *Ensifer*, comprising the former *Sinorhizobium* species and *Ensifer adhaerens* (Young, 2003), belongs to the *Alphaproteobacteria* and contains bacteria capable of nitrogen fixation in symbiosis with leguminous plants. Since *Ensifer* and *Sinorhizobium* represent synonymous genera (Martens *et al.*, 2007; Willems *et al.*, 2003; Young, 2003) and, as the oldest genus name, *Ensifer* has priority, we apply the *Ensifer* nomenclature according to Young (2003) for most *Sinorhizobium* species. A Request for an Opinion to grant priority to *Sinorhizobium* (Willems *et al.*, 2003) was denied by the Judicial Commission. Transfer of *Sinorhizobium morelense* to the genus *Ensifer* is not yet possible since this species is the subject of a pending Request for an Opinion (Euzéby & Tindall, 2004) and we therefore refer to this species here as '*S. morelense*'. Also, *Sinorhizobium americanum* has not been transferred to the genus *Ensifer* because this species was not described at the time of the request of Young (2003). Our data confirmed that MLSA of housekeeping genes is superior to 16S rRNA gene sequence analysis for *Ensifer* species discrimination (Martens *et al.*, 2007). Here, five additional housekeeping genes, *rpoB* (RNA polymerase, beta subunit), *atpD* (ATP synthase F1, beta subunit), *gap* (glyceraldehyde-3-phosphate dehydrogenase), *pnp* (polyribonucleotide nucleotidyltransferase) and *gyrB* (DNA gyrase B subunit), as well as the 23S rRNA gene, were examined. The phylogeny of the different genes was determined and results from the previous study were integrated in a large MLSA study. MLSA data were compared with DNA–DNA hybridization values and rRNA gene sequence data and the potential of MLSA for systematics and classification of strains was evaluated for the genus *Ensifer*.

METHODS

Strains and culture conditions. A total of 34 *Ensifer* strains were used in this study (Table 1): 27 strains representing all known former *Sinorhizobium* species (except for *Ensifer kummerowiae*, for which we could not obtain a bona fide strain) and seven strains representing the three different genomovars (A, B and C) of *E. adhaerens*. In addition, 14 rhizobial strains were included as reference strains. These additional strains represent the genera *Bradyrhizobium*, *Allorhizobium*, *Rhizobium*, *Mesorhizobium* and *Agrobacterium*. All

Table 1. Strains used and EMBL/GenBank/DDBJ accession numbers

Accession numbers of new data are listed in bold.

Strain	Other number	23S rRNA gene	*rpoB*	*pnp*	*atpD*	*gap*	*gyrB*	Source
Sinorhizobium americanum LMG 22684^T	CFNEI 156^T	**AM418707**	**AM295361**	**AM295474**	**AM418741**	**AM295395**	**AM418790**	*Acacia acatlensis*, Sierra de Huautla, Mexico
Ensifer arboris								
LMG 14919^T	HAMBI 1552^T	**AM418716**	**AM295380**	**AM295482**	**AM418767**	**AM295420**	**AM418815**	*Prosopis chilensis*, Kosti, Sudan
LMG 14917	HAMBI 1396	**AM418739**	**AM295381**	**AM295483**	**AM418776**	**AM295429**	**AM418824**	*Prosopis chilensis*, Kenya
LMG 19223	HAMBI 1704	**AM418738**	**AM295379**	**AM295481**	**AM418757**	**AM295410**	**AM418805**	*Acacia senegal*, Khartoum, Sudan
Ensifer fredii								
LMG 6217^T	USDA 205^T	AY244360	**AM295358**	**AM295458**	**AM418761**	**AM295414**	**AM418809**	*Glycine max*, Honan, China
LMG 8317	USDA 191	**AM418727**	**AM295357**	**AM295457**	**AM418758**	**AM295411**	**AM418806**	Soil, Shanghai, China
LMG 6218	USDA 206	**AM418729**	**AM295359**	**AM295459**	**AM418762**	**AM295415**	**AM418810**	*Glycine max*, Honan, China
LMG 6216	LMG 8317	**AM418735**	**AM295360**	**AM295460**	**AM418773**	**AM295426**	**AM418821**	Soil, Shanghai, China
Ensifer kostiensis								
LMG 19227^T	HAMBI 1489^T	**AM418717**	**AM295369**	**AM295479**	**AM418771**	**AM295424**	**AM418819**	*Acacia senegal*, Kosti, Sudan
LMG 14911	HAMBI 1502	**AM418736**	**AM295370**	**AM295480**	**AM418774**	**AM295427**	**AM418822**	*Acacia senegal*, Tendelti, Sudan
LMG 19225	HAMBI 1484	**AM418734**	**AM295368**	**AM295478**	**AM418770**	**AM295423**	**AM418818**	*Prosopis chilensis*, Kosti, Sudan
Ensifer medicae								
LMG 19920^T	R-916	**AM418714**	**AM295387**	**AM295484**	**AM418754**	**AM295407**	**AM418802**	*Medicago truncatula*, Aude, France
LMG 19921	R-481	**AM418741**	**AM295386**	**AM295486**	**AM418778**	**AM295431**	**AM490194**	*Medicago truncatula*
LMG 18864	HAMBI 1809	**AM418733**	**AM295385**	**AM295485**	**AM418769**	**AM295422**	**AM418817**	*Medicago truncatula*, Syria
Ensifer meliloti								
LMG 6133^T	NZP 4027^T	AF207786	**AM295384**	**AM295472**	**AM418760**	**AM295413**	**AM418808**	*Medicago sativa*, Virginia, USA
LMG 4289	INRA 2011	**AM418737**	**AM295383**	**AM295473**	**AM418775**	**AM295428**	**AM418823**	*Medicago sativa*
LMG 6130	NZP 4009	**AM418728**	**AM295382**	**AM295471**	**AM418759**	**AM295412**	**AM418807**	*Medicago sativa*, Australia
'*Sinorhizobium morelense*' LMG 21331^T	CFN E1007^T	**AM418715**	**AM295377**	**AM295454**	**AM418755**	**AM295408**	**AM418803**	*Leucaena leucocephala* cv. Cunningham in cultivated soils
Ensifer sp. LMG 20571	R-4955	**AM418718**	**AM295378**	**AM295461**	**AM418772**	**AM295425**	**AM418820**	Agricultural soil, Pittem, Belgium
Ensifer saheli								
LMG 7837^T	ORS 609^T	AY244368	**AM295362**	**AM295475**	**AM418756**	**AM295409**	**AM418804**	*Sesbania cannabina*, Dakar, Senegal
LMG 11864	ORS 600	**AM418740**	**AM295365**	**AM295477**	**AM418777**	**AM295430**	**AM418825**	*Sesbania pachycarpa*, Senegal
LMG 8310	ORS 611	**AM418731**	**AM295363**	**AM295476**	**AM418765**	**AM295418**	**AM418813**	*Sesbania grandiflora*, Dakar, Senegal
Ensifer terangae								
LMG 7834^T	ORS 1009^T	AY244369	**AM295366**	**AM295469**	**AM418764**	**AM295417**	**AM418812**	*Acacia laeta*, Dakar, Senegal
LMG 11859	ORS 52	**AM418732**	**AM295367**	**AM295470**	**AM418766**	**AM295419**	**AM418814**	*Sesbania rostrata*, Senegal
LMG 6464	ORS 51	**AM418730**	**AM295364**	**AM295468**	**AM418763**	**AM295416**	**AM418811**	*Sesbania rostrata*, Dakar, Senegal
Ensifer xinjiangensis								
LMG 17930^T	CCBAU 110^T	**AM418708**	**AM295355**	**AM295455**	**AM418745**	**AM295398**	**AM418793**	*Glycine max*, Xinjiang, China
R-16438	CCBAU 83834	**AM418726**	**AM295356**	**AM295456**	**AM418751**	**AM295404**	**AM418799**	Not known
Ensifer adhaerens gv. C								
LMG 20216^T	ATCC 33212^T	**AM418709**	**AM295374**	**AM295452**	**AM418746**	**AM295399**	**AM418794**	Soil, central Pennsylvania, USA

Table 1. cont.

Strain	Other number	23S rRNA gene	*rpoB*	*pnp*	*atpD*	*gap*	*gyrB*	Source
R-14067	ATCC 33499	**AM418722**	**AM295375**	**AM295450**	**AM418743**	**AM295396**	**AM418791**	Soil, central Pennsylvania, USA
LMG 20582	R-6387	**AM418723**	**AM295376**	**AM295451**	**AM418744**	**AM295397**	**AM418792**	Agricultural soil, Pittem, Belgium
E. adhaerens gv. A								
LMG 9954	BR819	**AM418725**	**AM295373**	**AM295466**	**AM418750**	**AM295403**	**AM418798**	*Leucaena leucocephala*, Brazil
LMG 10007	BR8606	**AM418711**	**AM295372**	**AM295465**	**AM418749**	**AM295402**	**AM418797**	*Pithecellobium dulce*, Brazil
R-9451	HAMBI 1631	**AM418724**	**AM295371**	**AM295467**	**AM418748**	**AM295401**	**AM418796**	*Sesbania grandiflora*, Sri Lanka
E. adhaerens gv. B R-7457	5D19	**AM418710**	**AM295347**	**AM295453**	**AM418747**	**AM295400**	**AM418795**	*Medicago sativa*, Spain
Mesorhizobium mediterraneum LMG 17148^T	UPM-Ca36^T	AY244363	**AM295350**	**AM295488**	**AM418768**	**AM295421**	**AM418816**	*Cicer arietinum* L., Spain
Bradyrhizobium elkanii LMG 6134^T	USDA 76^T	**AM418712**	**AM295348**	**AM295489**	**AM418752**	**AM295405**	**AM418800**	*Glycine max*, USA
Bradyrhizobium japonicum LMG 6138^T	NZP 5549^T	**AM418713**	**AM295349**	**AM295490**	**AM418753**	**AM295406**	**AM418801**	*Glycine max*, Japan
Rhizobium galegae LMG 6214^T	HAMBI 540^T	AF207783	**AM295389**	**AM295447**	**AM418779**	**AM295432**	**AM418826**	*Galegae orientalis*, Finland
Rhizobium giardinii R-4385^T	H152^T	**AM418719**	**AM295388**	**AM295448**	**AM418780**	**AM295433**	**AM418827**	*Phaseolus vulgaris*, France
Rhizobium gallicum R-4387^T	R602sp^T	AY244362	**AM295351**	**AM295463**	**AM418781**	**AM295434**	**AM418828**	*Phaseolus vulgaris*, France
Rhizobium huautlense LMG 18254^T	Wang S02^T	AY244375	**AM295390**	**AM295462**	**AM418782**	**AM295435**	**AM418829**	*Sesbania herbacea*, Sierra de Huautla, Mexico
Rhizobium leguminosarum LMG 14904^T	ATCC 10004^T	AY244361	**AM295352**	**AM295445**	**AM418783**	**AM295436**	**AM418830**	*Pisum sativum*, Illinois, USA
Allorhizobium undicola LMG 11875^T	ORS 992^T	**AM418720**	**AM295391**	**AM295464**	**AM418784**	**AM295437**	**AM418831**	*Neptunia natans*, Kaolack, Senegal
Agrobacterium radiobacter LMG 140^T	ATCC 19358^T	**AM418721**	**AM295393**	**AM295443**	**AM418785**	**AM295438**	**AM418832**	Not known
Rhizobium rhizogenes LMG 150^T	ATCC 11325^T	AF208480	**AM295353**	**AM295449**	**AM418786**	**AM295439**	**AM418833**	Apple
Agrobacterium rubi LMG 17935^T	ATCC 13335^T	AY244376	**AM295394**	**AM295444**	**AM418787**	**AM295440**	**AM418834**	*Rubus ursinis* var. *loganobaccus*, USA
Agrobacterium vitis LMG 8750^T	NCPPB 3554^T	AF209071	**AM295392**	**AM295487**	**AM418788**	**AM295441**	**AM418835**	*Vitis vinifera*, Australia
Rhizobium tropici LMG 9503^T	CIAT 899^T	AF208479	**AM295354**	**AM295446**	**AM418789**	**AM295442**	**AM418836**	*Phaseolus vulgaris*, Colombia

bacterial strains were grown on yeast extract mannitol agar (YMA) at 28 °C.

DNA preparation. Bacterial DNA was prepared using the alkaline lysis method as described by Baele *et al.* (2000).

Primers for amplification and sequencing. The following genes were studied: *rpoB* (RNA polymerase, beta subunit), *atpD* (ATP synthase F1, beta subunit), the 23S rRNA gene, *gap* (glyceraldehyde-3-phosphate dehydrogenase), *pnp* (polyribonucleotide nucleotidyltransferase) and *gyrB* (DNA gyrase B subunit). Primers for the amplification of the 23S rRNA gene were obtained from Van Camp *et al.* (1993). To design primers for PCR amplification and sequencing of the housekeeping genes, we used the corresponding sequences derived from the whole-genome sequences of related bacteria: *Agrobacterium tumefaciens* C58 (Goodner *et al.*, 2001; Wood *et al.*, 2001), *Ensifer meliloti* 1021 (Galibert *et al.*, 2001), *Mesorhizobium huakuii* MAFF 303099 (Kaneko *et al.*, 2000; Turner *et al.*, 2002), *Brucella melitensis* 16M^T (DelVecchio *et al.*, 2002) and *Bradyrhizobium japonicum* USDA 110 (Kaneko *et al.*, 2002). The gene sequences were compared using the BioNumerics 4.6 software

package (Applied Maths) in order to identify conserved regions for the development of suitable primers. The primers used are listed in Supplementary Table S1 (available in IJSEM Online).

PCR amplification and sequencing of the genes. PCR amplification was performed as described previously (Martens *et al.*, 2007). The cycling conditions are listed in Supplementary Table S1. The presence of PCR products and their concentration were verified by electrophoresis of 3 µl product on a 1 % agarose gel and staining with ethidium bromide. A molecular size marker (Smartladder-Eurogentec) was included to estimate the length of the amplification products.

The amplified products were purified using a Qiaquick PCR purification kit (Qiagen). The purified DNA was sequenced using the dideoxynucleotide chain-termination method with fluorescent ddNTPs (Applied Biosystems) on an ABI Prism 3100 capillary sequencer according to manufacturer's instructions (Applied Biosystems). Consensus sequences were constructed using the AutoAssembler software (Applied Biosystems). Accession numbers of new sequence data are listed in bold in Table 1.

Sequence data analyses. The TaxonGap software tool (Naser *et al.*, 2007) was applied to represent the resolution of the different genes within and between taxonomic units. For each gene and each species/genomovar, the amount of heterogeneity (sequence divergence within a species/genomovar) and the amount of separability (smallest amount of sequence divergence observed between a particular species/genomovar and the other species/genomovars; the species displaying the smallest amount of sequence divergence from the particular species is referred to as the closest neighbour taxon) were calculated. Distances used for the calculation of heterogeneity and separability values were determined using pairwise sequence alignments by the Needleman–Wunsch algorithm as implemented in BioNumerics 4.6.

Nucleotide sequence alignments were made using CLUSTAL_X (Thompson *et al.*, 1997) and RevTrans 1.4 (Wernersson & Pedersen, 2003), taking into account the corresponding amino acid alignments for protein-encoding genes. To assess the influence of noise due to saturation of the third codon position, we performed incongruence-length difference (ILD) tests (Farris *et al.*, 1995) as implemented in PAUP* version 4.0b10 (Swofford, 2002), using the different codon positions as separate partitions in 1250 replications. The same set of strains was used for all genes and sequence data for *Caulobacter crescentus* CB15, extracted from the complete genome sequence (Nierman *et al.*, 2001), were used as an outgroup. Neighbour-joining (NJ), maximum-parsimony (MP) and maximum-likelihood (ML) analyses were performed with PAUP*. NJ analyses were performed using the Kimura-2 correction and 1000 bootstrap (BT) replications; MP analyses were performed using the heuristic search option. For ML analyses, the optimal models of nucleotide substitution were estimated using the program MODELTEST 3.7 (Posada & Crandall, 1998) using both hierarchical likelihood ratio tests (hLRTs) and the Akaike information criterion (AIC) (Supplementary Table S2). When these options did not yield the same model, which was the case for the *rpoB*, *gap*, *pnp* and 23S rRNA genes, trees were constructed and compared using the different models. Since only negligible differences in tree topology and BT values were observed, only the trees constructed with the AIC model were used (Posada & Buckley, 2004). The MP trees were used as starting trees for the heuristic search procedure. BT analyses were performed using 1000 replications of heuristic searches for MP and 100 replications for ML. The ILD test implemented in PAUP* and using 1250 replicates was used to assess incongruence between datasets for the different genes.

DNA–DNA hybridization. DNA–DNA hybridizations were performed with *Ensifer fredii* strains LMG 6217^T and LMG 8317 and *Ensifer xinjiangensis* strains LMG 17930^T, R-16438 (=CCBAU 83834) and R-16439 (=CCBAU 83827). DNA was prepared as described by Willems *et al.* (2001) applying a slightly modified procedure of Marmur (1961). Hybridizations were carried out using a microplate method and biotinylated probe DNA (Ezaki *et al.*, 1989). Hybridizations were performed at 45 °C in 2× SSC in the presence of 50 % formamide (Willems *et al.*, 2001).

Correlation of DNA–DNA hybridization values with MLSA data. Similarity plots (scatter plots) between DNA–DNA hybridization values and sequence similarity values were constructed in BioNumerics 4.6. Correlation between values was calculated using Pearson's product-moment correlation coefficient (Supplementary Fig. S1).

RESULTS AND DISCUSSION

In the present study, nucleotide sequences of the *rpoB* (RNA polymerase, beta subunit), *atpD* (ATP synthase F1, beta subunit), *gap* (glyceraldehyde-3-phosphate dehydrogenase), *pnp* (polyribonucleotide nucleotidyltransferase) and *gyrB* (DNA gyrase B subunit) housekeeping genes and the 23S rRNA gene were determined for 34 *Ensifer* strains and 14 other rhizobial strains (Table 1). The genes selected are widely distributed, unique within the genome, of adequate length to be phylogenetically informative, located separately on the main chromosome (as assessed from the *E. meliloti* complete genome) and have a relatively high degree of conservation (as established from the literature) (Zeigler, 2003). Except for *pnp*, the housekeeping genes analysed in this study were reported previously as good taxonomic markers (Mollet *et al.*, 1997; Rönner *et al.*, 1991; Wertz *et al.*, 2003; Yamamoto & Harayama, 1995). Amplification was successful for all strains. The length of the amplified fragments was 474–489 bp for *atpD*, 798–804 bp for *gap*, 666–699 bp for *gyrB*, 954 bp for *rpoB*, 540 bp for *pnp* and 1884–1897 bp for the 23S rRNA gene. For *Ensifer medicae* LMG 19921, a highly divergent *gyrB* sequence (1169 bp; 57.7 and 59.2 % sequence similarity with *E. medicae* strains LMG 19920 and LMG 18864, respectively) was obtained which severely complicated the alignment. Query of the translated amino acid sequence via BLAST (Altschul *et al.*, 1997) against the NCBI bacterial database revealed 76 % similarity (53 % identity) with the GyrB sequence of an alphaproteobacterial *Sphingopyxis alaskensis* strain as the closest match. Horizontal transfer and subsequent recombination could be a possible explanation for this aberrant *gyrB* sequence. Another aberrant result was found for the 23S rRNA gene from *Ensifer arboris* LMG 14919^T, which contained a 98 bp insert near the 5′ end (total sequence length 1983 bp). Previously, Selenska-Pobell & Evguenieva-Hackenberg (1995) reported the presence of a highly variable 130 bp insert near the 5′ end of the 23S rRNA gene in some members of the *Rhizobiaceae*. However, the position and sequence of the insert were different from those of the 98 bp insert found in LMG 14919^T.

For analyses of the sequences, we also included the corresponding sequences retrieved from the complete

genome sequences of *Agrobacterium tumefaciens* C58 (Goodner *et al.*, 2001; Wood *et al.*, 2001), *Brucella melitensis* 16M^T (DelVecchio *et al.*, 2002), *Brucella suis* 1330^T (Paulsen *et al.*, 2002), *Caulobacter crescentus* CB15 (Nierman *et al.*, 2001), *Ensifer meliloti* 1021 (Galibert *et al.*, 2001), *Rhodopseudomonas palustris* CGA009 (Larimer *et al.*, 2004), *Mesorhizobium huakuii* MAFF 303099 (Kaneko *et al.*, 2000; Turner *et al.*, 2002) and *Bradyrhizobium japonicum* USDA 110 (Kaneko *et al.*, 2002). The lengths of the alignments used for individual gene analyses are listed in Supplementary Table S2. The alignments for *atpD*, *gap*, *gyrB* and the 23S rRNA gene contained gaps, whereas no gaps were present for *rpoB* or *pnp*. For the 23S rRNA gene, due to the intrinsically uncertain alignment because of low sequence similarity and length variations, a continuous region of 22 bases (positions 970–991 of the multiple alignment) was omitted from the analyses.

Individual gene analyses

The potential of the different genes to identify the *Ensifer* species/genomovars was assessed. Suitable molecular markers for identification purposes exhibit the smallest amount of heterogeneity within a species/genomovar and result in maximal separation between the different species/genomovars. All three codon positions were included in the individual gene analyses since no significant codon saturation was observed for the different codon positions (data not shown). For the calculation of the heterogeneity and separability values, *E. xinjiangensis* (two strains) and *E. fredii* (four strains) were considered as synonymous species (see below) since their gene sequences were identical (23S rRNA gene, *atpD*, *pnp* and *rpoB*) or very similar (98.4–100 % for *gap* and 97.2–100 % for *gyrB*) in all comparisons. However, a DNA–DNA hybridization value of 39 % was reported (Peng *et al.*, 2002) between the two type strains. We repeated these hybridizations and included two additional *E. xinjiangensis* strains, R-16438 and R-16439, and found hybridization values of 78–85 %. With a second *E. fredii* strain, LMG 8317, values were 74–89 %, thus establishing that *Ensifer xinjiangensis* is a later heterotypic synonym of *Ensifer fredii*. The close relationship and probable synonymy of *E. xinjiangensis* and *E. fredii* was reported previously based on sequence analyses of the 16S rRNA gene (Tan *et al.*, 1997), the internally transcribed spacer region (Kwon *et al.*, 2005) and housekeeping genes (Martens *et al.*, 2007).

Sequence data from this and our previous study (Martens *et al.*, 2007) were combined, and the heterogeneity and separability values were calculated for the 10 housekeeping genes (*atpD*, *dnaK*, *gap*, *glnA*, *gltA*, *gyrB*, *recA*, *rpoB*, *pnp*, *thrC*) and the 16S and 23S rRNA genes using TaxonGap (Naser *et al.*, 2007). Results are summarized in Fig. 1. For each *Ensifer* species (or genomovar), sequences of the same gene for the different strains included were highly similar and, as a consequence, heterogeneity values (indicated by the light-grey bars in Fig. 1) were low (sequence divergence ranged from 0–0.8 % for *atpD* to 0–2.8 % for *gyrB*; for the 16S and 23S rRNA genes, values were respectively 0–0.4 % and 0–0.8 %). This is mainly due to genuine low intraspecies/intragenomovar sequence variability, but also partially to inclusion of relatively few strains (two to four) per species/genomovar. *Ensifer* species displaying some intraspecies heterogeneity are *E. arboris*, *E. adhaerens* gv. A and C, *E. fredii*, *E. meliloti* and *Ensifer saheli*. Sequence divergence between *Ensifer* species for the housekeeping genes was clearly higher, ranging from 3.1–12.5 % for *glnA* to 5.8–20.5 % for *thrC*, which is reflected in the high separability values (indicated by the dark-grey bars in Fig. 1). In contrast, for the 16S and 23S rRNA genes, sequence divergence between species was only 0.2–2.1 % and 0.4–3.9 %, respectively. For each species/genomovar, there is a clear gap between the heterogeneity and separability values for each of the housekeeping genes. This implies that the *Ensifer* species and genomovars form distinct groups, well separated from each other, for all housekeeping gene sequences analysed. The housekeeping genes with the best capability to identify *Ensifer* strains, due to high separability and low heterogeneity values, are *gyrB*, *gltA*, *recA* and *thrC*. The rRNA genes, however, exhibit no or very little separability between species. As a consequence, rRNA gene sequence analysis does not always allow species identification: for example, 16S rRNA gene sequencing does not allow discrimination of *E. fredii* (including *E. xinjiangensis*) and *S. americanum* and 23S rRNA gene sequences can not separate *Ensifer* sp. LMG 20571 and '*S. morelense*'. Although the 23S rRNA gene contains more phylogenetic information than the 16S rRNA gene (Woese, 1987), housekeeping genes are more discriminatory and thus superior for the identification of strains from closely related lineages. The same conclusion could be drawn when determining the number of parsimony-informative sites to estimate of the amount of phylogenetic information contained in each gene (see Supplementary Table S2).

Initial NJ trees of the individual genes including all strains revealed a tight clustering of strains within each species/genomovar (data not shown), in line with low heterogeneity values computed by TaxonGap (Fig. 1). Therefore, only the type strains (except for *E. adhaerens*, where the three genomovars were included) were selected for further individual gene analyses in order to reduce computing time. The ML trees constructed from the individual *atpD*, *gap*, *gyrB*, *pnp*, *rpoB* and 23S rRNA gene alignments are shown in Fig. 2 (for the remaining gene trees, see Martens *et al.*, 2007). Considerable variation in tree topology was observed for all of the separate genes. In the *rpoB* and *gyrB* analyses, the different *Ensifer* species formed a single and significant cluster (BT values of 81 and 89 %, respectively). A close phylogenetic relationship was observed between *E. meliloti*, *E. medicae* and *E. arboris* for these genes (see also closest-neighbour analysis in Fig. 1). For *rpoB*, a tight clustering was also demonstrated for the species *E. fredii* (and the synonymous species *E. xinjiangensis*) and *S.*

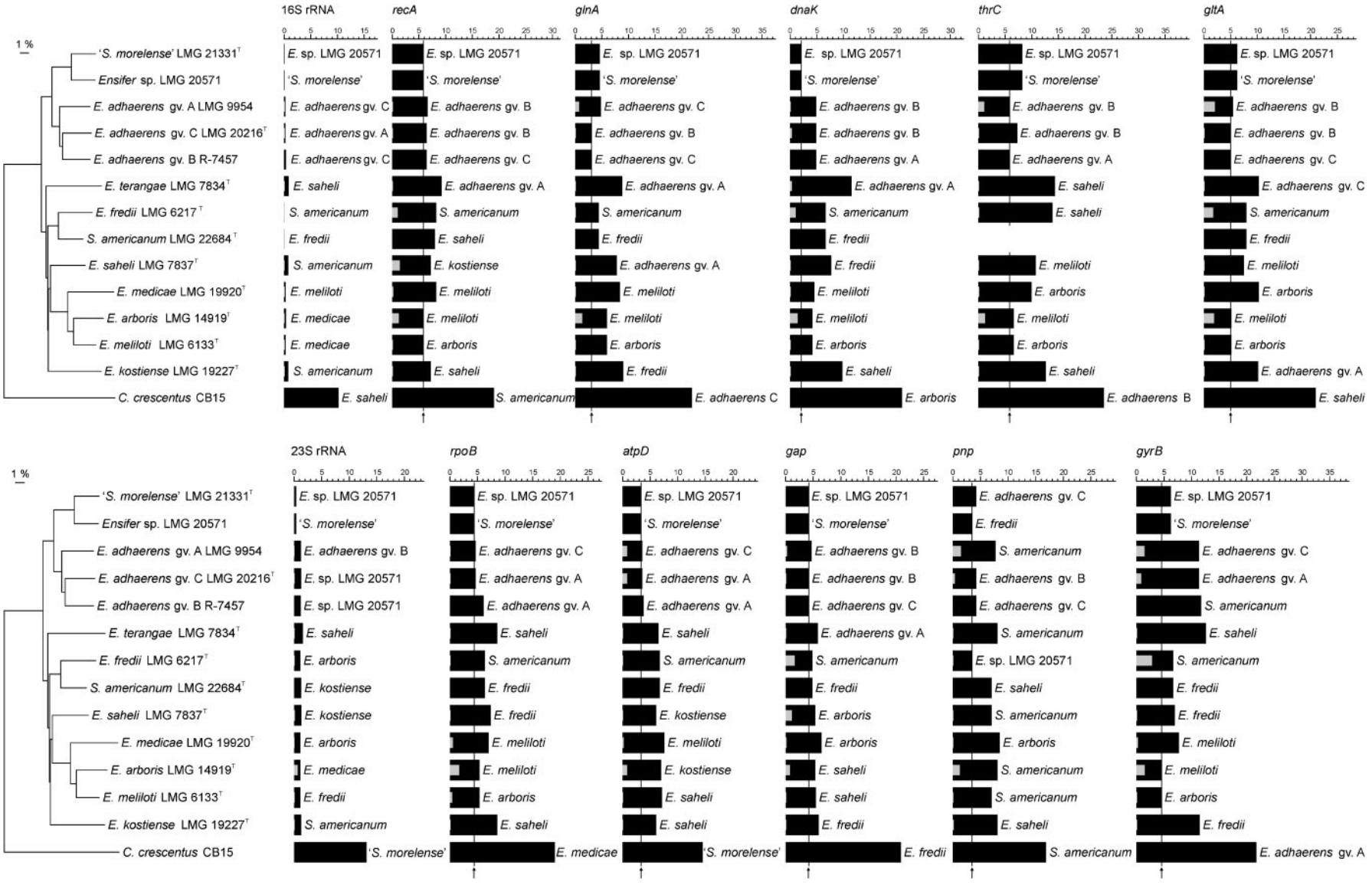

Fig. 1. Matrix of heterogeneity (light-grey bars) and separability (dark-grey bars) values with the different species/genomovars as matrix rows and the different genes as matrix columns, as calculated by TaxonGap. Species/genomovars were ordered according to their phylogenetic position in an ML tree calculated from the concatenated sequences of their type strains (when available) or another representative strain (see tree on the left). For each species/genomovar and each gene, the closest neighbour (i.e. the taxon with the smallest separability for this species and this gene) is listed to the right of the dark-grey bar. For each gene, the vertical line denotes the smallest separability recorded. Bars in the ML trees, 1 % estimated substitutions.

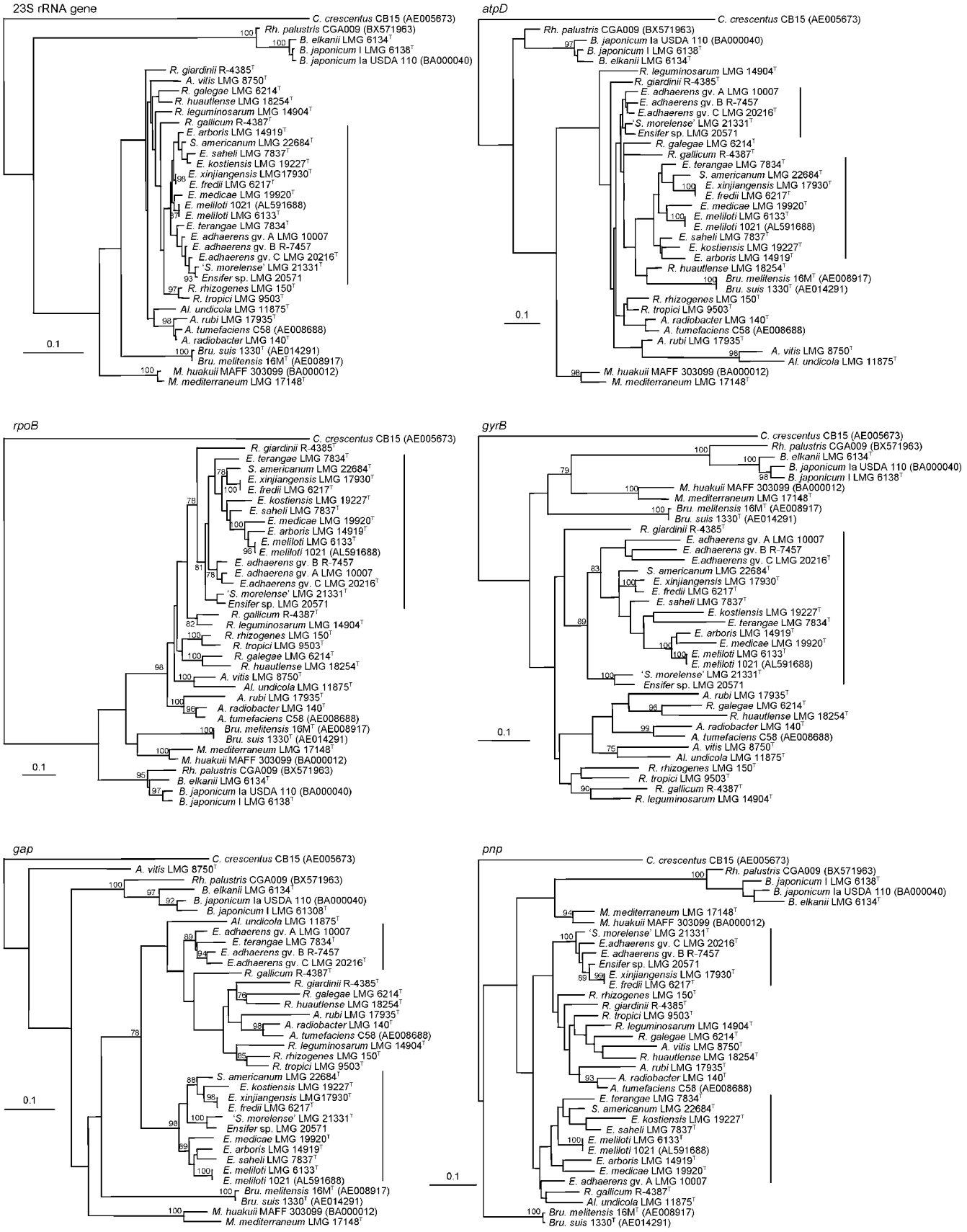
23S rRNA gene
C. crescentus CB15 (AE005673)
Rh. palustris CGA009 (BX571963)
B. elkanii LMG 6134T
B. japonicum I LMG 6138T
B. japonicum Ia USDA 110 (BA000040)
R. giardinii R-4385T
A. vitis LMG 8750T
R. galegae LMG 6214T
R. huautlense LMG 18254T
R. leguminosarum LMG 14904T
R. gallicum R-4387T
E. arboris LMG 14919T
S. americanum LMG 22684T
E. saheli LMG 7837T
E. kostiensis LMG 19227T
E. xinjiangensis LMG17930T
E. fredii LMG 6217T
E. medicae LMG 19920T
E. meliloti 1021 (AL591688)
E. meliloti LMG 6133T
E. terangae LMG 7834T
E. adhaerens gv. A LMG 10007
E. adhaerens gv. B R-7457
E.adhaerens gv. C LMG 20216T
'S. morelense' LMG 21331T
Ensifer sp. LMG 20571
R. rhizogenes LMG 150T
R. tropici LMG 9503T
Al. undicola LMG 11875T
A. rubi LMG 17935T
A. tumefaciens C58 (AE008688)
A. radiobacter LMG 140T
Bru. suis 1330T (AE014291)
Bru. melitensis 16MT (AE008917)
M. huakuii MAFF 303099 (BA000012)
M. mediterraneum LMG 17148T
0.1
atpD
C. crescentus CB15 (AE005673)
Rh. palustris CGA009 (BX571963)
B. japonicum Ia USDA 110 (BA000040)
B. japonicum I LMG 6138T
B. elkanii LMG 6134T
R. leguminosarum LMG 14904T
R. giardinii R-4385T
E. adhaerens gv. A LMG 10007
E. adhaerens gv. B R-7457
E.adhaerens gv. C LMG 20216T
'S. morelense' LMG 21331T
Ensifer sp. LMG 20571
R. galegae LMG 6214T
R. gallicum R-4387T
E. terangae LMG 7834T
S. americanum LMG 22684T
E. xinjiangensis LMG 17930T
E. fredii LMG 6217T
E. medicae LMG 19920T
E. meliloti LMG 6133T
E. meliloti 1021 (AL591688)
E. saheli LMG 7837T
E. kostiensis LMG 19227T
E. arboris LMG 14919T
R. huautlense LMG 18254T
Bru. melitensis 16MT (AE008917)
Bru. suis 1330T (AE014291)
R. rhizogenes LMG 150T
R. tropici LMG 9503T
A. radiobacter LMG 140T
A. tumefaciens C58 (AE008688)
A. rubi LMG 17935T
A. vitis LMG 8750T
Al. undicola LMG 11875T
M. huakuii MAFF 303099 (BA000012)
M. mediterraneum LMG 17148T
0.1
rpoB
C. crescentus CB15 (AE005673)
R. giardinii R-4385T
E. terangae LMG 7834T
S. americanum LMG 22684T
E. xinjiangensis LMG 17930T
E. fredii LMG 6217T
E. kostiensis LMG 19227T
E. saheli LMG 7837T
E. medicae LMG 19920T
E. arboris LMG 14919T
E. meliloti LMG 6133T
E. meliloti 1021 (AL591688)
E. adhaerens gv. B R-7457
E. adhaerens gv. A LMG 10007
E.adhaerens gv. C LMG 20216T
'S. morelense' LMG 21331T
Ensifer sp. LMG 20571
R. gallicum R-4387T
R. leguminosarum LMG 14904T
R. rhizogenes LMG 150T
R. tropici LMG 9503T
R. galegae LMG 6214T
R. huautlense LMG 18254T
A. vitis LMG 8750T
Al. undicola LMG 11875T
A. rubi LMG 17935T
A. radiobacter LMG 140T
A. tumefaciens C58 (AE008688)
Bru. melitensis 16MT (AE008917)
Bru. suis 1330T (AE014291)
M. mediterraneum LMG 17148T
M. huakuii MAFF 303099 (BA000012)
Rh. palustris CGA009 (BX571963)
B. elkanii LMG 6134T
B. japonicum Ia USDA 110 (BA000040)
B. japonicum I LMG 6138T
0.1
gyrB
C. crescentus CB15 (AE005673)
Rh. palustris CGA009 (BX571963)
B. elkanii LMG 6134T
B. japonicum Ia USDA 110 (BA000040)
B. japonicum I LMG 6138T
M. huakuii MAFF 303099 (BA000012)
M. mediterraneum LMG 17148T
Bru. melitensis 16MT (AE008917)
Bru. suis 1330T (AE014291)
R. giardinii R-4385T
E. adhaerens gv. A LMG 10007
E. adhaerens gv. B R-7457
E.adhaerens gv. C LMG 20216T
S. americanum LMG 22684T
E. xinjiangensis LMG 17930T
E. fredii LMG 6217T
E. saheli LMG 7837T
E. kostiensis LMG 19227T
E. terangae LMG 7834T
E. arboris LMG 14919T
E. medicae LMG 19920T
E. meliloti LMG 6133T
E. meliloti 1021 (AL591688)
'S. morelense' LMG 21331T
Ensifer sp. LMG 20571
A. rubi LMG 17935T
R. galegae LMG 6214T
R. huautlense LMG 18254T
A. radiobacter LMG 140T
A. tumefaciens C58 (AE008688)
A. vitis LMG 8750T
Al. undicola LMG 11875T
R. rhizogenes LMG 150T
R. tropici LMG 9503T
R. gallicum R-4387T
R. leguminosarum LMG 14904T
0.1
gap
C. crescentus CB15 (AE005673)
A. vitis LMG 8750T
Rh. palustris CGA009 (BX571963)
B. elkanii LMG 6134T
B. japonicum Ia USDA 110 (BA000040)
B. japonicum I LMG 61308T
Al. undicola LMG 11875T
E. adhaerens gv. A LMG 10007
E. terangae LMG 7834T
E. adhaerens gv. B R-7457
E.adhaerens gv. C LMG 20216T
R. gallicum R-4387T
R. giardinii R-4385T
R. galegae LMG 6214T
R. huautlense LMG 18254T
A. rubi LMG 17935T
A. radiobacter LMG 140T
A. tumefaciens C58 (AE008688)
R. leguminosarum LMG 14904T
R. rhizogenes LMG 150T
R. tropici LMG 9503T
S. americanum LMG 22684T
E. kostiensis LMG 19227T
E. xinjiangensis LMG17930T
E. fredii LMG 6217T
'S. morelense' LMG 21331T
Ensifer sp. LMG 20571
E. medicae LMG 19920T
E. arboris LMG 14919T
E. saheli LMG 7837T
E. meliloti LMG 6133T
E. meliloti 1021 (AL591688)
Bru. melitensis 16MT (AE008917)
Bru. suis 1330T (AE014291)
M. huakuii MAFF 303099 (BA000012)
M. mediterraneum LMG 17148T
0.1
pnp
C. crescentus CB15 (AE005673)
Rh. palustris CGA009 (BX571963)
B. japonicum I LMG 6138T
B. japonicum Ia USDA 110 (BA000040)
B. elkanii LMG 6134T
M. mediterraneum LMG 17148T
M. huakuii MAFF 303099 (BA000012)
'S. morelense' LMG 21331T
E.adhaerens gv. C LMG 20216T
E. adhaerens gv. B R-7457
Ensifer sp. LMG 20571
E. xinjiangensis LMG 17930T
E. fredii LMG 6217T
R. rhizogenes LMG 150T
R. giardinii R-4385T
R. tropici LMG 9503T
R. leguminosarum LMG 14904T
R. galegae LMG 6214T
A. vitis LMG 8750T
R. huautlense LMG 18254T
A. rubi LMG 17935T
A. radiobacter LMG 140T
A. tumefaciens C58 (AE008688)
E. terangae LMG 7834T
S. americanum LMG 22684T
E. kostiensis LMG 19227T
E. saheli LMG 7837T
E. meliloti LMG 6133T
E. meliloti 1021 (AL591688)
E. arboris LMG 14919T
E. medicae LMG 19920T
E. adhaerens gv. A LMG 10007
R. gallicum R-4387T
Al. undicola LMG 11875T
Bru. melitensis 16MT (AE008917)
Bru. suis 1330T (AE014291)
0.1

Fig. 2. Phylogenetic reconstructions based on individual analyses of the 23S rRNA, *atpD*, *rpoB*, *gyrB*, *gap* and *pnp* genes. Analyses were conducted using the ML method. BT values of 75 or more (using 100 replicates) are indicated at branching points. Accession numbers for complete genome sequences are listed. *Ensifer* strains are marked by vertical bars. *A.*, *Agrobacterium*; *Al.*, *Allorhizobium*; *B.*, *Bradyrhizobium*; *Bru.*, *Brucella*; *R.*, *Rhizobium*; *Rh.*, *Rhodopseudomonas*. Bars, 0.1 % estimated substitutions.

americanum (see also Fig. 1). In the 23S rRNA gene analysis, the single cluster encompassing all *Ensifer* strains was also apparent, but was supported by only a very low BT value (20 %). For the *atpD*, *pnp* and *gap* genes, aberrant groupings were found compared with other single-gene tree topologies from this and the previous study (Martens *et al.*, 2007). In the *atpD* tree topology, the genus *Ensifer* was composed of two separate and poorly supported clusters: one cluster contained '*S. morelense*' and the three genomovars of *E. adhaerens* (BT value 13 %), while the second cluster contained all other *Ensifer* species (BT value 50 %). In the case of *gap*, *Ensifer terangae* and the different *E. adhaerens* genomovars clustered together with all *Rhizobium* and most *Agrobacterium* strains (BT value 55 %), while all other *Ensifer* strains formed a single, separate clade (BT value 98 %). For *pnp*, the genus *Ensifer* was composed of two clusters: while one contained *E. fredii*, *E. adhaerens* gv. B and C, *Ensifer* sp. LMG 20571 and '*S. morelense*' (BT value 100 %), the other, poorly supported cluster contained the remaining *Ensifer* strains, but also *Allorhizobium undicola* and *Rhizobium gallicum* (BT value 25 %). Thus *atpD*, *gap* and *pnp* resulted in trees in which other taxa were grouped between or within the *Ensifer* clusters. This different placement of species in individual gene tree analyses may be due to different evolutionary histories of the genes, intragenomic rearrangements or horizontal gene transfer and subsequent recombination events (Charles *et al.*, 2005; Christensen *et al.*, 2004; Rokas *et al.*, 2003). Christensen & Olsen (1998) previously reported conflicting results when comparing *atpD* with other housekeeping genes and the 23S rRNA gene in *Salmonella*.

Composite tree

The ILD test (Farris *et al.*, 1995) was applied to assess congruence between the 12 genes. *S. americanum* LMG 22684^T, *Agrobacterium vitis* LMG 8750^T and *Allorhizobium undicola* LMG 11875^T were excluded from the comparison since no sequence was obtained for their *thrC*, *glnA* and *gltA* genes, respectively (Martens *et al.*, 2007). *E. medicae* LMG 19921 was also excluded from the analyses since it exhibited an aberrant *gyrB* sequence which complicated gene sequence alignment considerably. Different levels of significant congruence were found between *dnaK*, *glnA*, *gltA*, *gyrB*, *recA*, *rpoB* and *thrC* (Supplementary Table S3). These seven congruent genes coincide with the subset of genes that produced the most consistent phylogenetic placement of species (with some minor incongruence within clades), except for the *dnaK* gene, which exhibited a phylogeny with some aberrant clustering (Martens *et al.*, 2007). The congruent gene sequence alignments were concatenated. In line with single-gene sequence characteristics (Fig. 1), heterogeneity values of the seven concatenated gene sequences were low (sequence divergence ranging from 0 to 2.7 %) within an *Ensifer* species/genomovar, while sequence divergence was clearly higher at the interspecies/intergenomovar level (ranging from 5.5 to 14.2 %). The clear gap between intra- and interspecies sequence divergence values allows reliable identification of all species and results in clear species boundaries. Konstantinidis *et al.* (2006) demonstrated from a whole-genome comparison study in which the conserved core genes of several groups of organisms were analysed that the classical cut-off of 70 % DNA–DNA hybridization for species delineation corresponds to 96 % ANI. They also concluded that sequence similarity values of a concatenation of a random selection of six to eight genes should allow an accurate estimation of the total genome ANI value and give a significant prediction of whole-genome relatedness, even when the genes employed are among the worst-performing ones. This implies that similarity values from the concatenation of our housekeeping gene sequences may predict the total genome ANI values and, moreover, provide an easy tool to assess interorganismal relationships. In our study of seven concatenated genes, 2.7 % sequence divergence at the intraspecies level was deduced as the species delineation level. This can be regarded as corresponding to an ANI value of 97.3 %. Inclusion of additional strains for each species and/or more variable housekeeping genes could provide an even better correlation with the 96 % ANI value obtained by Konstantinidis *et al.* (2006).

A tree, including all examined *Ensifer* strains (except for *S. americanum*), was constructed for the concatenation of the seven congruent gene sequences applying the NJ, MP and ML methods. Regardless of which tree construction method was used, the same tree topology was obtained, and therefore only the ML tree is shown (Fig. 3). In line with most single-gene trees (Fig. 2), a close phylogenetic relationship was observed between *E. meliloti*, *E. medicae* and *E. arboris* (BT value 100 %) in the concatenated tree (Fig. 3). The combined analysis showed a cluster comprising all *Ensifer* strains, with two subclusters: one included '*S. morelense*' and the three *E. adhaerens* genomovars, while the other subcluster included all other *Ensifer* strains. This is consistent with most single-gene trees, although the BT values are not always significant (Martens *et al.*, 2007). However, in the concatenated tree, all mentioned clusters were supported by higher BT values than in the single-gene trees and were therefore more

robust. In addition, we created a concatenated tree of all 12 gene sequences (10 housekeeping and two rRNA genes) (data not shown). Compared to the results from the concatenation of the congruent genes, a nearly identical tree with similarly high BT values was found. The inclusion of non-congruent genes in the concatenation thus has little impact on the resulting tree topology. This is in line with the observation of Wertz *et al.* (2003) that, as more sequences were concatenated, the influence of genes with an aberrant signal was reduced and the underlying common phylogenetic signal was reinforced, as demonstrated by the increase in BT values.

Comparison of MLSA data with DNA–DNA reassociation data

A limited number of studies have compared analyses of housekeeping genes with DNA–DNA hybridization data. Yamamoto *et al.* (1999) demonstrated that the phylogenetic clustering of *Acinetobacter* strains based on sequences of a single gene, *gyrB*, was almost equivalent to genomic species delineation by DNA–DNA hybridization. Nørskov-Lauritsen *et al.* (2005) used the phylogeny of four housekeeping genes to study *Haemophilus*. They found that these phylogenies confirmed DNA–DNA hybridization groupings, whereas 16S rRNA gene phylogeny does not in all cases. Stackebrandt *et al.* (2007) found discrepancies between four housekeeping gene phylogenies for *Corallococcus* strains. The 16S rRNA gene phylogeny was only partially recovered, although it was better reflected in a concatenated tree. DNA–DNA hybridizations showed a larger diversity than reflected in the 16S rRNA gene phylogeny, but a clear comparison between housekeeping gene analyses and hybridization data was not made.

To evaluate the resolution of our MLSA data for taxonomic purposes, we compared them with a compilation of reported and new DNA–DNA hybridization results. Table 2 represents the DNA–DNA hybridization values available for *Ensifer* strains included in this study (de Lajudie *et al.*, 1994; Nick *et al.*, 1999; Peng *et al.*, 2002; Toledo *et al.*, 2003; Wang *et al.*, 2002; Willems *et al.*, 2003). Strains sharing over 70 % overall genome relatedness, and thus representing a single species, are grouped together.

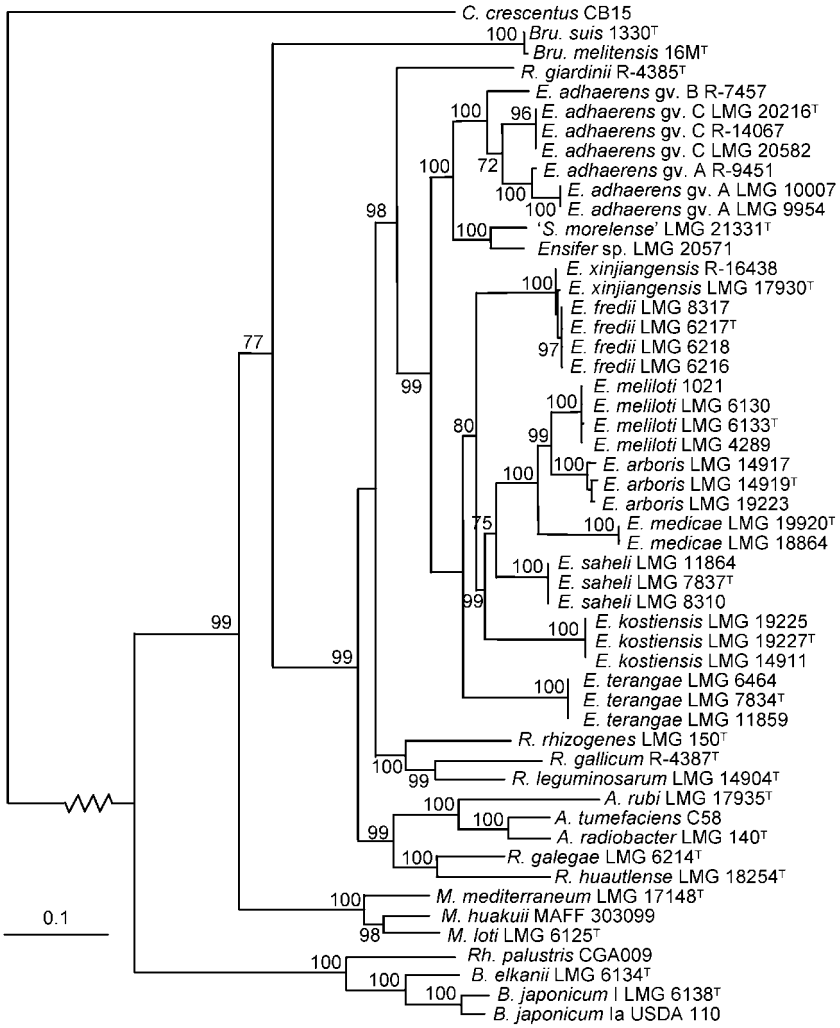

Fig. 3. Phylogenetic reconstruction based on the concatenated *recA*, *rpoB*, *gyrB*, *dnaK*, *glnA*, *gltA* and *thrC* gene sequences. Analyses were conducted using the ML method. BT values of 70 or more (using 100 replicates) are indicated at branching points. Bar, 0.1 % estimated substitutions.

The three *E. adhaerens* genomovars are related groups within *E. adhaerens* (DNA–DNA reassociation values between 50 and 70 %), distinguishable by genotypic but not by phenotypic tests (Willems *et al.*, 2003).

The linear Pearson's product-moment correlation coefficient was employed to permit a comparison of DNA–DNA reassociation data with our gene sequencing results (Supplementary Fig. S1). Product-moment (r) values were calculated between the DNA–DNA relatedness matrix and the corresponding similarity matrices of the respective gene sequences from *thrC*, *dnaK*, *gyrB*, *rpoB*, *glnA*, *gltA* and *recA*, a concatenation of these congruent genes and the 16S rRNA gene. Regression analysis showed a highly significant correlation between DNA–DNA hybridization values and sequence similarity values of the housekeeping genes, (r values ranging from 0.88 for *gyrB* to 0.93 for the seven concatenated genes). Correlation between DNA–DNA reassociation values and the 16S rRNA gene was lower (r=0.78). 16S rRNA gene sequence analysis can only reliably depict relationships to the species level for moderately related strains (below 97 % similarity) (Stackebrandt & Goebel, 1994). For example, 16S rRNA gene sequence similarities of *Ensifer* species may be as high as 100 % at DNA–DNA hybridization levels of 31 % (*S. americanum* LMG 22684^T and *E. fredii* LMG 6217^T). Correlation between similarity values for the 16S rRNA gene and the seven concatenated genes was also lower (r=0.78). The highest correlation (r=0.93) was found between DNA–DNA hybridization values and similarity values for the seven concatenated genes, which also supports the conclusion of Konstantinidis *et al.* (2006) that a concatenation of genes gives an accurate prediction of interorganismal relationships. The taxonomic resolution of housekeeping gene sequencing equals that of DNA–DNA hybridization; MLSA provides reliable information to the subspecies level. Indeed, at the intraspecies/genomovar level, relationships suggested by MLSA data (Fig. 3) corresponded with those revealed by DNA–DNA hybridization. For example, *E. adhaerens* gv. A strains LMG 9954 and LMG 10007 display 100 % sequence similarity (seven concatenated gene sequences) and shared 96 % DNA–DNA relatedness (mean value). Strains LMG 9954 and LMG 10007 displayed somewhat lower sequence similarity values with R-9451 (97.3 % similarity with both strains), corresponding with the slightly lower DNA–DNA hybridization values (mean values of 85 and 90 % respectively). At the interspecies/genomovar level, interorganismal relationships for highly related strains (sharing 50–70 % DNA–DNA relatedness) are more obvious from MLSA data. For example, *E. adhaerens* gv. A LMG 10007 and *E. adhaerens* gv. B R-7457 share 93.3 % sequence similarity for the concatenated gene sequences (compared with the cut-off value of 97.3 % for species level delineation; a clear gap is observed), while displaying a DNA–DNA relatedness of 67 % (cut-off value of 70 % for species-level delineation). A close relationship between the *E. adhaerens* genomovars is shown by both methods, but differentiation of the genomovars (distinction of the genomic species) is more clear from the MLSA data. In Supplementary Fig. S1, the gap between housekeeping gene sequence similarity values within and between species, i.e. the species boundary, is clear. From the DNA–DNA hybridization values, this gap between reassociation values within a species ($\geqslant$70 % relative DNA relatedness) and between species ($<$70 % relative DNA relatedness) is not clear at all. Goris *et al.* (2007) already noted that the 70 % relatedness rule for species delineation is rather arbitrary, since their DNA–DNA hybridization data show a continuous gradient of overall genetic relatedness rather than discrete species boundaries. For less closely related *Ensifer* strains ($<$50 % DNA–DNA relatedness), no information on particular interspecies/genomovar relationships is apparent from DNA–DNA hybridization. In contrast, our analyses of single and concatenated genes (seven or twelve) indicate some particular associations. For example, close relationships were observed between *S. americanum* and *E. fredii* (and *E. xinjiangensis*), between '*S. morelense*' and the three *E. adhaerens* genomovars and between *E. meliloti*, *E. medicae* and *E. arboris* (Figs 1, 2 and 3). These close relationships are not apparent from DNA–DNA reassociation levels (Table 2).

Since sequence similarity values and tree topologies (Figs 1, 2 and 3) were congruent with the genomic species previously delineated on the basis of DNA–DNA hybridization studies, our study indicates that even the single housekeeping gene analyses provide a robust species delineation that is at least equivalent and even superior to DNA–DNA hybridization (Supplementary Fig. S1). Delineating strains into species, based solely on the MLSA data and without prior knowledge of their classification, would result in the same genomic species. MLSA, like DNA–DNA hybridization, is a suitable technique for species delineation and for assessing relationships at the intraspecies level. MLSA surpasses DNA–DNA hybridization by its ability to give information on interspecies relationships and by providing clear species/genomovar boundaries.

Comparison of the MLSA data from our set of strains with DNA profiling methods such as rep-PCR and whole-genome dot-blot hybridization (Nick *et al.*, 1999) indicates clusters of the same genomic species. However, intraspecific genomic variation is more pronounced for the fingerprint techniques that cover a larger part of the genome. A selection of more strains per species and more variable genes to study intraspecific diversity, entering the field of multilocus sequence typing (MLST) (Maiden *et al.*, 1998), should provide higher variability within species and even better correlation with genomic methods. Since whole-genome methods such as rep-PCR take into account plasmids and non-essential genes and DNA, they can reveal strains that are intermediate between species that were delineated based on a selected set of significant characteristics. Because MLSA targets housekeeping genes, we would expect fewer intermediate strains and more clear-cut groupings.

Table 2. DNA relatedness between *Ensifer* strains assessed by DNA–DNA hybridization

Results taken from other studies are indicated as follows: *a*, Willems *et al.* (2003) (values obtained using a microplate hybridization method); *b*, Toledo *et al.* (2003) (microplate hybridization); *c*, Nick *et al.* (1999) (spectrophotometric); *d*, Peng *et al.* (2002) (spectrophotometric); *e*, Wang *et al.* (2002) (Southern hybridization of total DNA); *f*, de Lajudie *et al.* (1994) (spectrophotometric); *g*, Wang *et al.* (2002) (microplate hybridization).

Strain used for labelled probe	Binding (%) with unlabelled DNA from strain:																								
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
E. adhaerens gv. A																									
1. R-9451	**100**	70[a]	84[a]	64[a]	51[a]																				28[a]
2. LMG 9954	100[a]	**100**	100[a]		62[a]																				32[a]
3. LMG 10007	96[a]	92[a]	**100**	67[a]	51[a]	43[a]					24[a]											23[a]		31[a]	35[a]
4. *E. adhaerens* gv. B R-7457	62[a]		68[a]	**100**	54[a]	46[a]					28[a]											27[a]		34[a]	37[a]
E. adhaerens gv. C																									
5. LMG 20216[T]		62[a]	45[a]	53[a]	**100**	92[a]	97[a]				25[a]											25[a]		41[a]	46[a]
6. R-14067			49[a]	57[af]	96[a]	**100**	100[a]				25[a]											23[a]		40[a]	41[a]
7. LMG 20582					93[a]	93[a]	**100**																	41[a]	
8. *S. americanum* LMG 22684[T]								**100**	10[b]		31[b]		10[b]				13[b]		3[b]	8[b]					
E. arboris																									
9. LMG 14919[T]								10[b]	**100**	95[c]	2[b]						28[c]	28[c]		18[c]		0[d]		33[e]	
10. LMG 19223									95[c]	**100**	19[c]														
E. fredii																									
11. LMG 6217[T]			25[a]	29[a]	26[a]	23[a]		31[b]	2[b]	19[c]	**100**	90	21[b]	11[b]			23[cf]		2[b]/26[f]	5[b]	22[cf]	84	78	22[a]	29[a]
12. LMG 8317											90	**100**										88	74		
13. *E. kostiensis* LMG 19227[T]								10[b]			21[b]		**100**					20[c]				3[d]		34[e]	
E. medicae																									
14. LMG 19920[T]											11[b]			**100**	71[c]	80[c]	30[c]					28[d]		39[e]	
15. LMG 19921														71[c]	**100**	93[c]									
16. LMG 18864														80[c]	93[c]	**100**									
17. *E. meliloti* LMG 6133[T]								13[b]	28[c]		48[b]/ 23[cf]			30[c]			**100**		23[cf]		22[cf]	21[d]		17[e]	
E. saheli																									
18. LMG 7837[T]									28[c]				20[c]					**100**	89[f]			21[d]		31[e]	
19. LMG 8310								3[b]			2[b]/26[f]						23[cf]	89[f]	**100**		20[cf]				
E. terangae																									
20. LMG 7834[T]								8[b]	18[c]		5[b]									**100**	79[cf]	29[d]		31[e]	
21. LMG 6464											22[cf]						22[cf]		20[cf]	79[cf]	**100**				
E. xinjiangensis																									
22. LMG 17930[T]			22[a]	23[a]	22[a]	19[a]			0[d]		84	88	3[d]	28[d]			21[d]	21[d]		29[d]		**100**	79	19[a]	25[a]
23. R-16438											78	74										79	**100**		
24. '*S. morelense*' LMG 21331[T]			32[a]	39[a]	41[ag]	39[ag]	41[a]		33[e]		26[a]/19[e]		34[e]	39[e]			17[e]	31[e]		31[e]		23[a]/27[e]		**100**	66[a]
25. *Ensifer* sp. LMG 20571	25[a]	19[a]	28[a]	29[a]	37[a]	36[a]					23[a]											23[a]		52[a]	**100**

In conclusion, MLSA offers a very good, reliable alternative to DNA–DNA hybridization for the study of genomic relationships between bacteria from a particular group. It has the important advantage of yielding cumulative, exact data which, through database query, can be compared with sequence data from unknown organisms. MLSA of selected housekeeping genes, although not a genome-wide comparison technique, accurately predicts relationships between closely related organisms. It has great potential for species delineation and identification and for studying bacterial relationships at a wide range of evolutionary distances, from the intraspecies level to at least the genus level. For identification purposes, it seems prudent to study at least two independent housekeeping genes, since lateral gene transfer instances in a particular gene can not be excluded. A general identification strategy for new isolates could consist of initial partial sequencing of the 16S rRNA gene for genus-level identification. This remains valuable because of the large and comprehensive database available. This information can then guide the selection of suitable housekeeping genes (as a function of available reference data) for species identification. For phylogenetic analyses, more housekeeping genes should be analysed than for identification purposes. Incongruent phylogenies indicate possible horizontal transfer, whereas congruent phylogenies reflect the common history of genes. In the case of rhizobia, housekeeping genes with good capability to identify and classify strains are the *gyrB*, *gltA*, *recA* and *thrC* genes.

ACKNOWLEDGEMENTS

This work was performed in the framework of project QLK3-CT-2002-02097 funded by the Commission of the European Communities and project G.0156.02 funded by the Fund for Scientific Research – Flanders. A. W. is grateful for a post-doctoral fellowship of the Fund for Scientific Research – Flanders. We thank Manuel Delaere for his contribution to the phylogenetic analyses.

REFERENCES

Acinas, S. G., Marcelino, L. A., Klepac-Ceraj, V. & Polz, M. F. (2004). Divergence and redundancy of 16S rRNA sequences in genomes with multiple *rrn* operons. *J Bacteriol* **186**, 2629–2635.

Adekambi, T. & Drancourt, M. (2004). Dissection of phylogenetic relationships among 19 rapidly growing *Mycobacterium* species by 16S rRNA, *hsp65*, *sodA*, *recA* and *rpoB* gene sequencing. *Int J Syst Evol Microbiol* **54**, 2095–2105.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.

Amann, R. I., Lin, C. H., Key, R., Montgomery, L. & Stahl, D. A. (1992). Diversity among *Fibrobacter* strains: towards a phylogenetic classification. *Syst Appl Microbiol* **15**, 23–32.

Baele, M., Baele, P., Vaneechoutte, M., Storms, V., Butaye, P., Devriese, L. A., Verschraegen, G., Gillis, M. & Haesebrouck, F. (2000). Application of tRNA intergenic spacer PCR for identification of *Enterococcus* species. *J Clin Microbiol* **38**, 4201–4207.

Charles, L., Carbone, I., Davies, K. G., Bird, D., Burke, M., Kerry, B. R. & Opperman, C. H. (2005). Phylogenetic analysis of *Pasteuria penetrans* by use of multiple genetic loci. *J Bacteriol* **187**, 5700–5708.

Cho, J. C. & Tiedje, J. M. (2001). Bacterial species determination from DNA-DNA hybridization by using genome fragments and DNA microarrays. *Appl Environ Microbiol* **67**, 3677–3682.

Christensen, H. & Olsen, J. E. (1998). Phylogenetic relationships of *Salmonella* based on DNA sequence comparison of *atpD* encoding the beta subunit of ATP synthase. *FEMS Microbiol Lett* **161**, 89–96.

Christensen, H., Kuhnert, P., Olsen, J. E. & Bisgaard, M. (2004). Comparative phylogenies of the housekeeping genes *atpD*, *infB* and *rpoB* and the 16S rRNA gene within the *Pasteurellaceae*. *Int J Syst Evol Microbiol* **54**, 1601–1609.

Coenye, T., Gevers, D., Van de Peer, Y., Vandamme, P. & Swings, J. (2005). Towards a prokaryotic genomic taxonomy. *FEMS Microbiol Rev* **29**, 147–167.

de Lajudie, P., Willems, A., Pot, B., Dewettinck, D., Maestrojuan, G., Neyra, M., Collins, M. D., Dreyfus, B., Kersters, K. & Gillis, M. (1994). Polyphasic taxonomy of rhizobia – emendation of the genus *Sinorhizobium* and description of *Sinorhizobium meliloti* comb. nov., *Sinorhizobium saheli* sp. nov. and *Sinorhizobium terangae* sp. nov. *Int J Syst Bacteriol* **44**, 715–733.

DelVecchio, V. G., Kapatral, V., Redkar, R. J., Patra, G., Mujer, C., Los, T., Ivanova, N., Anderson, I., Bhattacharyya, A. & other authors (2002). The genome sequence of the facultative intracellular pathogen *Brucella melitensis*. *Proc Natl Acad Sci U S A* **99**, 443–448.

Euzéby, J. P. & Tindall, B. J. (2004). Status of strains that contravene Rules 27(3) and 30 of the Bacteriological Code. Request for an Opinion. *Int J Syst Evol Microbiol* **54**, 293–301.

Ezaki, T., Hashimoto, Y. & Yabuuchi, E. (1989). Fluorometric deoxyribonucleic acid-deoxyribonucleic acid hybridization in micro-dilution wells as an alternative to membrane filter hybridization in which radioisotopes are used to determine genetic relatedness among bacterial strains. *Int J Syst Bacteriol* **39**, 224–229.

Farris, J. S., Kallersjo, M., Kluge, A. G. & Bult, C. (1995). Constructing a significance test for incongruence. *Syst Biol* **44**, 570–572.

Fox, G. E., Wisotzkey, J. D. & Jurtshuk, P., Jr (1992). How close is close: 16S ribosomal RNA sequence identity may not be sufficient to guarantee species identity. *Int J Syst Bacteriol* **42**, 166–170.

Galibert, F., Finan, T. M., Long, S. R., Puhler, A., Abola, P., Ampe, F., Barloy-Hubler, F., Barnett, M. J., Becker, A. & other authors (2001). The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science* **293**, 668–672.

Gevers, D., Cohan, F. M., Lawrence, J. G., Spratt, B. G., Coenye, T., Feil, E. J., Stackebrandt, E., Van de Peer, Y., Vandamme, P. & other authors (2005). Re-evaluating prokaryotic species. *Nat Rev Microbiol* **3**, 733–739.

Goodfellow, M., Manfio, G. P. & Chun, J. (1997). Towards a practical species concept for cultivable bacteria. In *Species: the Units of Biodiversity*, pp. 25–59. Edited by M. F. Claridge & H. A. Dawah. London: Chapman & Hall.

Goodner, B., Hinkle, G., Gattung, S., Miller, N., Blanchard, M., Qurollo, B., Goldman, B. S., Cao, Y., Askenazi, M. & other authors (2001). Genome sequence of the plant pathogen and biotechnology agent *Agrobacterium tumefaciens* C58. *Science* **294**, 2323–2328.

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P. & Tiedje, J. M. (2007). DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* **57**, 81–91.

Holmes, D. E., Nevin, K. P. & Lovley, D. R. (2004). Comparison of 16S rRNA, *nifD*, *recA*, *gyrB*, *rpoB* and *fusA* genes within the family *Geobacteraceae* fam. nov. *Int J Syst Evol Microbiol* **54**, 1591–1599.

Jaspers, E. & Overmann, J. (2004). Ecological significance of microdiversity: identical 16S rRNA gene sequences can be found in bacteria with highly divergent genomes and ecophysiologies. *Appl Environ Microbiol* **70**, 4831–4839.

Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S., Watanabe, A., Idesawa, K., Ishikawa, K. & other authors (2000). Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. *DNA Res* **7**, 331–338.

Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T., Sasamoto, S., Watanabe, A., Idesawa, K., Iriguchi, M. & other authors (2002). Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA Res* **9**, 189–197.

Konstantinidis, K. T. & Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci U S A* **102**, 2567–2572.

Konstantinidis, K. T., Ramette, A. & Tiedje, J. M. (2006). Toward a more robust assessment of intraspecies diversity, using fewer genetic markers. *Appl Environ Microbiol* **72**, 7286–7293.

Kwon, S. W., Park, J. Y., Kim, J. S., Kang, J. W., Cho, Y. H., Lim, C. K., Parker, M. A. & Lee, G. B. (2005). Phylogenetic analysis of the genera *Bradyrhizobium*, *Mesorhizobium*, *Rhizobium* and *Sinorhizobium* on the basis of 16S rRNA gene and internally transcribed spacer region sequences. *Int J Syst Evol Microbiol* **55**, 263–270.

Larimer, F. W., Chain, P., Hauser, L., Lamerdin, J., Malfatti, S., Do, L., Land, M. L., Pelletier, D. A., Beatty, J. T. & other authors (2004). Complete genome sequence of the metabolically versatile photosynthetic bacterium *Rhodopseudomonas palustris*. *Nat Biotechnol* **22**, 55–61.

Maiden, M. C. J., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., Zhang, Q., Zhou, J. J., Zurth, K. & other authors (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* **95**, 3140–3145.

Marmur, J. (1961). A procedure for the isolation of deoxyribonucleic acid from microorganisms. *J Mol Biol* **3**, 208–218.

Martens, M., Delaere, M., Coopman, R., De Vos, P., Gillis, M. & Willems, A. (2007). Multilocus sequence analysis of *Ensifer* and related taxa. *Int J Syst Evol Microbiol* **57**, 489–503.

Mollet, C., Drancourt, M. & Raoult, D. (1997). *rpoB* sequence analysis as a novel basis for bacterial identification. *Mol Microbiol* **26**, 1005–1011.

Naser, S. M., Thompson, F. L., Hoste, B., Gevers, D., Dawyndt, P., Vancanneyt, M. & Swings, J. (2005). Application of multilocus sequence analysis (MLSA) for rapid identification of *Enterococcus* species based on *rpoA* and *pheS* genes. *Microbiology* **151**, 2141–2150.

Naser, S. M., Dawyndt, P., Hoste, B., Gevers, D., Vandemeulebroecke, K., Cleenwerck, I., Vancanneyt, M. & Swings, J. (2007). Identification of lactobacilli by *pheS* and *rpoA* gene sequence analyses. *Int J Syst Evol Microbiol* **57**, 2777–2789.

Nick, G., Jussila, M., Hoste, B., Niemi, R. M., Kaijalainen, S., de Lajudie, P., Gillis, M., de Bruijn, F. J. & Lindström, K. (1999). Rhizobia isolated from root nodules of tropical leguminous trees characterized using DNA-DNA dot-blot hybridisation and rep-PCR genomic fingerprinting. *Syst Appl Microbiol* **22**, 287–299.

Nierman, W. C., Feldblyum, T. V., Laub, M. T., Paulsen, I. T., Nelson, K. E., Eisen, J., Heidelberg, J. F., Alley, M. R. K., Ohta, N. & other authors (2001). Complete genome sequence of *Caulobacter crescentus*. *Proc Natl Acad Sci U S A* **98**, 4136–4141.

Nørskov-Lauritsen, N., Bruun, B. & Kilian, M. (2005). Multilocus sequence phylogenetic study of the genus *Haemophilus* with description of *Haemophilus pittmaniae* sp. nov. *Int J Syst Evol Microbiol* **55**, 449–456.

Owen, R. J. & Pitcher, D. (1983). Current methods for determining DNA-base composition and levels of DNA-DNA hybridization. *J Appl Bacteriol* **55**, R16–R16.

Paulsen, I. T., Seshadri, R., Nelson, K. E., Eisen, J. A., Heidelberg, J. F., Read, T. D., Dodson, R. J., Umayam, L., Brinkac, L. M. & other authors (2002). The *Brucella suis* genome reveals fundamental similarities between animal and plant pathogens and symbionts. *Proc Natl Acad Sci U S A* **99**, 13148–13153.

Peng, G. X., Tan, Z. Y., Wang, E. T., Reinhold-Hurek, B., Chen, W. F. & Chen, W. X. (2002). Identification of isolates from soybean nodules in Xinjiang Region as *Sinorhizobium xinjiangense* and genetic differentiation of *S. xinjiangense* from *Sinorhizobium fredii*. *Int J Syst Evol Microbiol* **52**, 457–462.

Posada, D. & Buckley, T. R. (2004). Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Syst Biol* **53**, 793–808.

Posada, D. & Crandall, K. A. (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818.

Rokas, A., King, N., Finnerty, J. & Carroll, S. B. (2003). Conflicting phylogenetic signals at the base of the metazoan tree. *Evol Dev* **5**, 346–359.

Rönner, S., Liesack, W., Wolters, J. & Stackebrandt, E. (1991). Cloning and sequencing of a large fragment of the *atpD* gene of *Pirellula marina* – a contribution to the phylogeny of *Planctomycetales*. *Endocytobiosis Cell Res* **7**, 219–229.

Rosselló-Mora, R. (2006). DNA-DNA reassociation methods applied to microbial taxonomy and their critical evaluation. In *Molecular Identification, Systematics, and Population Structure of Prokaryotes*, pp. 23–50. Edited by E. Stackebrandt. Heidelberg: Springer.

Selenska-Pobell, S. & Evguenieva-Hackenberg, E. (1995). Fragmentations of the large-subunit rRNA in the family *Rhizobiaceae*. *J Bacteriol* **177**, 6993–6998.

Stackebrandt, E. & Goebel, B. M. (1994). Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int J Syst Bacteriol* **44**, 846–849.

Stackebrandt, E., Frederiksen, W., Garrity, G. M., Grimont, P. A., Kämpfer, P., Maiden, M. C. J., Nesme, X., Rosselló-Mora, R., Swings, J. & other authors (2002). Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int J Syst Evol Microbiol* **52**, 1043–1047.

Stackebrandt, E., Päuker, O., Steiner, U., Schumann, P., Sträubler, B., Heibei, S. & Lang, E. (2007). Taxonomic characterization of members of the genus *Corallococcus*: molecular divergence versus phenotypic coherency. *Syst Appl Microbiol* **30**, 109–118.

Sullivan, J. T., Eardly, B. D., van Berkum, P. & Ronson, C. W. (1996). Four unnamed species of nonsymbiotic rhizobia isolated from the rhizosphere of *Lotus corniculatus*. *Appl Environ Microbiol* **62**, 2818–2825.

Swofford, D. L. (2002). PAUP*: phylogenetic analysis using parsimony (and other methods), version 4. Sunderland, MA: Sinauer Associates.

Tan, Z. Y., Xu, X. D., Wang, E. T., Gag, J. L., Martínez-Romero, E. & Chen, W. X. (1997). Phylogenetic and genetic relationships of *Mesorhizobium tianshanense* and related rhizobia. *Int J Syst Bacteriol* **47**, 874–879.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**, 4876–4882.

Thompson, F. L., Gevers, D., Thompson, C. C., Dawyndt, P., Naser, S., Hoste, B., Munn, C. B. & Swings, J. (2005). Phylogeny and molecular

identification of vibrios on the basis of multilocus sequence analysis. *Appl Environ Microbiol* **71**, 5107–5115.

Toledo, I., Lloret, L. & Martínez-Romero, E. (2003). *Sinorhizobium americanum* sp. nov., a new *Sinorhizobium* species nodulating native *Acacia* spp. in Mexico. *Syst Appl Microbiol* **26**, 54–64.

Turner, S. L., Zhang, X.-X., Li, F.-D. & Young, J. P. W. (2002). What does a bacterial genome sequence represent? Mis-assignment of MAFF 303099 to the genospecies *Mesorhizobium loti*. *Microbiology* **148**, 3330–3331.

Van Camp, G., Chapelle, S. & De Wachter, R. (1993). Amplification and sequencing of variable regions in bacterial 23S ribosomal RNA genes with conserved primer sequences. *Curr Microbiol* **27**, 147–151.

Wang, E. T., Tan, Z. Y., Willems, A., Fernández-López, M., Reinhold-Hurek, B. & Martínez-Romero, E. (2002). *Sinorhizobium morelense* sp nov., a *Leucaena leucocephala*-associated bacterium that is highly resistant to multiple antibiotics. *Int J Syst Evol Microbiol* **52**, 1687–1693.

Wayne, L. G., Brenner, D. J., Colwell, R. R., Grimont, P. A. D., Kandler, O., Krichevsky, M. I., Moore, L. H., Moore, W. E. C., Murray, R. G. E. & other authors (1987). International Committee on Systematic Bacteriology. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int J Syst Bacteriol* **37**, 463–464.

Wernersson, R. & Pedersen, A. G. (2003). RevTrans: multiple alignment of coding DNA from aligned amino acid sequences. *Nucleic Acids Res* **31**, 3537–3539.

Wertz, J. E., Goldstone, C., Gordon, D. M. & Riley, M. A. (2003). A molecular phylogeny of enteric bacteria and implications for a bacterial species concept. *J Evol Biol* **16**, 1236–1248.

Willems, A., Doignon-Bourcier, F., Goris, J., Coopman, R., de Lajudie, P., De Vos, P. & Gillis, M. (2001). DNA–DNA hybridization study of *Bradyrhizobium* strains. *Int J Syst Evol Microbiol* **51**, 1315–1322.

Willems, A., Fernández-López, M., Muñoz-Adelantado, E., Goris, J., De Vos, P., Martínez-Romero, E., Toro, N. & Gillis, M. (2003). Description of new *Ensifer* strains from nodules and proposal to transfer *Ensifer adhaerens* Casida 1982 to *Sinorhizobium* as *Sinorhizobium adhaerens* comb. nov. Request for an Opinion. *Int J Syst Evol Microbiol* **53**, 1207–1217.

Woese, C. R. (1987). Bacterial evolution. *Microbiol Rev* **51**, 221–271.

Wood, D. W., Setubal, J. C., Kaul, R., Monks, D. E., Kitajima, J. P., Okura, V. K., Zhou, Y., Chen, L., Wood, G. E. & other authors (2001). The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. *Science* **294**, 2317–2323.

Yamamoto, S. & Harayama, S. (1995). PCR amplification and direct sequencing of *gyrB* genes with universal primers and their application to the detection and taxonomic analysis of *Pseudomonas putida* strains. *Appl Environ Microbiol* **61**, 1104–1109.

Yamamoto, S., Bouvet, P. J. M. & Harayama, S. (1999). Phylogenetic structures of the genus *Acinetobacter* based on *gyrB* sequences: comparison with the grouping by DNA–DNA hybridization. *Int J Syst Bacteriol* **49**, 87–95.

Young, J. M. (2003). The genus name *Ensifer* Casida 1982 takes priority over *Sinorhizobium* Chen *et al.* 1988, and *Sinorhizobium morelense* Wang *et al.* 2002 is a later synonym of *Ensifer adhaerens* Casida 1982. Is the combination '*Sinorhizobium adhaerens*' (Casida 1982) Willems *et al.* 2003 legitimate? Request for an Opinion. *Int J Syst Evol Microbiol* **53**, 2107–2110.

Zeigler, D. R. (2003). Gene sequences useful for predicting relatedness of whole genomes in bacteria. *Int J Syst Evol Microbiol* **53**, 1893–1900.