Multilocus sequence analysis of root nodule isolates from *Lotus arabicus* (Senegal), *Lotus creticus*, *Argyrolobium uniflorum* and *Medicago sativa* (Tunisia) and description of *Ensifer numidicus* sp. nov. and *Ensifer garamanticus* sp. nov.

C. Merabet,1,2 M. Martens,3 M. Mahdhi,2,4 F. Zakhia,2† A. Sy,5 C. Le Roux,2 O. Domergue,2 R. Coopman,3 A. Bekki,1 M. Mars,4 A. Willems3 and P. de Lajudie2

1Laboratoire de Rhizobiologie, Université d’Oran, Es-Senia, Algeria
2IRD, UMR 113 Symbioses Tropicales et Méditerranéennes F-34398 Montpellier, France
3Laboratorium voor Microbiologie (WE10), Universiteit Gent, Belgium
4Faculté des Sciences, Gabès, Tunisia
5IRD-UCAD-ISRA, Laboratoire Commun de Microbiologie, Dakar, Senegal, West Africa

Nine isolates from *Argyrolobium uniflorum*, *Lotus creticus*, *Medicago sativa* (Tunisia) and *Lotus arabicus* (Senegal) were analysed by multilocus sequence analysis (MLSA) of five housekeeping genes (*recA*, *atpD*, *glnA*, *gltA* and *thrC*), the 16S rRNA gene and the nodulation gene *nodA*. Analysis of the individual and concatenated gene sequences demonstrated that the nine new strains constituted three stable, well-supported (bootstrap and gene sequence similarity values) monophyletic clusters, A, B and C, all belonging to the branch of the genus *Ensifer*, regardless of the phylogenetic reconstruction method used (maximum likelihood, maximum-parsimony, neighbour-joining). The three groups were further characterized by API 100 auxanographic tests, host specificity and *nodA* gene sequence analysis. On the basis of these data, clusters A and C are suggested as representing two novel species within the genus *Ensifer*, for which the names *Ensifer numidicus* sp. nov. (type strain ORS 1407T=LMG 24690T=CIP 109850T) and *Ensifer garamanticus* sp. nov. (type strain ORS 1400T=LMG 24692T=CIP 109916T) are proposed. The cluster B strains were assigned to *Ensifer adhaerens* genomovar A.

**INTRODUCTION**

Over the past three decades, bacterial taxonomy has been based on a polyphasic approach including both phenotypic and genotypic data with 16S rRNA gene sequence analysis and DNA–DNA hybridization data being considered as the cornerstones of genotypic characterization for descriptions of novel species (Vandamme et al., 1996; Stackebrandt et al., 2002; Coenye et al., 2005). However, many authors have reported drawbacks with these methods (Ueda et al., 1999; Stackebrandt et al., 2002; Schouls et al., 2003; Stackebrandt, 2006).
2003; van Berkum et al., 2003; Eardly et al., 2005). The ad hoc Committee for the Re-evaluation of the Species Definition in Bacteriology (Stackebrandt et al., 2002; Gevers et al., 2005) suggested the sequence analysis of a set of protein-encoding genes as alternative phylogenetic markers, or multilocus sequence analysis (MLSA). Several recent studies have confirmed that sequences from housekeeping genes can be used for identification at the species level (Wertz et al., 2003; Zeigler, 2003), for evolutionary population genetics and in taxonomy (Stepkowska et al., 2003; Vinuesa et al., 2005a,b; Bailly et al., 2006). The analysis of a small number of carefully selected gene sequences may equal or even surpass the precision of DNA–DNA hybridization for the quantification of genome relatedness and this approach thus has the potential to replace cumbersome DNA–DNA hybridizations (Zeigler, 2003, Martens et al., 2008). Mantelin et al. (2006) included MLSA in the description of four novel species of the genus Pseudomonas and Vinuesa et al. (2005c) described novel species of the genus Bradyrhizobium using phylogenetic analysis of three housekeeping genes, atpD, gltA and recA, combined with other classical genotypic and phenotypic analyses. ‘Ensifer mexicanus’ and ‘Sinorhizobium chiapane- cun’ were described using phenotypic analysis, phylogenies of the recA, gyrA, nolR, rpoB, rrs and symbiotic genes and confidence intervals of sequence similarity to estimate both inter- and intraspecies variation; all were in correlation with DNA–DNA hybridization data (Lloret et al., 2007; Rincón-Rosales et al., 2009). In addition, MLSA has been applied in diversity studies of the genera Enterococcus, Aeromonas and Ensifer (Sinorhizobium) (Naser et al., 2005; Soler et al., 2004; Bailly et al., 2006) and it has been reported as being a highly reproducible and economical method for rapid and reliable species identification.

The genera Sinorhizobium and Ensifer were recently rec-ognized as forming a single phylogenetic clade (Balkwill, 2005; Willems et al., 2003) and are now united and all species of the genus Sinorhizobium have been transferred to the genus Ensifer, in line with rule 38 of the Bacteriological Code (Young, 2003; Judicial Commission, 2008). The genus Ensifer currently includes eleven species (Wang et al., 2002; Wei et al., 2002; Young, 2003; Toledo et al. 2003). Two further species, ‘Ensifer mexicanus’ and ‘Sinorhizobium chiapane- cum’, have been described but the names have not yet been validly published (Lloret et al. 2007; Rincón-Rosales et al., 2009). Martens et al. (2007, 2008) recently demonstrated that the discriminatory power of MLSA for species identification and delineation is higher than 16S rRNA gene sequence analysis and DNA–DNA hybridization within the genus Ensifer.

In this study, MLSA was performed to complete the characterization of nine root nodule strains isolated from several legumes in Tunisia (Argyrolobium uniflorum, Lotus creticus, Medicago sativa) and Senegal (Lotus arabicus). Argyrolobium uniflorum (Tribe Genistaeae, Family Fabaceae) is an indigenous herb legume of the Mediterranean basin and is a pastoral and forage legume that is widely distributed in arid and semi-arid regions of Tunisia. This plant plays an important role in the maintenance of soil fertility, soil coverage and dune stability (Ferchichi, 1996). Lotus arabicus L. (tribe Loteae, family Fabaceae), syn. Lotus mossamedensis Welw. ex Baker, is a spontaneous annual herb legume of coastal dunes in Senegal. Medicago sativa is the most widely cultivated species of lucerne in the world and its symbiotic rhizobial strains have been extensively studied. Six of the isolates have previously been analysed by amplified rDNA restriction analysis (ARDRA) using seven enzymes, total cell protein SDS-PAGE analysis and 16S rRNA gene sequencing (Zakhia et al., 2004) and results suggested their separate positions in the genus Ensifer (pro synon. Sinorhizobium). According to the conclusions of Martens et al. (2007, 2008), the strains were subjected to MLSA using five housekeeping genes, the most discriminative for taxonomic purposes of the ten housekeeping genes evaluated in the genus Ensifer. Genes were analysed individually and as a concatenation using only the congruent genes estimated by the incongruence-length difference (ILD) test (Farris et al., 1995). Trees were constructed using three algorithmic methods for comparison. The characterization was completed by auxanographic tests, DNA–DNA hybridizations, host specificity investigations and sequence analysis of the symbiotic nodA gene for biobar description.

METHODS

Bacterial strains and culture media. The strains investigated are listed in Table 1. All strains were grown on yeast-mannitol agar (YMA) (Zakhia et al., 2004) at 28 °C. Isolates were stored at ~80 °C in 50 % (v/v) glycerol.

Host specificity. The isolates were tested for nodule formation on their original host plant. The seeds from Medicago sativa and Lotus arabicus were sterilized in 3 % (w/v) calcium hypochlorite and scarified by immersion in 96 % (v/v) sulphuric acid for 20 min. The seeds were then washed five times with sterile water and placed in 1 % (w/v) agar–water at 24–25 °C for germination. After 3–4 days, the seedlings were transferred to agar slant tubes (Bertrand, 1997) for root nodulation trials (5 plants per strain). Plants were grown under continuous light (20 W m−2) at 24 °C and inoculated with 1 ml exponential growth phase YM bacterial suspension. Roots were observed for nodule formation during the first 4 weeks after inoculation. The infectivity of strains isolated from Argyrolobium uniflorum has been previously reported by Zakhia (2004).

Genomic DNA isolation. Bacterial DNA was prepared as described by Zakhia et al. (2004) or alternatively by the alkaline lysis method (Baele et al., 2000).

DNA amplification. The 16S rRNA gene, internal fragments of five housekeeping genes (atpD, recA, glnA, thrC and glnA) and the nodA gene were amplified using primers listed in Supplementary Table S1 (available in IJSEM Online). PCR amplification of the atpD, recA, glnA, thrC and glnA genes was performed as described by Martens et al. (2007) with the exception that the total volume of the reaction mixture was 25 μl. PCR amplification of the 16S rRNA gene was performed according to Zakhia et al. (2004) and nodA gene amplification was conducted as described by Zakhia et al. (2006).

Gene sequencing and phylogenetic data analysis. The PCR products of the different genes were purified and sequenced as previously
Table 1. GenBank accession numbers for new sequence data

<table>
<thead>
<tr>
<th>Strain</th>
<th>SSU</th>
<th>recA</th>
<th>glnA</th>
<th>atpD</th>
<th>thrC</th>
<th>gltA</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. numidicus sp. nov.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORS 1444 (=LMG 24691= CIP 199850)</td>
<td>AM946555</td>
<td>AM946557</td>
<td>AM946560</td>
<td>AM946549</td>
<td>AM946586</td>
<td>AM946596</td>
<td>Argyrobium uniflorum, Tunisia, Zakhia et al. (2004)</td>
</tr>
<tr>
<td>ORS 1410 (=LMG 24736= CIP 199858)</td>
<td>AM946556</td>
<td>AM946557</td>
<td>AM946561</td>
<td>AM946550</td>
<td>NA</td>
<td>AM946595</td>
<td>Lotus catticus, Tunisia, Zakhia et al. (2004)</td>
</tr>
<tr>
<td>ORS 1407T (=LMG 24690T= CIP 199850T)</td>
<td>AM946554</td>
<td>AM946556</td>
<td>AM946560</td>
<td>AM946551</td>
<td>AM946585</td>
<td>AM946594</td>
<td>Argyrobium uniflorum, Tunisia, Zakhia et al. (2004)</td>
</tr>
<tr>
<td>E. garamanticus sp. nov.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORS 1400 (=LMG 24692T= CIP 199916T)</td>
<td>AM946555</td>
<td>AM946557</td>
<td>AM946558</td>
<td>AM946546</td>
<td>AM946582</td>
<td>AM946591</td>
<td>Argyrobium uniflorum, Tunisia, Zakhia et al. (2004)</td>
</tr>
<tr>
<td>ORS 1401 (=LMG 24693= CIP 199848)</td>
<td>AM946567</td>
<td>AM946574</td>
<td>AM946558</td>
<td>AM946547</td>
<td>AM946583</td>
<td>AM946592</td>
<td>Argyrobium uniflorum, Tunisia, Zakhia et al. (2004)</td>
</tr>
<tr>
<td>STM 354 (=LMG 24694= CIP 199849)</td>
<td>AM946568</td>
<td>AM946575</td>
<td>AM946559</td>
<td>AM946548</td>
<td>AM946584</td>
<td>AM946593</td>
<td>Medicago sativa, Tunisia, Zakhia et al. (2004)</td>
</tr>
<tr>
<td>E. adhaerens genovar A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORS 2154</td>
<td>AM946569</td>
<td>AM946579</td>
<td>AM946587</td>
<td>AM946553</td>
<td>AM946587</td>
<td>AM946598</td>
<td>Lotus arabicus, Senegal, this study</td>
</tr>
<tr>
<td>ORS 2133</td>
<td>AM946570</td>
<td>AM946580</td>
<td>AM946588</td>
<td>AM946554</td>
<td>AM946588</td>
<td>AM946597</td>
<td>Lotus arabicus, Senegal, this study</td>
</tr>
<tr>
<td>ORS 529</td>
<td>AM946571</td>
<td>AM946581</td>
<td>AM946589</td>
<td>AM946555</td>
<td>AM946589</td>
<td>AM946599</td>
<td>Lotus arabicus, Senegal, this study</td>
</tr>
</tbody>
</table>

API 100 galleries (BioMérieux) were used to test the utilization of carbon sources (carbohydrates, amino acids, organic acids) for bacterial growth. Inoculants were obtained from 36 h YMA slant cultures. After inoculation, the galleries were incubated at 30 °C and the results were determined after 1, 2, 4 and 7 days. The results of the auxanographic tests were scored as described previously (Kersters & De Ley, 1984). Jaccard similarity coefficients were calculated and an UPGMA analysis was performed using BioNumerics 4.6.

RESULTS

Nodulation tests

All of the strains studied induced efficient nodulation on their hosts of isolation. This has already been reported by Zakhia (2004) for strains ORS 1400T, ORS 1401, ORS 1407T, ORS 1444 (Argyrobium uniflorum) and for strain ORS 1410 (Medicago sativa). In this study, strains ORS 2154, ORS 2133 and ORS 529, isolated from Lotus arabicus, were tested and were effective on their plant of isolation.

16S rRNA gene sequence analysis

Nearly full-length 16S rRNA gene sequences (1340 bp) were determined for strains ORS 1410, ORS 2133, ORS 529, ORS 2154, ORS 1401 and STM 354. The 16S rRNA gene phylogenetic trees constructed using three methods (ML, MP, NJ) resulted in the same groupings. Therefore, only the resulting ML tree is shown in Fig. 1. All isolates studied were placed unambiguously within the class Alphaproteobacteria and within the clade of the reference species of the genus Ensifer. The isolates were subdivided...
into three clusters (clusters A, B and C) supported by high bootstrap values and displaying high intra-cluster sequence similarity.

Cluster A (bootstrap value 97 %) included isolates ORS 1444, ORS 1410 and ORS 1407 and formed a homogeneous group (sequence similarity values 100 %) that was equally distant to \textit{Ensifer arboris}, \textit{Ensifer medicae} and \textit{Ensifer meliloti} (99.7 % sequence similarity).

Cluster B strains (ORS 529, ORS 2154 and ORS 2133) also shared identical 16S rRNA gene sequences and formed a clade grouping together with the genomovars of \textit{Ensifer adhaerens} (¢99.6–99.8 % sequence similarity) and reference strains of \textit{Sinorhizobium morelense} (99.0–99.2 % sequence similarity), although low bootstrap values were displayed for this group (35 %).

Strains ORS 1400\textsuperscript{T}, ORS 1401 and STM 354 formed a homogeneous group (Cluster C, 100 % interstrain gene sequence similarity values) supported by a bootstrap value of 85 %, and clustered in the vicinity of \textit{Ensifer terangae} (48 % bootstrap value and 98.8 % sequence similarity). The reference strain of ‘\textit{E. mexicanus}’ LMG 23932, which was described recently (Lloret \textit{et al.}, 2007), was related to these taxa. It was not included in our analyses, but formed a distinct subclade, sharing sequence similarity values of 97.8–98.8 % with clusters A, B and C and 99.5 % with strains of \textit{E. terangae}.

\textbf{Sequence analysis of individual genes}

The five housekeeping gene fragments examined in this study had different lengths: \textit{glnA} (977 bp), \textit{gltA} (681 bp), \textit{recA} (550 bp), \textit{atpD} (461 bp), \textit{thrC} (636 bp). Within each group...
In the _atpD_ gene tree topology, _Agrobacterium tumefaciens_ grouped with _S. morelense_, the three genomovars of _E. adhaerens_ and all strains from cluster B (100 % bootstrap), although this cluster had low bootstrap support (32 %). A second cluster (bootstrap value 43 %) contained all other species of the genus _Ensifer_ and clusters A and C (both clusters 100 % bootstrap value).

In the case of the _gltA_ gene, all strains of the genus _Ensifer_, together with clusters A, B and C, formed a single separate but poorly supported clade (bootstrap value 38 %). Within this clade, all strains from cluster C formed a significant subclade with _E. terangae_ (100 % bootstrap). Another well-supported subgroup (bootstrap value 98 %) was formed by all strains from cluster B and _E. adhaerens_ genomovar A.

In the _glnA_ gene analysis, all strains of the genus _Ensifer_ (and clusters A, B and C) were grouped together with high bootstrap support (75 %) and the same two significant subclusters that appeared in the _gltA_ gene tree topology were found (bootstrap values 100 and 95 % for cluster _C/E. terangae_ and cluster _B/E. adhaerens_ genomovar A, respectively).

In the _thrC_ gene tree, all strains of the genus _Ensifer_ together with clusters A, B and C, formed a single, separate but less supported clade (bootstrap value 48 %), in which the same two subclades were again distinguished (bootstrap values 100 and 98 % for cluster _C/E. terangae_ and cluster _B/E. adhaerens_ genomovar A, respectively). In the final single gene tree, the _recA_ tree, all _Ensifer_ strains and clusters A, B and C again formed a separate, but poorly supported, clade (bootstrap value 27 %). A well-supported subgroup (bootstrap value 95 %) was formed by all strains from cluster B and _E. adhaerens_ genomovar A.

Thus, cluster A formed a separate, well-supported group (bootstrap value 100 %) in all single gene phylogenies with identical sequences for the three strains (100 %). Cluster A was located at different positions when compared with the reference species of the genus _Ensifer_ and no significant clusterings were apparent. The highest observed sequence similarities between cluster A strains and the reference species ranged between 86.9 % with _Ensifer saheli_ and _Ensifer kostiensis_ for the _thrC_ gene and 94.3 % with reference strains _Ensifer meliloti_ and _Ensifer medicae_ for the _atpD_ gene.

Cluster B grouped with _E. adhaerens_, more specifically _E. adhaerens_ genomovar A, with high bootstrap values in all housekeeping gene analyses, except for the _atpD_ gene. Sequence similarities with _E. adhaerens_ genomovar A strains ranged between 96.4 % for _thrC_ to 98.4 % for the _recA_ gene. With both other genomovars, the range was lower at 90.5 % (_gltA_) to 95.3 % (_glnA_).

Strains _ORS 1400^T_, _ORS 1401_ and _STM 354_ formed the monophyletic Cluster C in all gene phylogenies (bootstrap value ranging from 98 to 100 %; sequence similarity values 100 % for all genes). This cluster displayed sequence similarities with its closest neighbour _E. terangae_ in the range of 92.2 % (for _recA_) to 96.4 % (for _gltA_). Cluster C together with _E. terangae_ was also related to the _E. mexicanus_ reference strain _LMG 23932_ in the analysis of three housekeeping genes ( _gltA_, _glnA_ and _recA_ (not included in Supplementary Fig. S1). The sequence similarities between Cluster C strains and _E. mexicanus_ ranged from 92.2 % for _recA_, over 93.7 % for _gltA_, to 95.8 % for _glnA_. Between _E. terangae_ and _E. mexicanus_ somewhat comparable interspecies sequence similarity values amounted to 96.6 % for _recA_, 93.5 % for _gltA_ and 96.2 % for _glnA_. This indicated that _E. mexicanus_, _E. terangae_ and the isolates from Cluster C all represented distinct genomic species.

**Sequence analysis of concatenated housekeeping genes**

The _atpD_ gene was found not to be congruent with the other genes (P<0.01) while all other housekeeping genes showed compatible phylogenetic signals (data not shown). A tree was constructed from the concatenated sequences of the four congruent genes _recA_, _glnA_, _gltA_, and _thrC_ (P<0.01) estimated by the ILD test (Farris et al., 1995). The ILD test values ranged from 0.0120 for _thrC_ and _recA_ to 0.6288 for _thrC_ and _glnA_. In view of the ILD test results, the aligned sequences for _recA_, _thrC_, _gltA_ and _glnA_ were concatenated and an alignment of 2704 nucleotides was obtained (comprising 1486 invariant sites, 237 variable but parsimony uninformative sites and 981 parsimony informative sites).

The combined analysis showed a significant cluster (bootstrap value 97 %), comprising all strains of the genus _Ensifer_ together with clusters A, B and C. The cluster was subdivided into two closely related subclusters. One well-supported subcluster (bootstrap value 100 %) includes _S. morelense_, the three _E. adhaerens_ genomovars and cluster B. As in most single gene analyses, cluster B formed a significant subclade together with _E. adhaerens_ genomovar A. The second major subcluster (bootstrap value 99 %) comprised all other strains of the genus _Ensifer_ and clusters A and C. As found with most single gene trees, Cluster C formed a significant subbranch with _E. terangae_ (bootstrap value 100 %). Cluster A strains _ORS 1407^T_ and _ORS 1444_ also formed a reliable cluster (bootstrap value 100 %), which was well separated from all other genomic species. Since a _thrC_ sequence could not be obtained for strain _ORS 1410_, this strain could not be included in the concatenated housekeeping gene analysis. As indicated, all mentioned clusters were supported by higher bootstrap values in the concatenated tree than in the single gene trees and are therefore more robust.
Table 2. DNA–DNA hybridization values (%) between species of the genus Ensifer and strains of cluster B isolated from Lotus arabicus in Senegal

<table>
<thead>
<tr>
<th>Strain</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ORS 529</td>
<td>100</td>
<td>99</td>
</tr>
<tr>
<td>2. ORS 2133</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>E. adhaerens genovar A LMG 10007</td>
<td>74</td>
<td>75</td>
</tr>
<tr>
<td>E. adhaerens genovar B R-7457</td>
<td>60</td>
<td>59</td>
</tr>
<tr>
<td>E. adhaerens genovar C LMG 20216T</td>
<td>47</td>
<td>47</td>
</tr>
</tbody>
</table>

DNA–DNA hybridization

To determine to which E. adhaerens genovar Cluster B could be designated, we performed DNA–DNA hybridizations between two strains (ORS 529 and ORS 2133) of this cluster and three E. adhaerens reference strains, representing the three different genovars (Willems et al., 2003) (Table 2). In agreement with the gene sequence analyses, strains ORS 529 and ORS 2133 displayed high DNA–DNA hybridization values (99%), proving that the strains represent the same genomic species. The hybridization values of strains ORS 529 and ORS 2133 with E. adhaerens genovars B and C strains were below 70%, indicating that the two strains from Cluster B are not members of E. adhaerens genovar B or genovar C. However, DNA–DNA relatedness values between Cluster B strains and the E. adhaerens genovar A reference strain exceeded 70% which confirmed the assumption drawn from MLSA that Cluster B strains are strains of E. adhaerens genovar A.

nodA gene sequencing

The nodA gene was sequenced in strains representing both of the different clusters and host plants of isolation: two cluster C strains (ORS 1400T/Argyrolobium uniflorum and STM 354/Medicago sativa), one cluster B strain ORS 529 (isolated from Lotus arabicus, as were the two other strains of the cluster) and two Cluster A strains (ORS 1444/Argyrolobium uniflorum and ORS 1410/Lotus creticus). The nodA gene could not be amplified in strain ORS 1407T. The analysis of nodA gene sequences by the NJ method (Fig. 2) shows that strains isolated from Argyrolobium uniflorum, Lotus arabicus and Medicago sativa formed a separate cluster. This cluster was grouped with the nodA gene of E. meliloti and the E. medicae branch with high bootstrap support. In this study, the taxonomic position of nodA gene sequences of isolates from Argyrolobium uniflorum is reported for the first time.

Numerical analysis of auxanographic tests

All strains were tested for use of 99 substrates as the sole carbon source for growth, using API 100 galleries (see Supplementary Table S2 in IJSEM Online). Some discriminative features could be observed. In contrast to their closest phylogenetic neighbours, E. medicae LMG 19920T, E. meliloti LMG 6133T and E. arboris LMG 14919T, strains of Cluster A, did not grow on xylitol or DL-glycerate. Results for other substrates were variable between members of this cluster. About 28 substrates were utilized differently by strains ORS 1444, ORS 1410 and ORS 1407T and 13 substrates were only metabolized by strain ORS 1407T (dulcitol, D-lyxose, 1-O-methyl x-D-glucopyranoside, 3-O-methyl D-glucopyranose, D-gluconate, L-histidine, succinate, fumarate, ethanolamine, DL-β-hydroxybutyrate, L-aspasparte, L-alanine and propionate).

Cluster B strains used putrescine in contrast to the three genovars (A, B, C) of E. adhaerens. Seven substrates [(+)-l-sorbose, dulcitol, D-tagatose, citrate, 5-ketod-glucopionate, D-gluconate and glutarate] were not metabolized by members of Cluster B and were used diversely among the three genovars of E. adhaerens.

Cluster C, which appeared to be close to E. terangae in phylogenetic analyses, differed from this species in the test results for ten substrates. These ten substrates were assimilated by E. terangae LMG 7834T and not by any member of Cluster C. However, malonate was metabolized by all strains of Cluster C and not by E. terangae LMG 7834T.

Auxanographic data were analysed numerically (data not shown). Clusters A and C formed two related clusters but were separate from the other species of the genus Ensifer. Strains of Cluster B were included in E. adhaerens as observed in the genotypic analysis, with strain ORS 2154 close to Ensifer adhaerens genovar B strain R-7457.

DISCUSSION

In this study, we performed MLSA to characterize nine strains from Tunisia and Senegal which had been isolated from the root nodules of several legumes. Previously, six of them had been partially characterized by ARDRA, SDS-PAGE and 16S rRNA gene sequencing (Zakhia et al., 2004). Genes for MLSA were chosen according to previous taxonomic and phylogenetic studies: recA (recombinase A) and atpD (ATP synthase beta subunit) were used in a phylogenetic study of the genera Agrobacterium and Rhizobium (Gaunt et al., 2001); glnA (glutamine synthase) was used by Turner & Young (2000) and glnA (citrate synthase) was used by Hernandez-Lucas et al. (2004). In a study evaluating the taxonomic potential of ten housekeeping genes, Martens et al. (2007, 2008) found these genes and additionally thrC (threonine synthase) useful for identification and for the inference of phylogenetic relationships of species of the genus Ensifer. Phylogenetic analyses of the five housekeeping genes were compared with the 16S rRNA gene tree using three algorithmic methods (ML, MP and NJ). We completed the characterization with auxanographic tests and nodA gene sequences for biovar determination.

The nodA gene is a key gene for the establishment of symbiosis; it is present as a single copy and has a size of
590–660 bp (Moulin et al., 2004). Except for some photosynthetic species of the genus *Bradyrhizobia* (Giraud et al., 2007), all bacteria described thus far as being capable of establishing a symbiosis with legumes harbour nodulation (*nod*) and nitrogen fixation (*nif*, *fix*) genes. This opens up the possibility of screening for these genes as an alternative to nodulation tests. The encoded *ab*-unsaturated Nod factor protein is a key enzyme for Nod factor synthesis, transferring an acyl chain on a chitin oligomer, resulting in a biologically active NodA molecule. The sequence of the *nodA* gene provides information regarding the symbiotic characteristics of the rhizobium and may be predictive for the type of Nod factor produced and for host specificity (Lortet et al., 1996; Debelle et al., 2001). As the *nodA* gene is frequently plasmid-borne and has a propensity towards lateral gene transfer, the sequence of the gene is usually indicative of the biovar rather than the species affiliation (Haukka et al., 1998).

All phylogenies based on the three algorithmic methods unambiguously placed the nine root nodule strains within the Class *Alphaproteobacteria*, in the *Ensifer* clade. The nine strains were subdivided into three monophyletic clusters with a high intra-species sequence similarity of 100% for all gene sequence analyses.

Cluster A formed a separate cluster with variable positions relative to the reference species of the genus *Ensifer* in the phylogenies based on the 16S rRNA gene and all of the housekeeping genes. In the concatenated gene sequence analysis, Cluster A formed a single, separate cluster which was clearly distinguished from all of the other clusters. The sequence similarities between Cluster A strains and strains of related reference species supported this segregation. The gene sequence similarity values between members of Cluster A and the other genotypic clusters ranged between 82 and 94.3% in the different gene analyses (see Supplementary Fig. S1 in IJSEM Online). *Sinorhizobium chiapanecum*, related to *E. terangae*, was reported during the process of revision of the present work (Rincón-Rosales et al., 2009). The available recA and 16S rRNA gene sequences of this species were
retrieved and it was observed that both genes place ‘S. chiapanecum’, E. teranga and ‘E. mexicanus’ in a cluster (96.3–96.9 % (recA) and 99.3–99.7 % (16S RNA gene) internal similarity values), away from Cluster A, with interspecies similarity values ranging from 92.2–92.6 % (recA) and 98.4–98.9 % (16S RNA gene).

These values illustrate a clear gap in sequence similarity levels within Cluster A strains and between Cluster A and other reference strains, corresponding to the inter- and intra-species sequence similarity value gap observed by Martens et al. (2007, 2008) in their comparison of MLSA and DNA–DNA hybridization data. This indicates that Cluster A strains are well distinguished from the reference species of the genus Ensifer and therefore represent a new genomic species.

The auxanographic tests also confirmed that members of Cluster A represented a novel species of the genus Ensifer. The nodA gene sequence of a representative strain of Cluster A, strain ORS 1444, was examined in order to obtain some information regarding the Nod factor of Argyrolobium uniflorum microsymbionts. The results showed that this strain grouped in a separate cluster close to the E. meliloti/E. medicae nodA group. In view of our phenotypic and genotypic results and those obtained in previous studies (Zakhia et al., 2004), we propose a novel species of the genus Ensifer for the Cluster A strains with the name Ensifer numidicus sp. nov. Strain ORS 1407T is designated as the type strain.

In a parallel and independent work, Mahdhi et al. (2008) described root nodule isolates from Argyrolobium uniflorum in the same region of Tunisia. Among the most effective strains (potential candidates for inoculant production), 13 formed a homogeneous and separate group on the Ensifer branch when investigated by 16S rRNA gene PCR-RFLP. Three representative strains of this group were sequenced, and they shared identical nucleotide sequences with strain ORS 1444. This group may thus be considered as members of E. numidicus sp. nov.

Cluster B formed a monophyletic group in all phylogenies and grouped separately with E. adhaerens genovar A in the majority of the gene phylogenies. In the 16S rRNA gene tree, Cluster B strains were grouped together with the genomovars of E. adhaerens and S. morelense. In the single and concatenated housekeeping gene analyses, Cluster B grouped with E. adhaerens genovar A, supported by high bootstrap values. The analysis of the gene sequence similarities placed Cluster B in the gap between intra- and interspecies sequence similarities values, confirming that the members of Cluster B belong to the species E. adhaerens but with an uncertain genomovar position. Following the recommendation of Willems et al. (2003) for genomovar discrimination, DNA–DNA hybridization experiments were conducted. The results confirmed that members of Cluster B were different from genomovars B and C and belonged to E. adhaerens genovar A. Phenotypic and genotypic analyses clearly indicated that Cluster B grouped with E. adhaerens genovar A. All strains from Cluster B nodulate the plant from which they were isolated, Lotus arabicus. This is in contrast to all recognized members of E. adhaerens (Casida, 1982; Willems et al., 2003), even though Rogel et al. (2001) demonstrated that they may acquire nodulation capacity upon the introduction of symbiotic plasmids.

Cluster C formed a separate monophyletic group in all phylogenies, essentially in the E. teranga/E. mexicanus clade. A similar grouping was obtained using the four congruent genes (glnA, gltA, thrC and recA) in individual gene analyses and in the concatenated gene trees. Sequence similarity values indicated a clear gap between the similarity levels within Cluster C strains and between strains of Cluster C and other reference strains. Values between Cluster C strains and the reference strains were comparable with those at the interspecies level found by Martens et al. (2008) in their study comparing MLSA and DNA–DNA hybridization data. This indicated that Cluster C was distinct and formed a separate genus. In addition, the numerical analysis of auxanographic results confirmed that Cluster C was separate from all species of the genus Ensifer. In contrast to their closest phylogenetic neighbour E. teranga LMG 7834T, Cluster C strains were able to utilize malonate, but not p-hydroxybenzoate. Two representative strains of Cluster C, ORS 1400T and STM 354, isolated from two different legumes, exhibited nodA gene sequences which grouped together (bootstrap value 99 %) with those from Cluster A and B, in the E. meliloti and E. medicae branch.

The combination of the results obtained previously by Zakhia et al. (2004) and in this study by MLSA and additional 16S rRNA and nodA gene sequence analyses and auxanographic tests indicate that Cluster C represents a novel genospecies in the E. teranga/E. mexicanus clade, for which we propose the name Ensifer garamanticus sp. nov., with strain ORS 1400T as the type strain.

Additional DNA–DNA hybridization experiments to establish the separate species status of clusters A and C were not performed due to the clarity of the MLSA evidence. In this study, the highest gene sequence similarity level with other species of the genus Ensifer obtained for Cluster A was 94.3 % (atpD). For Cluster C, this value was 96.4 % (gltA). In view of the finding that strains of Cluster B, with housekeeping gene sequence similarity levels of 96.4–98.4 % with E. adhaerens genomovars B and C, produced DNA–DNA relatedness levels below the species threshold (Table 2, 47–60 %), even lower DNA–DNA binding between strains that have lower gene sequence similarity levels in housekeeping genes would be expected. This information is also in line with a previous elaborate comparison of housekeeping gene sequences and DNA–DNA hybridization data (Martens et al., 2008) where it was shown that MLSA of selected housekeeping genes can accurately predict relationships between closely related organisms. It is concluded Clusters A and C represent two novel species of the genus Ensifer species, for which the names of Ensifer numidicus and Ensifer garamanticus are proposed, respectively.
Description of *Ensifer numidicus* sp. nov.

*Ensifer numidicus* (nu.mi’di.cus. L. masc. adj. *numidicus* pertaining to the country of Numidia, Numidian, the Roman denomination of the region in North Africa from which the majority of the organisms were isolated).

Short, aerobic, Gram-negative, non-spore-forming rods. Strains grow on yeast-mannitol medium on which they form white slightly mucous colonies after 3 days at 28 °C. At the molecular level, this species can be differentiated by phylogenetic analysis based on several housekeeping genes (*recA*, *glnA*, *gltA*, *thrC* and *atpD*) and 16S rRNA gene sequencing. Its phylogenetic neighbours are *E. medicae*, *E. meliloti* and *E. arboris*. Detailed phenotypic features for strains of the species are given in Supplementary Table S2. Can be distinguished from *E. medicae*, *E. meliloti* and *E. arboris* by the differential use of a combination of growth substrates. Does not grow on xylitol or DL-glycerate.

Additional strains tolerate pH values of 6–9, 1–2 % (w/v) and strain STM 4036 is the most efficient. The type strain is ORS 1400T (Cluster 1) ampicillin (100) and streptomycin (100), but not by any member of *Ensifer garamanticus* (Cluster C). Malonate is metabolized by all strains and not by *E. terangae* LMG 7834T.

The type strain is ORS 1400T (=LMG 24690T=CIP 109916T). The DNA G+C content of the type strain is 62.4 % (HPLC).

A group of 13 *Sinorhizobium* sp. strains reported by Mahdhi *et al.* (2008), AB1, AB3, AB5, AB6, AB10, AM1, AM2, AM3, AS1, AS2, AS3, AS4, AS5, may be considered as belonging to *Ensifer numidicus* and are included as members of this species. These strains were isolated independently from *Argyrolobium uniflorum* in South Tunisia. The strains form a homogeneous group by PCR-RFLP of the 16S rRNA gene. Three representative strains of this group STM 4034 (=AB1), STM 4036 (=AB3), STM 4039 (=AS1), share identical 16S rRNA gene sequences with strains of *E. numidicus* as isolated in this study. The three latter strains form the most effective symbiosis with *Argyrolobium uniflorum* and strain STM 4036 is the most efficient. The additional strains tolerate pH values of 6–9, 1–2 % (w/v) NaCl and grow at 40 °C, but not at 45 °C. They are sensitive to (µg ml⁻¹) ampicillin (100) and streptomycin (100), but resistant to kanamycin (100) and nalidixic acid (100).

Description of *Ensifer garamanticus* sp. nov.

*Ensifer garamanticus* (ga.ra.man’ti.cus. L. masc. adj. *garamanticus*, pertaining to Garamante, Garamantian, the country of Garamantes, from which the strains were isolated).

Short, aerobic, Gram-negative, non-spore-forming rods. The strains form white and slightly mucous colonies on YMA after 48–72 h incubation at 28 °C.

Strains efficiently nodulate *Argyrolobium uniflorum* (ORS 1400T, ORS 1401) and *Medicago sativa* (STM 354). Can be distinguished from other species by phylogenetic analysis based on several housekeeping genes (*recA*, *glnA*, *gltA*, *thrC* and *atpD*) genes and 16S rRNA gene sequencing. Malonate is used by all strains for growth, but is not used by the closest phylogenetic neighbour *E. terangae*. Detailed phenotypic features for all strains are given in Supplementary Table S2. In the API 100 gallery, ten substrates are used for growth by *E. terangae* LMG 7834T but not by any member of *Ensifer garamanticus* (Cluster C). Malonate is metabolized by all strains and not by *E. terangae* LMG 7834T.

The group formed by *Sinorhizobium* sp. strains STM 4015, STM 4016, STM 4027, STM 4031, STM 4032, isolated from *Genista saharae* in South Tunisia by Mahdhi *et al.* (2007), share identical 16S rRNA gene sequence with the type strain of *E. garamanticus*. They may be thus be considered as members of this species. These strains generally tolerate high temperatures (40 °C), high pH values (7–12) and high NaCl concentrations (1–4 %, w/v) for growth. They nodulate their plant of isolation, so that *Genista saharae* may also be considered as a potential host plant of *Ensifer garamanticus*.

ACKNOWLEDGEMENTS

The authors are grateful to F. Boukhatem, B. Dreyfus, A. Galiana, L. Moulin, G. Béna, B. Hamed Bey, L. Ameziane and L. Bouchentouf for helpful discussions. C. M., F. Z., A. B. and A. S. are indebted to the Département Soutien-Formation (DSF) of the Institut de Recherche pour le Développement (IRD), France, for a PhD fellowship and ESCD grants and M. Mosbah for the AUF grant. A. W is grateful to the Fund for Scientific Research – Flanders for a postdoctoral fellowship. This work was supported by the Commission of the European Communities RT-program contract BACDIVERS QLRT-2001-02097.

REFERENCES


cum sp. nov. which has common symbiotic genes with Sinorhizobium mexicanum. FEMS Microb Ecol 67, 103–117.


