Characterization and genus identification of the rhizobial symbiont from *Caragana arborescens* in Western Canada

Judicaël Moukoumi[2], Russell K. Hynes[1, 4], Timothy J. Dumonceaux[1], Jennifer Town[1], and Nicolas Bélanger[3]

[1]Agriculture and Agri-Food Canada, Saskatoon Research Centre, Saskatoon, SK;

[2]Department of Soil Science, University of Saskatchewan, Saskatoon, SK;

[3]Centre d'étude de la forêt, UER Science et technologie, Téluq, Université du Québec à Montréal, QC

[4]Corresponding author: Russell K. Hynes, Agriculture and Agri-Food Canada, Saskatoon Research Centre, 107 Science Place, Saskatoon, SK, S7N OX2

Tel: 306-956-7638; Fax.: 306-956-7247

E-mail address: russell.hynes@agr.gc.ca

Sequences determined in this study were deposited into GenBank with the following accession numbers: for 16S: JF819654-JF819658; for recA: JF819659-JF819661; for cpn60: JF819662-JF81966 and KC521416-KC521435.

Abstract

Naturally-occurring nitrogen-fixing symbionts from root nodules of caragana, (*Caragana arborescens*), growing in central Saskatchewan were isolated following surface sterilizing caragana root nodules, squashing and spreading the contents on yeast-extract mannitol medium. The symbiotic nature of the strains was confirmed following inoculation onto surface sterilized *C. arborescens* seed in a gnotobiotic Leonard jar system. The *Rhizobium* isolates from *C. arborescens* root nodules are intermediate in growth rate (g) (mean of 5 isolates g = 6.41) as compared to fast growers, *Rhizobium leguminosarum* NRG457 (g: 4.44), *R. tropici* 899 (g: 3.19) and *Sinorhizobium meloliti* BALSAC (g: 3.45) but faster than the slow grower *Bradyrhizobium japonicum* USDA 110 (g: 13.86) and similar to *Mesorhizobium amorpheae* (g: 7.76). Nitrogen derived from fixation by measuring changes in $\delta^{15}N$ natural abundance in plant tissue confirmed the effectiveness of the strains; approximately 80% N_2 from fixation. Strain identification was carried out by determining the sequences of three genes, 16S rRNA-encoding genes, cpn60, and recA. This analysis determined that the symbiotic partner of Canadian *C. arborescens* belong to genus Mesorhizobium and seem more related to *M. loti* than to previously described Caragana symbionts like *M. caraganae*. This is the first report of *Mesorhizobium* sp nodulating *C. arborescens* in western Canada.

Keywords: *Mesorhizobium*, *Caragana arborescens*, nitrogen fixation, cpn60

Introduction

Caragana arborescens, commonly referred to as caragana or Siberian pea-shrub, is a woody legume in the family *Fabaceae* and native to north-eastern Europe and central Asia. *Caragana arborescens* was introduced to North America in 1752 and is the favored species in western Canadian prairie for shelterbelts for seeded fields and farm homes (Dietz *et al.* 2008). Its endurance to cold and drought conditions, long life span, ability to fix nitrogen, and height of 4-5 m made it an acceptable choice. *C. arborescens* tolerance to cold was reported by Hensley and Carpenter (1979). The lowest activation temperature of nodules from the several legumes tested were those of *C. arborescens*, i.e. 3-5 °C.

Symbiotic interactions involving leguminous plants and rhizobia result in the differentiation of root or stem tissue into nodules where these bacteria reduce nitrogen gas to nitrogenous compounds which enter the plant's metabolic machinery. Early studies by Beijerinck suggested that the root nodule isolates from *C. arborescens* were similar to the slow growing rhizobia from cowpea, soybean or lupin (Allen and Allen 1939). Gregory and Allen (1953) reported markedly different growth rates and other physiological differences from fourteen rhizobial isolates from root nodules of *C. arborescens*. Plant cross-inoculation groupings placed the rhizobia that nodulate *C. arborescens* in the *Lotus* group (Jarvis et al. 1982). Further refinements in the nomenclature of *Rhizobium* spp. resulted in the transfer of the *C. arborescens* symbiont to *Mesorhizobium* gen nov (Jarvis et al. 1997). In a phylogenetic tree based on the 16S rRNA-encoding gene, Chen et al. (2008) reported that *Mesorhizobium* sp CCBAU 11226, isolated from *C. arborescens* roots, was closely-related to *M. tianshanense* USDA 3592 (>99.3% sequence identity at the 16S rRNA-encoding locus). To our knowledge, the

identity of the rhizobia species nodulating *C. arborescens* in Canada has not been studied or reported.

In 2007, field plots were established in Saskatchewan, Canada, with pure and mixed plantations of willow (*Salix miyabeana*) and *C. arborescens*, to examine the potential of species association as a strategy for sustained productivity and reduced CO_2 footprint (Moukoumi et al. 2012). Examination of field grown *C. arborescens* revealed that the roots were nodulated, whereas it was determined with the $\delta^{15}N$ natural abundance method that as much as 30% of total nitrogen in *C. arborescens* leaves on poor soils originated from the atmosphere (i.e. N obtained from biological N_2-fixation). These results suggest that the soil had a competent population of rhizobia. The objectives of this study were to: 1) isolate naturally occurring nitrogen-fixing symbionts from *C. arborescens* root nodules, 2) confirm N_2 fixation activity by the $\delta^{15}N$ natural abundance method, and 3) apply 16S rRNA, *recA*, and *cpn*60 sequence analysis to identify the rhizobial symbiont from *C. arborescens*.

Materials and methods

Isolation of putative *Rhizobium* strains from *C. arborescens* root nodules

Nodules were collected from the roots of *C. arborescens* plants carefully extracted from field plots in three former agricultural fields in central Saskatchewan; two associated with the University of Saskatchewan, Saskatoon and a third in Harris, Sk, 90 km southwest of Saskatoon. Field site characteristics are reported in Moukoumi et al. (2012). Thirty root nodules were surface sterilized by immersing and with shaking in 95% ethanol for 20 sec, 5 min in 20% sodium hypochlorite (commercial Clorox bleach diluted 5 × in sterile dH_2O) and rinsed 7 times in sterile dH_2O. Surface sterilized nodules

were squashed, using flame-sterilized forceps, and the nodule contents were spread over the surface of yeast extract-mannitol agar (YMA); g/1000mL dH_20: mannitol, 10; K_2HPO_4, 0.5; $MgSO_4.7H_2O$, 0.2; NaCl, 0.1; $CaCO_3$, 0.01; yeast extract (Difco, Detroit) 0.5; agar, 15g, in Petri dishes and incubated at 23 °C (Weaver and Frederick 1982). After 4 d colonies of putative *Rhizobium* strains were restreaked on YMA and incubated at 23°C. This process was repeated 2 additional times to insure purity before inoculating YM broth (Weaver and Frederick 1982). After 4 days of incubation with shaking at 150 rpm at 23 °C, the cultures were streaked onto YMA and nutrient agar (Difco, Detroit) for additional purity checks. Five putative isolates named Laura, Harris, Sovereign, Thrasher and Kernan were retained and preserved in a -80 °C freezer.

Generation time (g) studies – comparison to reference Rhizobiaceae isolates

One hundred mL of YM broth in 250 mL baffled flasks was inoculated with 0.2 mL of stationary phase cultures of each isolate and reference isolates from other genera of rhizobia. All isolates were incubated at 23 °C with shaking at 150 rpm. Population densities of rhizobial isolates Laura, Harris, Sovereign, Thrasher and Kernan, *Sinorhizobium meloliti* BALSAC, *Rhizobium leguminosarun* NRG457 (Agriculture Agri-Food Canada), *R. tropici* 899, a.k.a. CIAT899 (University of Minnesota), *Bradyrhizobium japonicum* USDA 110 (kindly provided by Guoping Yang, BeckerUnderwood Inc.), and *M. amorpheae* (DSM No. 21831) were determined by serial dilution and plating on YMA. Generation time (g) was determined by first calculating the number of generations (n) that occurred during the logarithmic phase of growth (N and No: final and original population densities) from the equations:

$$n = (\log_{10} N - \log_{10} No)/\log_{10} 2$$

Generation time was calculated from the equation: g = t/n

where t is elapsed time from N and No. Mean generation times were calculated from triplicated growth experiments with each isolate.

Leonard jar studies

C. arborescens seeds, provided by the Agroforestry Development Centre (Indian Head, Sk.), were surface sterilized using a 95% ethanol solution for 25 sec followed by a 20% sodium hypochlorite (as above) solution for 20 min. Following 7 sterile dH_2O rinses, the seed was placed in a laminar flow hood to dry. The seeds were then pre-germinated on moistened sterile filter paper and two seedlings were later transplanted into sterilized Leonard jar assemblies (Vincent 1970). The upper part of the assembly, 300 mL, was filled with acid washed sand and the lower part of the assembly, 500 mL, contained N-free plant nutrient solution (Hoagland and Arnon 1950). Putative *Rhizobium* isolates grown in YM broth and incubated on a reciprocal shaker (150 rpm) at 23°C for 4 days were applied in 1 mL aliquots to each seedling. Treatments included *C. arborescens* seedlings inoculated individually with strains Laura, Harris, Sovereign, Thrasher and Kernan and the non-inoculated control. *C. arborescens* was grown in a Conviron plant growth room with day/night temperatures of 23 and 20°C and 10 hours of light and 14 hours of darkness. After 90 days, *C. arborescens* shoots were cut from the roots, dried at 65ºC for 3 days, weighed and finely ground before $^{15}N/^{14}N$ and total C and N analysis.

Foliage sampling and analyses of nitrogen (N), carbon (C) and N isotopes

The oven-dried samples were coarsely ground prior to leaf N and C determination by dry combustion and infrared detection using a Leco CNS 2000 Analyzer (LECO Corporation, St. Joseph, MI). Sub-samples were then further ground (<60μm) and 1.0 ±

0.15 mg was analyzed for $^{15}N/^{14}N$ ratios using a continuous flow isotope ratio mass spectrometer interfaced with a RoboPrep Sample Converter (Europa Scientific, Crewe, UK). The working standard for $^{15}N/^{14}N$ was pea grain (atom $\%^{15}N$ content = 0.36726). The deviation of the sample $^{15}N/^{14}N$ ratio from that of the atmosphere ($\delta^{15}N$) was calculated as follows:

$$\delta^{15}N = ((R_{sample} / R_{standard}) - 1) \times 1000$$

Molecular phylogenetic studies of the putative *Rhizobium* strains

DNA was extracted from 200 µL of an overnight culture of each isolate using a modified bead-beating protocol as described by Hill et al. (2005a). Final DNA pellets were dissolved in 100 µL of 10 mM Tris-Cl pH 8.0 with 1mM EDTA. Two µL of each extract were used as a template in a PCR with primers F1 and R2 (Dorsch and Stackebrandt 1992), which amplify nucleotides 11-536 of the 16S rRNA-encoding gene (numbering according to *E. coli* sequence). *recA* gene fragments were amplified using primers *recA*63F and *recA*555R as described by Gaunt et al. (2001). 16S rRNA and *recA* amplicons were sequenced directly using the amplification primers. For amplification of the *cpn60* universal target (*cpn60* UT) (Hill et al. 2004), a cocktail of primers H279/H280 and H1612/H1613 was used as described (Hill et al. 2005b) with an annealing temperature of 50°C. Amplicons were generated using combinatorial enhancer solution (CES) according to Ralser et al. (2006) in order to improve the representation of all copies of *cpn60* from each isolate. *cpn60* UT PCR products were ligated into pGEM-T Easy PCR product cloning vector (Promega, Madison, WI) according to the manufacturer's directions prior to sequencing, and a total of 48 *cpn60* UT clones was

sequenced for each isolate. Sequences were analyzed using pregap v1.5/gap v4.10 (Staden package) to generate contigs that were manually checked for quality and trimmed of primer sequences.

Taxonomic analysis. 16S rRNA, *recA*, and *cpn60* UT sequences were aligned using the default parameters of clustalw (Thompson et al. 1994) to determine sequence identities. Phylogenetic analysis was performed using the neighbor joining method (Saitou at al. 1987) with bootstrapping using 500 replicates. Phylogenetic trees were generated using MEGA4 (Tamura et al. 2007).

Whole-genome alignments and determination of average nucleotide identity (ANI) by MumMer (ANIm), ANI by BLAST (ANIb), and tetranucleotide frequency correlations (Tetra) were determined using JSpecies as described by Richter and Rossello-Mora (2009). *Mesorhizobium* genomic sequences were obtained from NCBI (*M. ciceri* biovar biserrulae WSM1271, *M. australicum* WSM2073, *M. opportunistum* WSM2075, and *M. loti* MAFF303099) or from the Joint Genome Initiative Genome Portal (http://genome.jgi-psf.org/) for *M. loti* WSM1293, *M. loti* R7A, *M. loti* R88b, *M. loti* NZP2037, *M. loti* R7A, and *Mesorhizobium* sp. BNC1. Sequences of all *cpn60* UT copies that were not available in GenBank or in cpnDB were retrieved from the draft genomes using BLASTp with the *M. loti cpn60* UT peptide sequence (cpnDB ID b4797) as the query.

Data analysis

Means of triplicate samples of each treatment plus/minus standard errors are reported.

Results

Isolation of putative *Rhizobium* strains from *C. arborescens* root nodules

Rhizobium-like colonies were randomly selected from YMA plates previously streaked with the contents from squashed root nodules of *C. arborescens.* Visual characterization of the five putative isolates of *Rhizobium* was as follows: colonies were opaque white, raised and round and 0.5 to 1 mm in diameter after 6 days growth at 23°C.

Generation time (g) studies – comparison to reference Rhizobiaceae isolates

Generation times (g) of the five rhizobia isolates were compared to fast, intermediate and slow growing Rhizobiaceae isolates (Table 1). The isolates from *C. arborescens* root nodules were intermediate in growth rate as compared to the fast growing *R. leguminosarum* NRG457, g: 4.44; *R. tropici* 899, g: 3.19 and *S. meloti* BALSAC, g: 3.45 and slow growing *B. japonicum* USDA110, g: 13.86 (Table 1). The mean g for the *C. arborescens* root nodules isolates was 6.41, similar to that of an intermediate growing *M. amorpheae*, g: 7.76. Acid production by the *C. arborescens* root nodules isolates, *S. meloliti* BALSAC, *R. tropici* 899, *R. leguminosarum* NRG457 and alkali production by *B. japonicum* USDA 110 in YM broth were noted (data not presented).

Leonard jar studies

Nodules with red interiors developed on *C. arborescens* roots, indicative of the presence of leghemoglobin and effective nitrogen-fixing activity. *C. arborescens* root nodules were elongate, approximately 3 by 1.5 mm and typically there were 20 -30 nodules per plant. All inoculated plants growing in a N-free medium (Hoagland and Arnon 1950) had dark green leaves, suggesting that these plants were supplied N by symbiotic nitrogen fixation. Non-inoculated *C. arborescens* plants displayed stunted

182 growth, the roots did not have nodules and the leaves were yellow, indicative of N
183 deprivation.

184 Nitrogen and C concentrations and N isotopes data generated from the leaves
185 collected from the nodulated and non-nodulated *C. arborescens* are presented in Table 2.
186 Nodulated plants had on average 4.9 and 3.2 times higher percent N and C than the non-
187 nodulated (non-inoculated) plants (Table 2). $\delta^{15}N$ varied from -1.84 to -1.44 for the
188 nodulated plants. $\delta^{15}N$ of the non-inoculated plants was 0.8 indicative that all of the N
189 was derived from the seed.

190 **Molecular taxonomic studies of the putative *Rhizobium* strains**

191 Comparison of the 16S rRNA-encoding genes of the five isolates to the
192 corresponding reference strains suggested that the isolates formed two groups within the
193 genus *Mesorhizobium* (Figure 1). The 16S rRNA-encoding sequences of isolates Laura
194 and Kernan were closely related to *M. amorphae*, *M. australicum* and *Mesorhizobium*
195 isolates CCBAU 11231, CCBAU 11214, and CCBAU 11208 identified by Chen et al.
196 (2008). Isolates Harris, Sovereign and Thrasher had identical 16S rRNA-encoding
197 sequences that were closely related to a group of *Mesorhizobium* spp. including *M.*
198 *caraganae*, *M. loti*, and CCBAU 11226 (Chen et al. 2008). Although the 16S rRNA-
199 encoding sequences of all the isolates and reference strains formed distinct clustering
200 patterns that were consistent with those described by Chen et al. (2008), the sequences
201 were highly similar to one another, sharing sequence identities > 97% (data not shown).

202 We also examined two protein-encoding genes, *cpn60* and *recA*, for determining
203 the taxonomic affiliations of the isolates. Since *cpn60* is present in multiple copies in the
204 root-nodulating *Alphaproteobacteria* (Lund 2009), we examined all of the currently

205 available complete genomes for *Mesorhizobium* spp. Representative results are shown in
206 Table 3. Determination of the genomic identity parameters described by Richter and
207 Rossello-Mora (2009) revealed that most of these sequenced strains can be identified as
208 separate species, consistent with their current taxonomic designations, although their
209 close phylogenetic relationship is reflected in the generally high Tetra scores that were
210 observed (Table 3). Among the 5 *M. loti* genomes that were compared, however, only
211 one pair, *M. loti* MAFF303099 and *M. loti* R7A, had sufficiently similar genomic
212 sequences to be considered the same species according to the metrics suggested by
213 Richter and Rossello-Mora (2009) (Table 3). Turner et al. (2002) suggested that strain
214 MAFF303099 has been mis-assigned to *M. loti* and should instead be classified as *M.*
215 *huakuii* biovar loti. Unfortunately the lack of a complete genome sequence for the latter
216 strain precludes its inclusion in this whole-genome sequence analysis. All of the other
217 pairs of *M. loti* genomes fell below the sequence similarity cutoffs defined by Richter and
218 Rossello-Mora (2009) for grouping into the same species (Table 3 and data not shown).
219 Among all of the strains, the number of copies of *cpn60* varied, from as few as 2 copies
220 in *Mesorhizobium* spp. BNC1 to as many as 5 in *M. loti* MAFF303099. Some of these
221 copies shared a very high sequence identity (up to 1.0), despite being from distinct
222 species as shown by JSpecies (Table 3). Other *cpn60* copies were highly distinct from
223 one another, with pairwise *cpn60* UT sequence identities of 0.74-0.76.
224 Clones corresponding to the *cpn60* UT were retrieved for all isolates. The number
225 of distinct *cpn60* copies that were retrieved for each isolate varied from 3 (isolate
226 Sovereign) to 7 (isolate Thrasher) (Table 4). In all cases, the *cpn60* copies from a single
227 isolate were most closely related to 2 or 3 different *Mesorhizobium* spp., with sequence

identities ranging from 87.2% to 98.7% (Table 4). Phylogenetic analysis of all of the *cpn60* copies observed from all isolates revealed that several of the *cpn60* sequences from the isolates were identical or nearly identical to one another (supplemental Figure 1[1]). Only one *cpn60* sequence was observed in all of the strains (Sovereign *cpn60*-2, Laura *cpn60*-4, Harris *cpn60*-3, Thrasher *cpn60*-2, Kernan *cpn60*-2) This *cpn60* sequence shared ~94% sequence identity with 3 of the *cpn60* copies from *M. loti* MAFF303099. A few other *cpn60* sequences were shared by 2 or 3 isolates (supplemental Figure 1[1]). Generally the *cpn60* sequences of the *C. arborescens*-nodulating isolates described here tended to cluster together (Supplemental Figure 1[1]).

Finally, to place these isolates in the context of reference strains of *Mesorhizobium* as well as the isolates reported by Chen et al. (2008) and Yan et al. (2007) the sequences of *recA* fragments for isolates Sovereign, Thrasher and Kernan were determined. These sequences were all identical to one another and phylogenetic analysis placed these isolates in a clade with *M. loti* LMG 6125 as well as *M. ciceri* (Figure 2). The isolates reported here appeared to be distinct from the strains isolated by Chen et al. that form a symbiotic relationship with *C. arborescens* (Chen et al. 2008, Yan et al. 2007).

Discussion

Mesorhizobium isolates were collected from *C. arborescens* root nodules grown in soil containing naturally occurring and effective rhizobia. This study is part of a larger study examining species association and nitrogen contribution from N-fixing *C. arborescens* to *Salix miyabeana* (Moukoumi et al. 2012). Wind carrying soil laden with

[1] Supplemental data are available with this article through the journal web site at

effective rhizobia for caragana from near-by (2 km or more) rows of caragana is the likely source of inoculum since these agricultural sites had never been planted with caragana. The origin of this rhizobial population is unknown, but presumably was introduced to Saskatchewan soils from northern Asia with *C. arborescens* over 1 century ago (Cram 1969). The importance of an effective rhizobial symbiont for survival and establishment of *C. arborescens* was highlighted when *C. arborescens* was introduced in New Zealand for revegetation of semiarid areas (Wills 1982). Nodulated by ineffective rhizobia, *C. arborescens* was chlorotic, slow to establish and displayed low survival rates (Wills 1982). The long history of cultivating *C. arborescens* in Saskatchewan to mitigate wind erosion of soils has resulted in the soil becoming enriched with effective rhizobial symbionts. Indeed, Vlassak et al. (1973) reported in a Saskatchewan field study using the acetylene reduction technique that *C. arborescens* fixed 335 μg N/g soil /h. In the present study, the contribution of fixed nitrogen and identity of the rhizobial symbiont of *C. arborescens* growing in western Canada was determined by $\delta^{15}N$ natural abundance method and sequence analysis of selected genes.

Mesorhizobium spp. from this study exhibited an intermediate growth rate (g) similar to that described of other rhizobia isolated from *C. arborescens* root nodules from diverse locations and similar to *Mesorhizobium amorpheae*; slower than, for example, *Rhizobium leguminosarum* NRG457, *R. tropici* 899 and *Sinorhizobium meloliti* BALSAC, but faster than *Bradyrhizobium japonicum* USDA 110 (Gregory and Allen 1953, Jarvis et al. 1982, Willems, 2006). *Mesorhizobium* spp. from this study formed effective nodules on *C. arborescens* roots and augmented the allocation of N into the leaves in a gnotobiotic system (Vincent 1970).

Plant tissue from *C. arborescens* inoculated with isolates Laura, Harris, Sovereign, Thrasher and Kernan had much higher total N content and a negative $\delta^{15}N$ signal indicative of effective symbiotic N_2 fixation. The ^{15}N natural abundance method is based on the premise that plants that do not fix atmospheric N_2 have a positive ^{15}N content while N_2-fixing plants have a lower or negative ^{15}N content. Non-fixing plants extract soil N in the form of NH_4^+ and NO_3^- (high ^{15}N content) while N_2-fixing plants rely on atmospheric N_2 (low ^{15}N content). Consequently, non-fixing plants have positive $\delta^{15}N$ content and N_2-fixing plants dilute the $\delta^{15}N$, yielding negative $\delta^{15}N$ content (Galiana et al. 2002).

Jensen (1967) reported rhizobia that nodulated *Lotus* spp. formed a cross-inoculation group that included several other legumes including *C. arborescens*. The new species, *Rhizobium loti* that formed nodules on *C. arborescens*, first appeared in Bergey's Manual of Systematic Bacteriology Volume 1 in 1984 but was later transferred to the new genus, *Mesorhizobium* (Jarvis et al. 1997). *Mesorhizobium* spp. have been identified by 16S rRNA gene sequence analysis as being responsible for forming nodules on *C. arborescens* in China and Russia (Chen et al. 2008, Baymiev et al. 2010).

Determination of the 16S rRNA-encoding sequences of the isolates revealed clustering patterns that were consistent with those reported by Chen et al. (2008), with isolates Laura and Kernan apparently more similar to *M. amorphae* and *M. australicum*, and the other isolates clustering with a number of different *Mesorhizobium* spp. that have been reported to nodulate *C. arborescens*, including *M. caraganae* (Guan et al. 2008) and *Mesorhizbium* spp. CCBAU 11226 (Chen et al 2008). These results suggested that the

295 isolates represented two groups of strains, but the high level of sequence conservation at
296 the 16S locus among these strains made it difficult to discern their taxonomic relatedness.
297 Since the high similarity of the 16S rRNA-encoding sequences of the isolates
298 and reference strains made it difficult to determine the taxonomic assignment of the
299 isolates, two protein-encoding genes were examined. The 555 nucleotide (185 amino
300 acid) *cpn60* UT can predict taxonomic assignments using a genomic sequence identity
301 prediction model (Verbeke et al. 2011) and has been proposed as a DNA barcode marker
302 for Bacteria due to its high sequence diversity and barcode gap compared to 16S rRNA-
303 encoding genes (Links et al. 2012). Furthermore, *cpn60* UT sequences can be compared
304 to cpnDB, a publicly available database of *cpn60* UT sequences (www.cpndb.ca; Hill et
305 al. 2004). Despite these advantages, *Mesorhizobium* spp. and other root-nodulating
306 *Alphaproteobacteria* harbor multiple, distinct copies of the *cpn60* gene, some of which
307 are thought to play a role in the root-nodulating phenotype (Lund 2009, Shimoda et al.
308 2008, Okubo et al. 2012, Wallington and Lund 1994). This has the potential to
309 complicate the assignment of taxonomic identities using *cpn60* gene sequences since
310 gene transfer events among otherwise distinct species that have similar functional roles
311 may occur. We therefore compared the *cpn60* UT sequences for each of the reference
312 strains for which all copies of *cpn60* are known. This analysis revealed that certain *cpn60*
313 copies shared very high sequence identities among demonstrably distinct species of
314 *Mesorhizobium* (Table 3), which may indeed reflect relatively recent horizontal gene
315 transfer events. However, we could find no evidence in cpnDB that any of these
316 sequences were found in any genus besides *Mesorhizobium* spp., which suggests that
317 such gene transfer events may be restricted to taxonomically closely related

318 *Mesorhizobium* spp. The *cpn6*0 UT may therefore be useful as a taxonomic marker for
319 *Mesorhizobium* spp., but it seems unable to resolve below the genus level in this case,
320 unlike for most other taxonomic groups (Links et al. 2012). This is likely true of other
321 root-nodulating *Alphaproteobacteria* as well.
322 Comparing the *cpn60* UT sequences of the isolates to one another and to
323 reference sequences revealed that the various *cpn60* copies of the isolates tended to
324 cluster together, with several cases of identical *cpn60* UT sequences coming from
325 different isolates. Notably, one sequence was shared in all 5 of the isolates (supplemental
326 Figure 1) which was most closely related (94% sequence identity) to one of the *cpn60*
327 genes of *M. loti* MAFF303099 (cpnDB ID b1085). It is possible that this copy of *cpn60*
328 plays a role in the nodulation host range of these isolates, since they were found in
329 distinct geographic regions of Saskatchewan but nodulate the same plant species.
330 *recA* has been examined as a sequence that could be used to predict genomic
331 sequence identity (Zeigler 2003), and has been shown to display similar phylogenies to
332 the 16S rRNA-encoding gene for root-nodulating *Alphaproteobacteria* with somewhat
333 higher sequence divergence (Gaunt et al. 2001). *recA* analysis indicated that isolates
334 Sovereign, Thrasher and Kernan form a group with *M. loti* and *M. ciceri* and are
335 apparently distinct from the Chinese isolates reported by Chen et al. (2008) and Yan et al.
336 (2007). Consistent with this, isolates Sovereign and Thrasher also had identical 16S
337 rRNA-encoding gene fragments. However, while *recA* is typically a better predictor of
338 species assignment than 16S rRNA-encoding genes (Ziegler 2003), the 16S rRNA-
339 encoding gene fragment of isolate Kernan was only 98% identical to the *recA* gene
340 fragment of isolates Sovereign and Thrasher (Figure 1 and 2). This discrepancy could be

due to the fact that full-length *recA* genes were not determined in this study. Altogether, molecular taxonomic analysis using the 3 genes investigated consistently placed all of the isolates within the genus *Mesorhizobium*, although none of the three markers was able to conclusively assign these isolates to a known species. These results underscore the difficulties in assigning taxonomic identities to Mesorhizobium spp., even with whole genome sequences available (Sahgal and Johri 2006). It is likely that the determination of the whole or partial genomic sequences of the isolates as suggested by Richter and Rossello-Mora (2009) would help to resolve this.

Acknowledgements

This project is supported by the Saskatchewan Ministry of Agriculture Strategic Research Program: Soils and Environment-Soil Biological Processes and in part by a NSERC Strategic Project Grant. We thank Dr.'s Richard Farrell, Department of Soil Science, University of Saskatchewan for his contributions to N dynamics, including ^{15}N natural abundance work in this study and S. Hemmingsen, NRC-Plant Biotechnology Institute for assistance with DNA sequencing. This manuscript is dedicated to the late Dr. Peter Graham who provided the inspiration to complete this work.

References

Allen, O.N. and Allen, E.K. 1939. Root nodule bacteria of some tropical leguminous plants. Soil Sci. **47:** 63-76.

Baymiev, A.K., Ptitsyn, K.G., and Baimiev, A.K. 2010. Influence of the introduction of *Caragana arborescens* on the composition of its root-nodule bacteria. Microbiology **79**: 115-120.

Chen, W.F., Guan, S.H., Zhao, C.T., Yan, X.R., Man, C.X., Wang, E.T., and Chen, W.X. 2008. Different *Mesorhizobium* species associated with *Caragana* symbiotic genes and have common host ranges. FEMS Microbiol Lett **283**: 203-209.

Cram, W.H. 1969. Breeding and genetics of caragana. The Forestry Chronicle **45**: 400-401.

Dietz, D.R., Slabaugh, P.E. and Bonner, F.T. 2008. *Caragana arborescens* Lam. *In* The Woody Plant Seed Manual, Agriculture Handbook No. 727, *Edited by* F. T. Bonner and R. P. Karrfalt. U.S. Department of Agriculture, Forest Service, Washington, DC. pp. 321-323.

Dorsch, M., and Stackebrandt, E. 1992. Some modifications in the procedure of direct sequencing of PCR amplified 16S rDNA. J. Microbiol. Meth. **16:** 271-279.

Galiana A., Balle P., Kang, A.N., Domenach, A,M. 2002. Nitrogen fixation estimated by the 15N natural abundance method in *Acacia mangium* Willd. inoculated with *Bradyrhizobium* sp. and grown in silvicultural conditions. Soil Biol. Biochem. **34:** 251–262.

380 Gaunt, M. W., Turner, S. L., Rigottier-Gois, L., Lloyd-Macgilp, S. A. and Young, J. P.
381 2001. Phylogenies of atpD and recA support the small subunit rRNA-based
382 classification of rhizobia. Int. J. Syst. Evol. Microbiol. **51:** 2037-2048.
383 Gregory, K.F. and Allen, O.N. 1953. Physiological variations and host plant specificities
384 of rhizobia isolated from *Caragana arborescens* L. Can. J. Bot. **31**: 730-738.
385 Guan, S.H., Chen, W.F., Wang, E.T., Lu, Y.L., Yan, X.R., Zhang, X.X., and Chen, W.X.
386 2008. *Mesorhizobium caraganae* sp. nov., a novel rhizobial species nodulated with
387 *Caragana* spp. in China. Int. J. Syst. Evol. Microbiol. **58:** 2646 - 2653.
388 Hensley D.L., and Carpenter P.L. 1979. The effect of temperature on N_2 fixation (C_2H_2
389 reduction) by nodules of legume and actinomycete-nodulated woody species. Bot.
390 Gaz. **140**: 558-564.
391 Hill, J. E., Penny, S. L., Crowell, K. G., Goh, S. H and Hemmingsen, S. M**.** 2004.
392 cpnDB: A Chaperonin Sequence Database. Genome Res. **14:**1669-1675.
393 Hill, J. E., Hemmingsen, S. M., Goldade, B. G., Dumonceaux, T. J., Klassen, J., Zijlstra,
394 R. T., Goh, S. H., and Van Kessel, A. G. 2005 (a). Comparison of ileum microflora
395 of pigs fed corn-, wheat-, or barley-based diets by chaperonin-60 sequencing and
396 quantitative PCR. Appl. Environ.Microbiol. **71:** 867-875.
397 Hill, J.E., Town, J.R., and Hemmingsen, S.M. 2005 (b). Improved template
398 representation in *cpn60* polymerase chain reaction (PCR) product libraries
399 generated from complex templates by application of a specific mixture of PCR
400 primers. Environ. Microbiol. **8**: 741-746.
401 Hoagland, D.R. and Arnon, D.I. 1950. The water culture method for growing plants
402 without soil. California Agricultural Experimental Station Circular. No. 347.

403 Jarvis, B.D.W., Pankhurst, C.E., and Patel, J.J. 1982. *Rhizobium loti*, a new species of
404 legume root nodule bacteria. Int. J. Syst. Bact. **32**: 378-380.
405 Jarvis, B.D.W., Van Berkum, P., Chen, W.X., Nour, S.M., Fernandez, M.P., Cleyet-
406 Marel, J.C., and Gillis, M. 1997. Transfer of *Rhizobium loti*, *Rhizobium huakuii*,
407 *Rhizobium ciceri*, *Rhizobium mediterraneum*, and *Rhizobium tianshanense* to
408 *Mesorhizobium* gen. nov. Int. J. Syst. Bact. **47**: 895-898.
409 Jensen, H.L. 1967. Mutual host plant relationships in two groups of legume root nodule
410 bacteria (*Rhizobium* spp). Arch. Mikrobiol. **59**: 174-179.
411 Links, M.G., Dumonceaux T.J., Hemmingsen S.M., and Hill J.E. 2012. The chaperonin-
412 60 universal target is a barcode for bacteria that enables *de novo* assembly of
413 metagenomic sequence data. PLoS One.;7(11):e49755.
414 Lund, P. A. 2009. Multiple chaperonins in bacteria--why so many? FEMS Microbiol.
415 Rev. **33:** 785-800.
416 Moukoumi, J., Farrell, R.E., Van Rees, K.J.C., Hynes, R.K., Bélanger, N. 2012. Growth
417 and Nitrogen dynamics of juvenile short rotation intensive cultures of pure and
418 mixed *Salix miyabeana* and *Caragana arborescens*. Bioenergy Research **5**: 719-
419 732.
420 Okubo, T., Tsukui, T., plus 46 others. 2012. Complete genome sequence of
421 *Bradyrhizobium* sp. S23321: insights into symbiosis evolution in soil oligotrophs.
422 Microbes and environments. JSME **27**: 306-315.
423 Ralser, M., Querfurth, R., Warnatz, H.J., Lehrach, H., Yaspo, M.L., and Krobitsch, S.
424 2006. An efficient and economic enhancer mix for PCR. Biochem Biophys Res
425 Commun. **347**: 747-51.

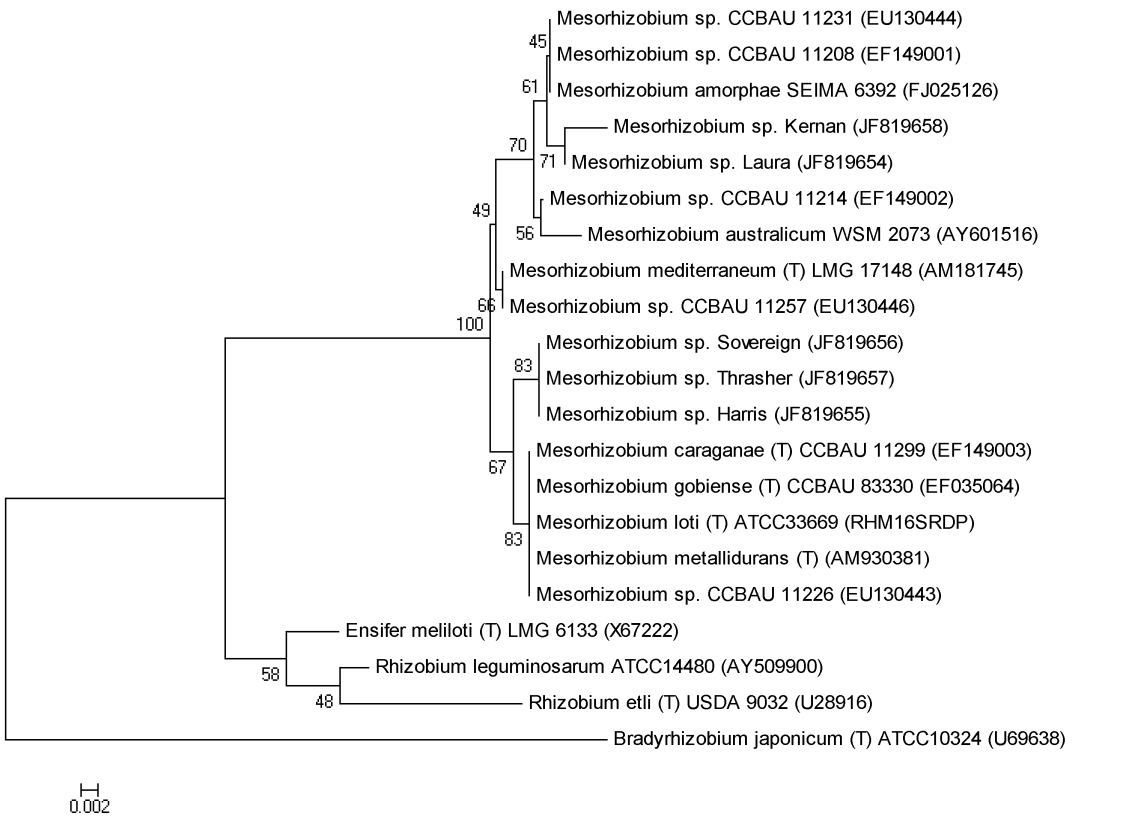
426 Richter, M. & Rossello-Mora, R. 2009. Shifting the genomic gold standard for the
427 prokaryotic species definition. Proc. Nat. Acad. Sci. USA. **106**: 19126-19131.
428 Sahgal, M., and Johri, B.N. 2006. Taxonomy of rhizobia: Current status. Current Science
429 90: 486-487.
430 Saitou, N. and Nei, M. 1987. The neighbor-joining method: A new method for
431 reconstructing phylogenetic trees. Mol. Biol. Evol. **4**: 406-425.
432 Shimoda, Y., Shinpo, S., Kohara, M., Nakamura, Y., Tabata, S., Sato, S. 2008. A large
433 scale analysis of protein-protein interactions in the nitrogen-fixing bacterium
434 *Mesorhizobium loti*. DNA research: an international journal for rapid publication of
435 reports on genes and genomes **15**: 13-23.
436 Tamura, K., Dudley, J., Nei, M., and Kumar**,** S. 2007. MEGA4: Molecular Evolutionary
437 Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. **24:**1596-1599.
438 Thompson, J. D., Higgins, D. G. and Gibson, T. J. 1994. CLUSTAL W: improving the
439 sensitivity of progressive multiple sequence alignment through sequence weighting,
440 position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**
441 4673-4680.
442 Turner S. L., Zhang X.X., Li, F.D., Young, J.P. 2002.What does a bacterial genome
443 sequence represent? Mis-assignment of MAFF 303099 to the genospecies
444 *Mesorhizobium loti*. Microbiology. **148**: 3330-3331.
445 Verbeke, T.J., Sparling, R., Hill, J.E., Links, M.G., Levin, D., and Dumonceaux, T.J.
446 2011. Predicting relatedness of bacterial genomes using the chaperonin-60
447 universal target (cpn60 UT): Application to *Thermoanaerobacter* species. Syst.
448 Appl. Microbiol. **34**: 171-179.

449 Vincent, J.M. 1970. A practical manual for the study of root-nodule bacteria,
450 International Biological Program Handbook No. 15, Blackwell Scientific
451 Publications Ltd., Oxford, UK.
452 Vlassak, K., Paul, E., and Harris, R. 1973. Assessment of biological nitrogen fixation in
453 grassland and associated sites. Plant Soil **38**: 637-649.
454 Wallington, E. J. and Lund, P. A. 1994. *Rhizobium leguminosarum* contains multiple
455 chaperonin (cpn60) genes. Microbiology **140**: 113-122.
456 Weaver, R.W. and Frederick, L.R. 1982. *Rhizobium*. *In* Methods of Soil Analysis, Part 2.
457 Chemical and Microbiological properties – Agronomy Monograph No. 9, Madison,
458 WI, USA. pp 1043-1070.
459 Willems, A., Hoste, B., Tang, J., Janssens, D., and Gillis, M. 2001. Differences between
460 subcultures of the *Mesorhizobium loti* type strain from different culture collections.
461 Syst. Appl. Microbiol. **24**: 549-553.
462 Wills, B.J. 1982. Rhizobial inoculation of Siberian pea-shrub (*Caragana arborescens*
463 Lamb.) for semi-arid revegetation. New Zeal. J. Exp. Agr. **10**: 353-354.
464 Yan, X.R., Chen, W.F., Fu, J.F., Lu, Y.L., Xue, C.Y., Sui, X.H., Li, Y., Wang, E.T., and
465 Chen, W.X. 2007. *Mesorhizobium* spp. are the main microsymbionts of Caragana
466 spp. grown in Liaoning Province of China. FEMS Microbiol. Lett. **271**: 265-273.
467 Zeigler, D. R. 2003. Gene sequences useful for predicting relatedness of whole genomes
468 in bacteria. Int. J. Syst. Evol. Microbiol. **53**, 1893-1900.

Figure legends

Figure 1. Phylogenetic tree of *Mesorhizobium* and related species based on a gene fragment corresponding to nucleotides 11-536 of the *E. coli* 16S rRNA-encoding gene. Strains are identified by their GenBank accession numbers (in parentheses). The percentage of 500 replicate trees in which the taxa clustered together is shown next to the corresponding nodes.

Figure 2. Phylogenetic analysis of the *recA* sequences of *Mesorhizobium* isolates from this report, placed in the context of the isolates reported by Chen et al (2008) and of selected reference strains. Sequences were all trimmed to the same length (426 bp) prior to alignment. The percentage of 500 replicate trees in which the taxa clustered together is shown next to the branches.



479

480 **Figure 1.**

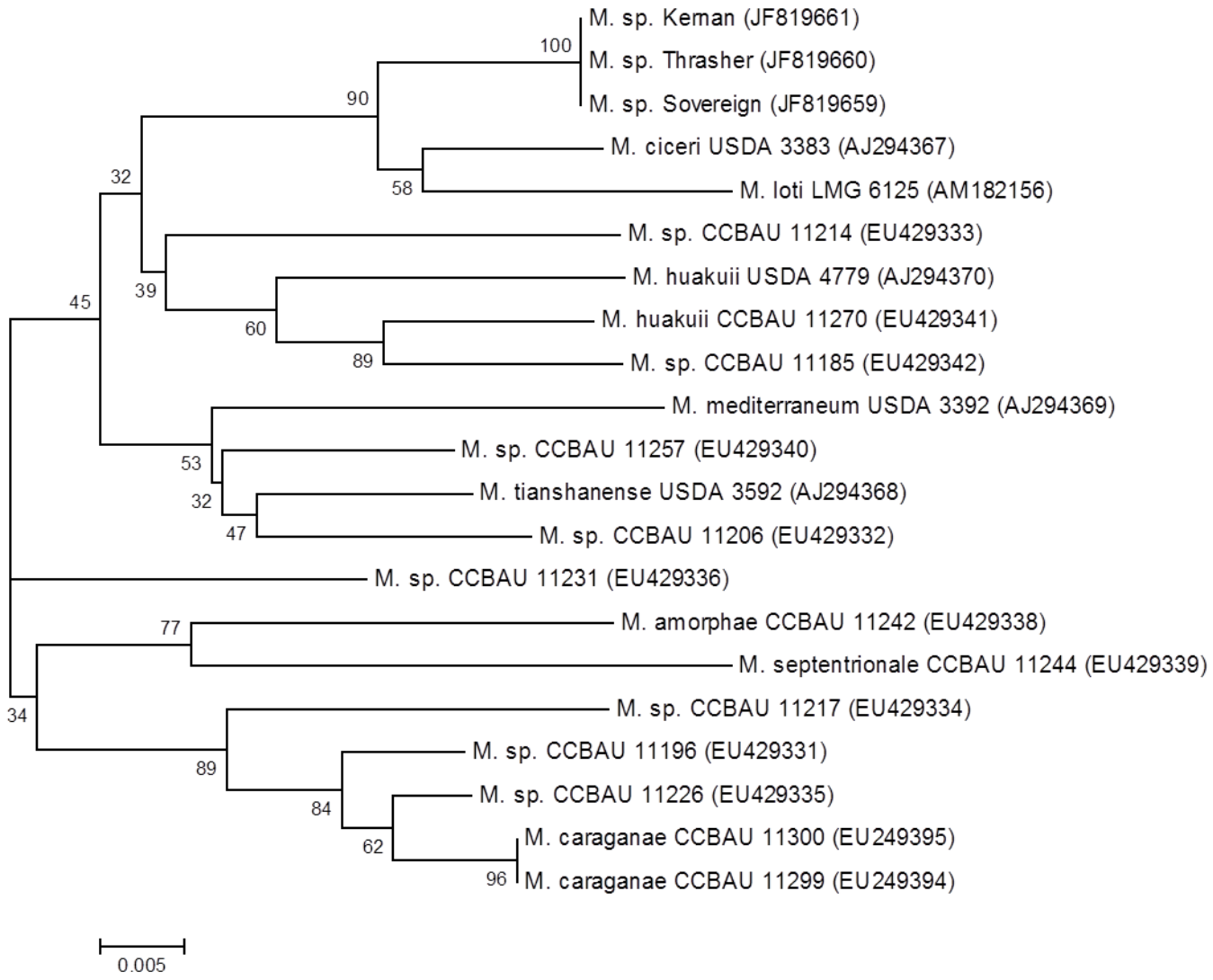

M. sp. Kernan (JF819661)
M. sp. Thrasher (JF819660)
M. sp. Sovereign (JF819659)
M. ciceri USDA 3383 (AJ294367)
M. loti LMG 6125 (AM182156)
M. sp. CCBAU 11214 (EU429333)
M. huakuii USDA 4779 (AJ294370)
M. huakuii CCBAU 11270 (EU429341)
M. sp. CCBAU 11185 (EU429342)
M. mediterraneum USDA 3392 (AJ294369)
M. sp. CCBAU 11257 (EU429340)
M. tianshanense USDA 3592 (AJ294368)
M. sp. CCBAU 11206 (EU429332)
M. sp. CCBAU 11231 (EU429336)
M. amorphae CCBAU 11242 (EU429338)
M. septentrionale CCBAU 11244 (EU429339)
M. sp. CCBAU 11217 (EU429334)
M. sp. CCBAU 11196 (EU429331)
M. sp. CCBAU 11226 (EU429335)
M. caraganae CCBAU 11300 (EU249395)
M. caraganae CCBAU 11299 (EU249394)
100
90
58
32
39
60
89
45
53
32
47
77
34
89
84
62
96
0.005

Figure 2.

491 **Table 1**. Generation times (g) of Rhizobaceae isolates used in this study.

Strain	g (hours)	SE
Mesorhizobium loti Laura	5.68	0.16
Mesorhizobium loti Harris	6.84	0.51
Mesorhizobium loti Sovereign	7.16	0.36
Mesorhizobium loti Thrasher	6.62	0.25
Mesorhizobium loti Kernan	5.76	0.19
Sinorhizobium melioti BALSAC	3.45	0.72
Bradyrhizobium japonicium USDA 110	13.86	2.08
Rhizobium leguminosarum NRG457	4.44	0.08
Rhizobium tropici 899[1]	3.19	0.70
Mesorhizobium amorpheae	7.76	0.86

492 Generation times determined in Yeast extract-mannitol broth, 23 °C, shaking at 150rpm; see
493 materials and methods for additional details. [1]CIAT899.
494

Table 2. Percent N, C and N_2 derived from the atmosphere from caragana leaves grown from seeds in Hoaglands nitrogen-free media with and without *Mesorhizobium* isolates.

Treatment	% N	% C	% N_2 derived from atmosphere
Non-inoculated	0.6±0.1	13.8±2.9	0
Mesorhhizobium sp Laura	3.9±0.3	44.5±0.7	83.0±2.1
Mesorhizobium sp Harris	2.8±0.4	44.8±0.4	78.5±4.5
Mesorhizobium sp Sovereign	2.6±0.4	44.1±0.8	76.6±5.5
Mesorhizobium sp Thrasher	2.9±0.1	44.5±0.2	80.1±1.0
Mesorhizobium sp Kernan	3.2	45.2	80.0

Data represent the mean of 3 reps, 2 for the non-inoculated ± the standard error and 1 for *Mesorhizobium sp.* Kernan.

Table 3. Determination of ANIm, ANIb, and Tetra scores using JSpecies along with corresponding genomic sequence identity predictions based on the *cpn60* UT and a one-gene *recA*-based model. Only representative results are shown for the 5 complete *M. loti* genomes (MAFF303099, WSM1293, R88b, NAP2037, and R7A) that were examined.

		analysis by JSpecies[1]				*cpn60* UT sequence identity	
genome	target genome	ANIm	ANIb	Tetra	conclusion	highest	lowest
M. loti MAFF303099 (5)	*M.australicum* WSM 2073 (3)	88.26	86.77	0.99708	different spp	0.95	0.75
M. loti MAFF303099 (5)	*M. ciceri* bv. biserrulae (4)	87.82	85.82	0.99828	different spp	0.96	0.76
M. loti MAFF303099 (5)	*M. opportunistum* (4)	89.12	87.84	0.99813	different spp	0.96	0.74
M. loti MAFF303099 (5)	*Mesorhizobium* sp. BNC1 (2)	82.78	71.83	0.89382	different spp	0.83	0.74
M. loti MAFF303099 (5)	*M. loti* R7A (4)	98.53	98.46	0.99936	same spp.	0.99	0.76
M. loti MAFF303099 (5)	*M. loti* WSM1293 (3)	87.89	85.83	0.99845	different spp	0.94	0.76
M. loti WSM1293 (3)	*M. loti* R88b (3)	87.77	85.86	0.99781	different spp	0.95	0.76
M.australicum WSM 2073 (3)	*M. ciceri* bv. biserrulae (4)	88.57	86.92	0.9966	different spp	1.00	0.77
M.australicum WSM 2073 (3)	*M. opportunistum* (4)	89.35	88.2	0.99853	different spp	1.00	0.76
M.australicum WSM 2073 (3)	*Mesorhizobium* sp. BNC1 (2)	83.12	71.64	0.89391	different spp	0.85	0.75
M. ciceri bv. biserrulae (4)	*M. opportunistum* (4)	89.34	89.89	0.99834	different spp	1.00	0.76
M. ciceri bv. biserrulae (4)	*Mesorhizobium* sp. BNC1 (2)	82.9	71.5	0.89034	different spp	0.86	0.76
Mesorhizobium sp. BNC1 (2)	*M. opportunistum* (4)	82.96	71.83	0.89066	different spp	0.86	0.76

[1]JSpecies was used to determine ANIm, ANIb, and Tetra values as described by Richter and Rossello-Mora (2009). An ANIm/ANIb of 95% with Tetra >0.99 is used to designate two strains as the same species.

506 **Table 4.** cpnDB nearest neighbor for each copy of *cpn60* analyzed for all isolates. *cpn60* copies are numbered arbitrarily. Matches are
507 identified by their cpnDB ID numbers and, if applicable, by their GenBank accession numbers.

Isolate	cpnDB nearest neighbor	fasta %ID
Mesorhizobium spp. Laura		
cpn60-1	b4797 NC_002678 *Mesorhizobium loti* MAFF303099	92.4
cpn60-2	b19969 AGSN01000124 *Mesorhizobium amorphae* CCNWGS0123	87.4
cpn60-3	b1085 NC_002679 *Mesorhizobium loti* MAFF303099	87.2
cpn60-4	b1085 NC_002679 *Mesorhizobium loti* MAFF303099	94.1
cpn60-5	b19969 AGSN01000124 *Mesorhizobium amorphae* CCNWGS0123	96.4
Mesorhizobium spp. Harris		
cpn60-1	b19971 AGSN01000157 *Mesorhizobium amorphae* CCNWGS0123	92.6
cpn60-2	v12449 *Mesorhizobium amorphae* DSM 21831	89.9
cpn60-3	b1085 NC_002679 *Mesorhizobium loti* MAFF303099	94.1
cpn60-4	b18426 CP002447 *Mesorhizobium ciceri* biovar biserrulae	98.7
cpn60-5	b19967 AGSN01000015 *Mesorhizobium amorphae* CCNWGS0123	86.3
Mesorhizobium spp. Sovereign		
cpn60-1	b19971 AGSN01000157 *Mesorhizobium amorphae* CCNWGS0123	92.6
cpn60-2	b1085 NC_002679 *Mesorhizobium loti* MAFF303099	94
cpn60-3	b18426 CP002447 *Mesorhizobium ciceri* biovar biserrulae	98.7
Mesorhizobium spp. Thrasher		
cpn60-1	b19971 AGSN01000157 *Mesorhizobium amorphae* CCNWGS0123	87.5
cpn60-2	b1085 NC_002679 *Mesorhizobium loti* MAFF303099	94.1
cpn60-3	19971 AGSN01000157 *Mesorhizobium amorphae* CCNWGS0123	92.6
cpn60-4	v12449 *Mesorhizobium amorphae* DSM 21831	89.9
cpn60-5	b18426 CP002447 *Mesorhizobium ciceri* biovar biserrulae	98.7
cpn60-6	b18427 CP002447 *Mesorhizobium ciceri* biovar biserrulae	95.3
cpn60-7	v12449 *Mesorhizobium amorphae* DSM 21831	96
Mesorhizobium spp. Kernan		

cpn60-1	b19969 AGSN01000124 *Mesorhizobium amorphae* CCNWGS0123	91
cpn60-2	b1085 NC_002679 *Mesorhizobium loti* MAFF303099	94.1
cpn60-3	b19969 AGSN01000124 *Mesorhizobium amorphae* CCNWGS0123	98.2
cpn60-4	b4797 NC_002678 *Mesorhizobium loti* MAFF303099	92.3