

<https://helda.helsinki.fi>

Using amplicon sequencing of rpoB for identification of inoculant rhizobia from peanut nodules

Mousavi, S. Abdollah

2022-02

Mousavi , S A , Gao , Y , Penttinen , P , Frostegård , Å , Paulin , L & Lindström , K 2022 , ' Using amplicon sequencing of rpoB for identification of inoculant rhizobia from peanut nodules ' , Letters in Applied Microbiology , vol. 74 , no. 2 , pp. 204-211 . <https://doi.org/10.1111/lam.13599>

<http://hdl.handle.net/10138/353916>

<https://doi.org/10.1111/lam.13599>

acceptedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

ORIGINAL ARTICLE

Using amplicon sequencing of *rpoB* for identification of inoculant rhizobia from peanut nodules

S.A. Mousavi^{1,2}, Y. Gao^{1,3,4}, P. Penttinen⁵ , Å. Frostegård³, L. Paulin⁶ and K. Lindström¹

- 1 Ecosystems and Environment Research Programme, Faculty of Biological and Environmental Sciences and Helsinki Institute of Sustainability Science (HELSUS), University of Helsinki, Helsinki, Finland
- 2 Division of Ecology and Evolutionary Biology, Department of Biology, University of Turku, Turku, Finland
- 3 Department of Chemistry, Biotechnology and Food Sciences, Norwegian University of Life Sciences, Ås, Norway
- 4 International Research and Development Center, Kingenta Ecological Engineering Group Co. Ltd, Linyi, China
- 5 College of Resources, Sichuan Agricultural University, Chengdu, China
- 6 DNA Sequencing and Genomics Laboratory, Institute of Biotechnology, University of Helsinki, Helsinki, Finland

Significance and Impact of the Study: Legume crops are often inoculated with effective rhizobia. To assess the success of the inoculant, it is necessary to distinguish it from closely related strains. Methods used until now do not provide sufficiently specific information on the identity of the inoculant strain. Here, we have developed a method that is based on amplicon sequencing of the bacterial housekeeping gene *rpoB*, encoding the beta-subunit of the RNA polymerase. The analyses of the sequence data showed that the method reliably identified bradyrhizobial strains in nodules, at least at the species level, and could be used to assess the competitiveness of the inoculant compared to other bradyrhizobia.

Keywords

field experiment, identification, inoculum, nodule, peanut, pouch experiment, Rhizobia, *rpoB* amplicon sequencing.

Correspondence

Petri Penttinen, Department of Microbiology, College of Resources, Sichuan Agricultural University, No. 211 Huimin Road, Chengdu 611130, China.
E-mail: petri.penttinen@helsinki.fi

2021/0177: received 19 August 2021, revised 22 October 2021 and accepted 28 October 2021

doi:10.1111/lam.13599

Abstract

To improve the nitrogen fixation, legume crops are often inoculated with selected effective rhizobia. However, there is large variation in how well the inoculant strains compete with the indigenous microflora in soil. To assess the success of the inoculant, it is necessary to distinguish it from other, closely related strains. Methods used until now have generally been based either on fingerprinting methods or on the use of reporter genes. Nevertheless, these methods have their shortcomings, either because they do not provide sufficiently specific information on the identity of the inoculant strain, or because they use genetically modified organisms that need prior authorization to be applied in the field or other uncontained environments. Another possibility is to target a gene that is naturally present in the bacterial genomes. Here we have developed a method that is based on amplicon sequencing of the bacterial housekeeping gene *rpoB*, encoding the beta-subunit of the RNA polymerase, which has been proposed as an alternative to the 16S rRNA gene to study the diversity of rhizobial populations in soils. We evaluated the method under laboratory and field conditions. Peanut seeds were inoculated with various *Bradyrhizobium* strains. After nodule development, DNA was extracted from selected nodules and the nodulating rhizobia were analysed by amplicon sequencing of the *rpoB* gene. The analyses of the sequence data showed that the method reliably identified bradyrhizobial strains in nodules, at least at the species level, and could be used to assess the competitiveness of the inoculant compared to other bradyrhizobia.

Introduction

Use of biological nitrogen fixation is essential when aiming at sustainable intensification in agriculture. Symbiotic nitrogen fixation (SNF) in legumes is the main biological process in agriculture, contributing to 100–122 Tg N per year, compared with industrial nitrogen fixation, through which approximately 140 Tg N per year is added to agricultural fields (Herridge *et al.* 1984). SNF takes place in legume root nodules in which bacteria called rhizobia perform the reduction of N₂ gas to ammonia (Lindström and Mousavi 2020). To optimize the symbiosis, rhizobia are often added to the legume crop as biofertilizer in inoculant preparations carrying selected elite rhizobia. Optimal functioning of the symbiosis requires that the selected strain(s) are adapted to local soils and able to outcompete indigenous rhizobia during the nodulation.

Inoculant quality control should ensure that the majority of the legume root nodules have been nodulated by the rhizobial strain(s) used as inoculant. Thus, there is a need for rapid and accurate strain identification from root nodules formed in the field (Tas *et al.* 1995). Identification of indigenous bacteria capable of entering the nodules would give additional information about rhizobial ecology in the field, useful for the development of sustainable intensification.

Through the years, a diversity of methods has been introduced and tested for the purpose. Intrinsic antibiotic resistance fingerprinting was introduced as an identification method in 1980 (Beynon and Josey 1980). In the 1990s and 2000s, the improvement of molecular techniques revolutionized bacterial identification methods. PCR-based fingerprinting methods, for example, PCR with either species-specific primers, rep-PCR, RFLP and RAPD along with the use of recombinant marker or reporter genes such as *gusA*, *lacZ*, *luc* or *lux*, or DNA-DNA dot-blot hybridization became popular genotyping methods (Tas *et al.* 1994; Nick *et al.* 1999; Pitkääjärvi *et al.* 2003; Sikora and Redžepović 2003). The methods enabled a better differentiation among closely related bacterial strains and the detection of a higher rhizobial diversity in nodules than before (Sikora and Redžepović 2003). These methods have been used to identify bacteria occupying the nodules, to monitor specific cultivable bacterial strains in field releases and to study the persistence, population dynamics and competitiveness for nodulation of rhizobial strains (Tas *et al.* 1996; Pitkääjärvi *et al.* 2003). However, these methods do not discriminate between strains in nodules accommodating multiple strains.

Amplicon sequencing of the 16S rRNA gene is nowadays the most extensively used method in molecular identification and classification studies. However, because of the fairly low resolution of 16S rRNA genes, amplicon

sequencing targeting other housekeeping or core genes was recommended by, for example, Vos *et al.* (2012) and Zhang *et al.* (2017). The *recA* and *rpoB* genes are commonly used phylogenetic markers with a genetic divergence that can give a better resolution of bacterial communities than the 16S rRNA gene (Vos *et al.* 2012; Gołębiewski and Tretyn 2020). Because of the wide use of these genes in many studies, extensive databases are now available for identification and classification purposes. Vos *et al.* (2012) proposed that *rpoB* could be used as a good marker for studying intra-species level bacterial diversity focusing on specific taxonomic groups. Recently, Zhang *et al.* (2017) used primers *rpoB*1479-F and *rpoB*1831-R to amplify *rpoB* gene fragments and suggested that amplicon sequencing targeting the *rpoB* gene was a useful tool for identification and classification of rhizobia isolated from soybean nodules belonging to the genus *Bradyrhizobium*.

Here, we investigated the potential of using the *rpoB* gene as a marker for identification of *Bradyrhizobium* inoculants in legume nodules. The purpose was to design a convenient and reliable tool for specific tracing of inoculants in field studies, with the capacity of discriminating the inoculant from indigenous bradyrhizobia and in addition give information of their diversity. The study was part of a project aiming at the development of inoculants for peanut (*Arachis hypogaea*) in the Shandong province in China, a major peanut growing area. Out of 62 species in the genus *Bradyrhizobium*, the species *B. arachidis*, *B. guangdongense*, *B. guangxiense* and *B. lablabi* have all been isolated from peanut plants growing in China (Chang *et al.* 2011; Wang *et al.* 2013; Li *et al.* 2015). We constructed a database containing *rpoB* sequences of 311 bradyrhizobial reference strains, followed by the application of amplicon sequencing targeting the *rpoB* gene in DNA isolated from the nodules of peanut plants (i) inoculated with bradyrhizobial strains in axenic culture in a growth chamber; (ii) uninoculated plants growing in local farmers' fields in the Shandong province in China and (iii) inoculated with bradyrhizobial strains in a field experiment in the same area.

Results and discussion

The growth chamber experiment was done as a first test of the feasibility of the method. Firstly, seeds were inoculated with one single *Bradyrhizobium* strain. The results from the single-strain inoculations showed that when any of the strains *Bradyrhizobium* sp. AC101b, CCBAU 53344, CCBAU 53363, HAMBI 2142, HAMBI 2130, and HAMBI 2125 or HAMBI 2135 was used as a single inoculant, 99.4–99.9% of the *rpoB* sequences obtained from the nodules were assigned to the inoculant (Table 1, Table S1),

Table 1 Sequence identities and relative abundances of the *rpoB* gene amplicons in the growth chamber experiment. The relative abundances are calculated as percentage of the total abundance. Only percentages above 0.10 are shown. The relative abundances of the detected inoculants are in bold

Reference	Inoculant												
	AC101b	CCBAU 53344	CCBAU 53363	HAMBI 2142	HAMBI 2130	HAMBI 2125	HAMBI 2135	HAMBI 2125+	HAMBI 2130	CCBAU 53363	CCBAU 53344	HAMBI 2142+	Negative control
<i>Bradyrhizobium arachidis</i> LMG 26795													82.3
CCBAU 53344/CCBAU 53363*	0.5	99.9	99.8							89.0	88.9		
<i>Bradyrhizobium</i> sp. HAMBI 2125						99.8		1.9					
<i>B.</i> sp. HAMBI 2130/AC101b†	99.4				99.9			98.0		10.8			
<i>Bradyrhizobium</i> sp. HAMBI 2133													17.4
<i>Bradyrhizobium</i> sp. HAMBI 2135							99.9						
<i>Bradyrhizobium</i> sp. HAMBI 2142				99.9							8.7		
<i>Rhizobiales</i> unclassified											2.4		0.2

*The sequences of the *rpoB* gene of these two strains, both of which belong to *B. guangxiense*, are 100% identical.

†The sequences of the *rpoB* gene of these two strains are 100% identical.

thus demonstrating that the nodules were occupied almost exclusively by that strain. Secondly, seeds were inoculated with a 50 : 50% mixture of two strains to assess the capability of the method to discriminate multiple strains. In all those cases, both strains were identified in the nodules, but one of them was strongly dominating over the other. *Bradyrhizobium* sp. HAMBI 2130 accounted for 98.0% and HAMBI 2125 only 1.9% of the relative abundances. Strain CCBAU 53344 accounted for the main part of the relative abundance, both when co-inoculated with strain AC101b (89.0% vs 10.8%) and with strain HAMBI 2142 (88.9% vs 8.7%). Concluding absolute abundances from amplicon sequencing data is possible by using internal standards or reference frames, or additional methods like quantitative PCR or flow cytometry (Morton *et al.* 2019). Because the amplicon sequencing data are compositional and maintains the relationships between parts (Gloor *et al.* 2017; McLaren *et al.* 2019), the ratios of taxa in the samples provided information on the ratios of inoculants in the nodules. Because the method allowed successful identification of closely related strains, assessing the competitiveness in a further study with appropriate replicates and methods suitable for analysing compositional data was deemed feasible.

The use of closely placed pouches in a growth chamber is efficient for first studies of nodule formation using different bacterial strains/legume varieties, but there is a risk of cross-contamination between pouches. In our

experiment, the negative control plants, which were inoculated with sterile distilled water, had formed nodules due to contamination from adjacent pouches. The nodules in the control treatments were nodulated mostly by the two strains *B. arachidis* and *B. ottawaense* used in a parallel study. Despite the shortcomings of these results regarding the negative controls, they provided a first indication that the *rpoB* gene can be used to distinguish closely related strains of bradyrhizobia and that it can be used to compare the competitiveness of different strains for nodulation.

To assess the applicability of the *rpoB* amplicon sequencing method for detection of inoculated rhizobia in field conditions, we first analysed the rhizobial communities inside nodules collected from peanut plants growing on eight sites at local farms close to Linshu, Shandong, China (Table 2, Table S2). The *rpoB* communities at different sites were scattered in the principal component analysis (PCA) plot (Fig. S1), suggesting relatively large variation in indigenous peanut nodulating bradyrhizobia on a local scale. The sequences with highest relative abundances from the sites L2, L3 and L6 were assigned to *B. guangdongense*. Most of the sequences from the sites L1 and L8 were assigned to unclassified bradyrhizobial species, and approximately 11–12% of the sequences were assigned to *B. guangdongense*. In site L5, almost 100% of the sequences were assigned to unclassified bradyrhizobia. In site L7, 80, 7 and 7% were assigned to *B. arachidis*, *B. ottawaense* and unclassified

Table 2 Sequence identities and relative abundances of the *rpoB* gene amplicons in the local farms in Linshu, China. The relative abundances are as percentage of the total abundance. Only percentages above 0.10 are shown

Reference	Sample							
	L1	L2	L3	L4	L5	L6	L7	L8
<i>Bradyrhizobium</i> unclassified	0.7	0.8	8.0	60.1	91.7	0.9	5.0	6.5
<i>B. arachidis</i> USDA 3384							80.5	
<i>B. guangdongense</i> CCBAU 51649	11.9	98.6	91.6	39.7	0.2	97.4		10.9
<i>B. ottawaense</i> CCBAU 15635							7.2	
<i>Bradyrhizobium</i> sp. unclassified	87.3	0.6			6.6	1.5	7.3	82.6
<i>Rhizobiales</i> unclassified					0.8	0.1		
<i>Rhizobium</i> unclassified					0.2			
<i>Paraburkholderia phymatum</i> LMG 21445			0.2		0.2			

bradyrhizobia, respectively. In site L4, most sequences were assigned to unclassified bradyrhizobia and around 40% to *B. guangdongense*. Based on our results, *B. guangdongense* might be considered as one of the most common peanut-nodulating rhizobia in the sampling region.

In the field experiment, the *rpoB* community that nodulated the uninoculated peanuts (negative control) and the peanut plants inoculated with AC70c, AC87j1 and HAMBI 2115 resembled those in the local fields (Fig. S1). However, inoculants AC70c and AC87j1 were successfully identified from the nodules, with 71 and 37% of the sequences, respectively, assigned to the inoculant (Table 3, Table S3). In addition, inoculant strain HAMBI 2115 was also identified, yet only 1.25% of the sequences were assigned to the inoculant. Almost all the sequences obtained from peanut nodules inoculated with strain HAMBI 2127 were assigned to the inoculant strain, making the *rpoB* community to resemble those in the growth chamber experiment. The differences in relative abundances between the inoculants may reflect differences in competitiveness.

In this study, we present a method for identification of inoculant strains and indigenous rhizobia from peanut nodules. Using the 16S rRNA marker, the probability of correct species identification differs significantly between genera and species based on their sequence similarity (Winand *et al.* 2020). For most of the inoculant bradyrhizobia, the variable regions V1-V3, V4-V5, V6-V8 and V7-V8 of 16S rRNA gene were identical and over 99%

Table 3 Sequence identities and relative abundances of the *rpoB* gene amplicons in the field experiment. The relative abundances are as percentage of the total abundance. Only percentages above 0.10 are shown. The relative abundances of the detected inoculants are in bold

Reference	Inoculant				
	AC70c	HAMBI 2115	HAMBI 2127	AC87j1	Negative control
<i>Bradyrhizobium guangdongense</i> CCBAU 51649	27.6	96.5		61.4	93.9
<i>B. guangxiense</i> CCBAU 53363		1.9			
<i>Bradyrhizobium</i> sp. HAMBI 2115		1.3			
<i>Bradyrhizobium</i> sp. HAMBI 2127			99.9		
<i>Bradyrhizobium</i> sp. AC70c	71.1				
<i>Bradyrhizobium</i> sp. AC87j1				37.7	
<i>Bradyrhizobium</i> sp. unclassified					
<i>Rhizobiales</i> unclassified	1.1			0.7	5.9

similar to those of *B. guangdongense* CCBAU 51649 (Fig. S2), showing that distinguishing the inoculants and native strains may have been unfeasible using the 16S rRNA marker. Instead, the *rpoB* amplicon sequences separated the inoculants except HAMBI 2130 and AC101b that carried identical sequences (Fig. S2). Vos *et al.* (2012) proposed that the *rpoB* marker could be used as a complement to 16S rRNA gene for high throughput microbial diversity studies. In line with Zhang *et al.* (2017), who proposed that applying amplicon sequencing of the *rpoB* gene is a straightforward method to study rhizobial populations in soils, the results of this study showed that the *rpoB* marker is a promising target gene for identification bacteria inside nodules at the species level. With appropriate replicates, this method should be applicable to study competitiveness of different inoculants inside the nodules of inoculated legumes.

Materials and methods

Growth chamber experiment

Two common peanut cultivars growing in China, Bai Sha and 150, were used in the growth chamber experiment. Peanut seeds were sterilized in 96% ethanol for 1 min and 4% NaClO for 3–5 min and rinsed with sterile distilled water four times (Hungria *et al.* 2016). The seeds were germinated at 25°C in the dark for 2 days. The

germinated seeds were inoculated by soaking them in a suspension containing 1×10^8 CFU per ml of the inoculant rhizobial strain (Table S4) for 2 min. Seeds in the negative control were soaked in autoclaved H₂O. One cv. Bai Sha and one cv. 150 seed, inoculated with the same inoculum (see below), were placed in a CYG seed germination pouch (16.50 cm wide by 18.00 cm high), (Mega International Ltd., Minneapolis, MN), with two pouches per treatment. The plants were placed in a growth chamber (MLR-350; SANYO Electric Co. Ltd., Osaka, Japan) set to the following conditions: a 2-h photo period with a temperature of 25°C, a 10-h photo period with a temperature of 29°C, a 2-h photo period with a temperature of 25°C and a 10-h period at 25°C without light. The pouches were irrigated with either autoclaved Milli-Q water or Jensen solution with macronutrients (following Jensen, 1942; Yates *et al.* 2016). Nodules were harvested 50 days after inoculation. Eight nodules from the four plants of each treatment were collected and pooled together. The nodules were treated as described below.

Field experiment and sampling from local farms

In the field experiment, the peanut seeds were inoculated with *Bradyrhizobium* strains HAMBI 2115, HAMBI 2127, AC70c and AC87j1. To prepare the inocula, 35 ml of a 1×10^8 CFU per ml bacterial broth culture was injected to a bag with gamma-sterilized peat-based carrier (Elomestari Oy, Tornio, Finland) and sterile distilled water was injected to a carrier bag for the negative control treatment. The bags were incubated at 28°C for 5 days. The inocula were mixed with 100 ml water and the seeds were soaked in the inoculum suspensions before sowing (following the procedure provided by Elomestari Oy). The plants were grown under normal field conditions, applying a randomized complete block design with four replicates. Plants were collected after flowering (beginning of pod forming) and nodules from eight root systems per treatment were pooled together.

Peanut plants from local peanut farms were collected from eight sites close to Linshu, Shandong, China (Table S5). The nodules from two plants per site were collected.

DNA isolation, PCR and sequencing

The nodules from the field experiment and the local fields were kept in tubes containing silica gel until DNA extraction. These nodules were washed in autoclaved ddH₂O to remove soil particles on the surface of nodules. The nodules from all experiments were surface sterilized in 96% ethanol for 1 min and in 4% NaClO for 3–5 min. The nodules were crushed with tweezers sterilized in 70%

ethanol, and 250 mg from the interior of the nodules were used for DNA extraction using NucleoSpin Soil kit (Macherey-Nagel GmbH Co. KG, Duren, Germany). The amplification of *rpoB* gene was performed in two PCR steps. A new set of primers were constructed based on rpoB1479F and rpoB1831R designed by Zhang *et al.* (2017) by adding truncated Illumina TruSeq adapter sequences to the 5' end of those primers. The primers Illum_rpoB1479F (5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCTGATCGARACGCCGGAAGG-3') and Illum_rpoB1831R (5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGCATGTTTCGARCCCAT-3') were used for amplification of the *rpoB* fragments from the DNA extracted from nodules. Amplification was initiated with denaturation at 98°C for 30 s, followed either by 15–17 cycles for samples from the chamber experiment or 17–25 cycles for samples from field experiments at 98°C for 30 s, 65 or 70°C for 30 s and 72°C for 30 s, and a final extension at 72°C for 5 min (Table S6). The amplicons were checked using a Universal Hood II gel documentation unit (Bio-Rad Laboratories, Inc., Hercules, CA). The amplicons were purified using Exonuclease I and FastAP Thermosensitive Alkaline Phosphatase (ThermoFisher Scientific, Waltham, MA). An aliquot (1–3 µl) was used as template in an index-PCR reaction where dual indexes were introduced for each sample. The used dual-indexes were selected using the Barcosel program (Somervuo *et al.* 2018). The obtained amplicons were pooled, purified, quantified and paired-end sequenced at the institute of Biotechnology, University of Helsinki (Finland) using Illumina's MiSeq platform and v 3 600 cycle reaction kit (San Diego, CA).

Reference data set

The sequences of reference strains were retrieved from JGI Integrated Microbial Genome and Microbiomes system (<https://img.jgi.doe.gov/>; Chen *et al.* 2021), Zhang *et al.* (2017) and our ongoing genomic project (JGI Gold study ID: Gs0134353). To sequence the *rpoB* gene of the strains AC101b, AC86d, AC70c and CCBAU 53344 for the reference data set, the strains were cultured in YMB medium (Vincent 1970) for 5 days at 28°C followed by DNA extraction using the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany). The *rpoB* gene was amplified using primers rpoB-1561F (5'-TCGCAGTTCATGGACCAGR-3') and rpoB-2453R (5'-GTAGCCGTTCCA SGGCATG-3') that were designed manually by applying Bioedit ver. 7.0.5.3 (Hall 1999). The reaction mixture for PCR included 0.25 µl Takara Ex Taq and 4 µl dNTP for one reaction (50 µl) with 1 µl DNA template. The amplification program included denaturation at 95°C for 2 min, followed by 39 cycles at 95°C for 15 s, 63°C for

15 s and 72°C for 60 s, and a final extension at 72°C for 7 min. The amplified fragments were checked in agarose gel and the target bands were cut and purified using Omega Bio-tek E.Z.N.A. gel extraction kit (Omega Bio-Tek Inc., Norcross, GA). Products were sequenced at GATC Biotech Company (Ebersberg, Germany) using Sanger sequencing. The sequences of the *rpoB* gene of all the reference strains (311 strains in total) were aligned applying CLUSTALW (Larkin *et al.* 2007) in BIOEDIT ver. 7.0.5.3 (Hall 1999).

Data analyses

The amplicon sequences were analysed using MOTHUR ver. 1.44.1 (Schloss *et al.* 2009) at CSC—IT Center for Science, Finland. Amplicon sequences were clustered into operational taxonomic units (OTUs) at 99% similarity. OTUs were assigned to taxa using the in-house built *rpoB* sequence database. Out of the 4184 OTUs, 42 OTUs with mean abundance greater than 10 were selected for PCA. Zeros in the relative abundance data were replaced using the count zero multiplicative method, and the relative abundance data were converted to proportions with the zCompositions package in R 3.6.3 (Palarea-Albaladejo and Martı́n-Fernández 2015). The data were transformed to their centred log ratios with the compositions R package (van den Boogaart *et al.* 2014). PCA was done with R 3.6.3 and visualized using ggplot2 and ggfortify R packages (Tang *et al.* 2016; Wickham 2016; Horikoshi and Tang 2018). To assess the possibility to distinguish the detected bradyrhizobia using 16S rRNA gene amplicon sequencing, the available 16S rRNA genes sequences (Table S7) were subjected to *in silico* PCR using primers for the variable regions V1–V3, V4–V5, V6–V8 and V7–V8 (Table S8) in MOTHUR. The *in silico* amplicons were aligned, distance matrices were calculated using the Jukes-Cantor model, and neighbour joining trees were constructed using R packages seqinr, DECIPHER, Rphylip and APE (Paradis *et al.* 2004; Charif and Lobry 2007; Revell and Chamberlain 2014; Wright 2016). For comparison, the *rpoB* amplicons of the same strains were analysed similarly.

Acknowledgements

The authors wish to thank Kingenta Ecological Engineering Group co. Ltd., Linyi, China for financial support (to ASM and YG) and Dr Li Xinzhu, Yan Tingshuai, Liu Jinyan, Xu Wenfeng, Gu Zhiguang and Hu Hongtao (Kingenta) for support with the field experiments. SAM was also funded by Finnish Cultural foundation and Finska Vetenskaps-Societeten. YG is grateful for a scholarship from China Scholarship Council (CSC). We thank Dr

Xinhua Sui at China Agricultural University for providing *B. guangxiense* and Petri Leinonen (Ceo of Elomestari Oy) for providing us carriers that were used for inoculation in the field experiment. The CSC—IT Center for Science, Finland, provided computational resources.

Conflict of Interest

No conflict of interest has been declared.

Author contributions

Seyed Abdollah Mousavi: Conceptualization, Methodology, Formal analysis, Investigation, Writing—Original Draft; Yuan Gao: Formal analysis, Investigation, Writing—Original Draft; Petri Penttinen: Conceptualization, Formal analysis, Writing—Original Draft; Åsa Frostegård: Writing—Original Draft, Supervision, Project administration, Funding acquisition; Lars Paulin: Methodology, Resources; Kristina Lindström: Writing—Original Draft, Supervision, Project administration, Funding acquisition.

Data Availability Statement

Raw data analysed in this study are deposited in the Sequence Read Archive (SRA) database (accession no. SAMN17174480–SAMN17174507 in the BioProject PRJNA641629).

References

- Beynon, J. and Josey, D. (1980) Demonstration of heterogeneity in a natural population of *Rhizobium phaseoli* using variation in intrinsic antibiotic resistance. *Microbiology* **118**, 437–442.
- van den Boogaart, K., Tolosana, R. and Bren, M. (2014) *compositions: Compositional Data Analysis*. R package version 1.40-1. Vienna, Austria: R Foundation for Statistical Computing.
- Chang, Y.L., Wang, J.Y., Wang, E.T., Liu, H.C., Sui, X.H. and Chen, W.X. (2011) *Bradyrhizobium lablabi* sp. nov., isolated from effective nodules of *Lablab purpureus* and *Arachis hypogaea*. *Int J Syst Evol Microbiol* **61**, 2496–2502.
- Charif, D. and Lobry, J.R. (2007) SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In *Structural Approaches to Sequence Evolution* ed. Bastolla, U., Porto, M., Roman, E. and Vendruscolo, M. pp. 207–232. Berlin: Springer Science & Business Media.
- Chen, I.A., Chu, K., Palaniappan, K., Ratner, A., Huang, J., Huntemann, M., Hajek, P., Ritter, S. *et al.* (2021) The IMG/M data management and analysis system v. 6.0: new tools and advanced capabilities. *Nucleic Acids Res* **49**, D751–D763.
- Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V. and Egozcue, J.J. (2017) Microbiome datasets are

- compositional: and this is not optional. *Front Microbiol* **8**, 2224.
- Gołębiewski, M. and Tretyn, A. (2020) Generating amplicon reads for microbial community assessment with next-generation sequencing. *J Appl Microbiol* **128**, 330–354.
- Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41**, 95–98.
- Herridge, D., Roughley, R. and Brockwell, J. (1984) Effect of rhizobia and soil nitrate on the establishment and functioning of the soybean symbiosis in the field. *Aust J Agric Res* **35**, 149–161.
- Horikoshi, M. and Tang, Y. (2018) Data visualization tools for statistical analysis results. v0. 1.0. <http://CRAN.R-project.org/package=ggfortify>. R package version 0.4 1, 28.
- Hungria, M., O'Hara, G., Zilli, J., Araujo, R., Deaker, R. and Howieson, J. (2016) Isolation and growth of rhizobia. In *Working with Rhizobia* ed. Howieson, J.G. and Dilworth, M.J. pp. 39–60. Canberra: Australian Centre for International Agricultural Research.
- Jensen, H. (1942) Nitrogen fixation in leguminous plants. I. General characters of root nodule bacteria isolated from species of *Medicago* and *Trifolium* in Australia. *Proc Linn Soc NSW* **66**, 98–108.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M. *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948.
- Li, Y.H., Wang, R., Zhang, X.X., Young, J.P.W., Wang, E.T., Sui, X.H. and Chen, W.X. (2015) *Bradyrhizobium guangdongense* sp. nov. and *Bradyrhizobium guangxiense* sp. nov., isolated from effective nodules of peanut. *Int J Syst Evol Microbiol* **65**, 4655–4661.
- Lindström, K. and Mousavi, S.A. (2020) Effectiveness of nitrogen fixation in rhizobia. *Microb Biotechnol* **13**, 1314–1335.
- McLaren, M.R., Willis, A.D. and Callahan, B.J. (2019) Consistent and correctable bias in metagenomic sequencing experiments. *Elife* **8**, e46923.
- Morton, J.T., Marotz, C., Washburne, A., Silverman, J., Zaramela, L.S., Edlund, A., Zengler, K. and Knight, R. (2019) Establishing microbial composition measurement standards with reference frames. *Nat Commun* **10**, 1–11.
- Nick, G., Jussila, M., Hoste, B., Niemi, R.M., Kaijalainen, S., de Lajudie, P., Gillis, M., de Bruijn, F.J. *et al.* (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.
- Palarea-Albaladejo, J. and Martín-Fernández, J.A. (2015) zCompositions: R package for multivariate imputation of left-censored data under a compositional approach. *Chemometr Intell Lab Syst* **143**, 85–96.
- Paradis, E., Claude, J. and Strimmer, K. (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290.
- Pitkäjärvi, J., Räsänen, L.A., Langenskiöld, J., Wallenius, K., Niemi, M. and Lindström, K. (2003) Persistence, population dynamics and competitiveness for nodulation of marker gene-tagged *Rhizobium galegae* strains in field lysimeters in the boreal climatic zone. *FEMS Microbiol Ecol* **46**, 91–104.
- Revell, L.J. and Chamberlain, S.A. (2014) Rphylop: an R interface for PHYLIP. *Methods Ecol Evol* **5**, 976–981.
- Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B. *et al.* (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**, 7537–7541.
- Sikora, S. and Redžepović, S. (2003) Genotypic characterisation of indigenous soybean rhizobia by PCR-RFLP of 16S rDNA, rep-PCR and RAPD analysis. *Food Technol Biotechnol* **41**, 61–67.
- Somervuo, P., Koskinen, P., Mei, P., Holm, L., Auvinen, P. and Paulin, L. (2018) BARCOSEL: a tool for selecting an optimal barcode set for high-throughput sequencing. *BMC Bioinform* **19**, 257.
- Tang, Y., Horikoshi, M. and Li, W. (2016) ggfortify: unified interface to visualize statistical results of popular R packages. *R J* **8**, 478–489.
- Tas, E., Kaijalainen, S., Saano, A. and Lindström, K. (1994) Isolation of a *Rhizobium galegae* strain-specific DNA probe. *Microb Releases* **2**, 231–237.
- Tas, E., Leinonen, P., Saano, A., Rasanen, L.A., Kaijalainen, S., Piippola, S., Hakola, S. and Lindstrom, K. (1996) Assessment of competitiveness of rhizobia infecting *Galega orientalis* on the basis of plant yield, nodulation, and strain identification by antibiotic resistance and PCR. *Appl Environ Microbiol* **62**, 529–535.
- Tas, E., Saano, A., Leinonen, P. and Lindstrom, K. (1995) Identification of *Rhizobium* spp. in peat-based inoculants by DNA hybridization and PCR and its application in inoculant quality control. *Appl Environ Microbiol* **61**, 1822–1827.
- Vincent, J.M. (1970) *A Manual for the Practical Study of the Root-nodule Bacteria*. (International Biological Programme Handbook 15). Oxford: Blackwell.
- Vos, M., Quince, C., Pijl, A.S., de Hollander, M. and Kowalchuk, G.A. (2012) A comparison of *rpoB* and 16S rRNA as markers in pyrosequencing studies of bacterial diversity. *PLoS One* **7**, e30600.
- Wang, R., Chang, Y.L., Zheng, W.T., Zhang, D., Zhang, X.X., Sui, X.H., Wang, E.T., Hu, J.Q., *et al.* (2013) *Bradyrhizobium arachidis* sp. nov., isolated from effective nodules of *Arachis hypogaea* grown in China. *Syst Appl Microbiol* **36**, 101–105.
- Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.
- Winand, R., Bogaerts, B., Hoffman, S., Lefevre, L., Delvoeye, M., Van Braekel, J., Fu, Q., Roosens, N.H. *et al.* (2020)

- Targeting the 16S rRNA gene for bacterial identification in complex mixed samples: comparative evaluation of second (Illumina) and third (Oxford Nanopore Technologies) generation sequencing technologies. *Int J Mol Sci* **21**, 298.
- Wright, E.S. (2016) Using DECIPHER v2. 0 to analyze big biological sequence data in R. *R J* **8**, 352.
- Yates, R., Howieson, J., Hungria, M., Bala, A., O'Hara, G. and Terpolilli, J. (2016) Authentication of rhizobia and assessment of the legume symbiosis in controlled plant growth systems. In *Working with Rhizobia* ed. Howieson, J.G. and Dilworth, M.J. pp. 73–108. Canberra: Australian Centre for International Agricultural Research.
- Zhang, X.X., Guo, H.J., Jiao, J., Zhang, P., Xiong, H.Y., Chen, W.X. and Tian, C.F. (2017) Pyrosequencing of *rpoB* uncovers a significant biogeographical pattern of rhizobial species in soybean rhizosphere. *J Biogeogr* **44**, 1491–1499.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Principal component analysis of *rpoB* communities in peanut nodules.

Figure S2. Sequence diversity of bradyrhizobial strains.

Table S1. Sequence identities and relative abundances of the *rpoB* gene amplicons in the growth chamber experiment.

Table S2. Sequence identities and relative abundances of the *rpoB* gene amplicons in the local farms in Linshu, China.

Table S3. Sequence identities and relative abundances of the *rpoB* gene amplicons in the field experiment.

Table S4. List of strains used in growth chamber treatments.

Table S5. Nodule samples collected from local farms and the place of isolation.

Table S6. Annealing temperature and the number of cycles of the first round of PCR.

Table S7. 16S rRNA gene sequences subjected to *in silico* PCR.

Table S8. Primers applied in *in silico* PCR.