Appl Microbiol Biotechnol (2015) 99:5547–5562
DOI 10.1007/s00253-015-6515-3

Ribosomal protein biomarkers provide root nodule bacterial identification by MALDI-TOF MS

Dominik Ziegler · Joël F. Pothier · Julie Ardley · Romain Kouakou Fossou · Valentin Pflüger · Sofie de Meyer · Guido Vogel · Mauro Tonolla · John Howieson · Wayne Reeve · Xavier Perret

Received: 12 January 2015 / Revised: 19 February 2015 / Accepted: 28 February 2015 / Published online: 18 March 2015

Abstract Accurate identification of soil bacteria that form nitrogen-fixing associations with legume crops is challenging given the phylogenetic diversity of root nodule bacteria (RNB). The labor-intensive and time-consuming 16S ribosomal RNA (rRNA) sequencing and/or multilocus sequence analysis (MLSA) of conserved genes so far remain the favored molecular tools to characterize symbiotic bacteria. With the development of mass spectrometry (MS) as an alternative method to rapidly identify bacterial isolates, we recently showed that matrix-assisted laser desorption ionization (MALDI) time-of-flight (TOF) can accurately characterize RNB found inside plant nodules or grown in cultures. Here, we report on the development of a MALDI-TOF RNB-specific spectral database built on whole cell MS fingerprints of 116 strains representing the major rhizobial genera. In addition to this RNB-specific module, which was successfully tested on unknown field isolates, a subset of 13 ribosomal proteins extracted from genome data was found to be sufficient for the reliable identification of nodule isolates to rhizobial species as shown in the putatively ascribed ribosomal protein masses (PARPM) database. These results reveal that data gathered from genome sequences can be used to expand spectral libraries to aid the accurate identification of bacterial species by MALDI-TOF MS.

Keywords Bacterial fingerprints · Phylogeny · Cluster analysis · Rhizobia · Legume nodules · GEBA-RNB

Electronic supplementary material The online version of this article (doi:10.1007/s00253-015-6515-3) contains supplementary material, which is available to authorized users.

D. Ziegler · R. K. Fossou · M. Tonolla · X. Perret (✉)
Department of Botany and Plant Biology, Microbiology Unit, Sciences III, University of Geneva, 30 quai Ernest-Ansermet, CH-1211 Geneva 4, Switzerland
e-mail: xavier.perret@unige.ch

D. Ziegler · V. Pflüger · G. Vogel
Mabritec, Lörracherstrasse 50, CH-4125 Riehen, Switzerland

J. F. Pothier
Environmental Genomics and Systems Biology Research Group, Institute for Natural Resources Sciences, Zurich University of Applied Sciences (ZHAW), Grüental, P.O. Box, CH-8820 Wädenswil, Switzerland

J. Ardley · S. de Meyer · J. Howieson · W. Reeve
Center for Rhizobium Studies, Murdoch University, Murdoch, Australia

M. Tonolla
Laboratory of Applied Microbiology, Department of Environment Construction and Design, University of Applied Sciences of Southern Switzerland, via Mirasole 22a, CH-6500 Bellinzona, Switzerland

Introduction

Soil bacteria, collectively known as Rhizobia or root nodule bacteria (RNB), form beneficial nitrogen-fixing associations with most leguminous plants, including important crops such as soybean (*Glycine max*), bean (*Phaseolus vulgaris*), pea (*Pisum sativum*), chickpea (*Cicer arietinum*), pigeon pea (*Cajanus cajan*), and pastures such as clovers (*Trifolium* spp.) or *Medicago* spp. These associations provide significant sustainable inputs of reduced nitrogen (N) to agricultural systems that are estimated to be ca. 50–70 Tg/year (Herridge et al. 2008). A key aspect of symbiotic nitrogen fixation is the development of specialized structures called nodules on the roots (or more rarely stems) of host plants. Within the nodule cells, RNB differentiate into bacteroids, which reduce atmospheric nitrogen (N_2) to ammonia (Oldroyd and Downie 2008; Udvardi and Poole 2013). Rhizobial infection and nodule organogenesis require a coordinated exchange of molecular

signals between the host and its symbiotic bacteria (Gibson et al. 2008; Perret et al. 2000).

Although most of the approximately 20,000 species of the *Leguminosae* are able to nodulate (Doyle 2011), only a small fraction of their rhizobial associations have been characterized. In contrast to legumes, which are monophyletic (Werner et al. 2014), molecular techniques have demonstrated that rhizobia are polyphyletic, belonging to diverse bacterial genera of the *Alphaproteobacteria* and *Betaproteobacteria* (Masson-Boivin et al. 2009). In RNB, symbiotic ability is conferred by genes for nodulation and nitrogen fixation that can be acquired via lateral transfer of plasmids or genomic islands (Broughton 2003; Freiberg et al. 1997; Moulin et al. 2004; Nandasena et al. 2007; Sullivan and Ronson 1998). To date, 16 rhizobial genera and over 100 species have been validly described (ICSP Subcommittee on the taxonomy of *Rhizobium* and *Agrobacterium*—diversity, phylogeny, and systematics; http://edzna.ccg.unam.mx/rhizobial-taxonomy/node/4), and these numbers are expected to increase as more symbioses are studied.

Phenotypic methods of identifying and typing bacteria, such as colony morphology and serotyping, have largely been superseded by molecular markers. Sequence analysis of the 16S ribosomal RNA (rRNA) gene allows the construction of valid bacterial phylogenies (Schleifer 2009), but lacks resolving power at and below the species level, due to the extent of gene conservation (Willems 2006). Multilocus sequence analysis (MLSA) of conserved core genes was proposed as a robust and reliable alternative method for classifying bacteria to species level (Gevers et al. 2005; Hanage et al. 2006) and has been successfully used to delineate species within a number of genera including *Ensifer*, formerly *Sinorhizobium* (Martens et al. 2008), and *Bradyrhizobium* (Rivas et al. 2009). However, incongruent phylogenies have been reported among several genes commonly used in MLSA, along with evidence of lateral gene transfer or intergenic recombination, resulting in unclear taxonomic resolution (Tian et al. 2012; Zhang et al. 2012). No single genealogical reconstruction or typing method currently encompasses all levels of bacterial diversity, from domain to strain. Moreover, molecular marker analysis is costly and laborious, making it unsuitable for use in high throughput, broad-scale studies of rhizobial taxonomy, phylogeny, biogeography, or population genetics.

Alternatively, matrix-assisted laser desorption ionization (MALDI) time-of-flight (TOF) mass spectrometry (MS) has increasingly replaced DNA-based procedures for the reliable identification of bacteria, in particular, in diagnostic laboratories responsible for characterizing clinical isolates (Seng et al. 2009). MALDI-TOF MS offers the possibility of accurately resolving bacterial identity to the genus, species, and subspecies levels in some taxa and has successfully been applied to diverse rhizobial strains grown on solid media (Ferreira et al. 2011; Sánchez-Juanes et al. 2013; Ziegler et al. 2012) and found inside legume nodules (Ziegler et al. 2012). In most direct MALDI-TOF MS analyses, bacterial cells are lysed to release the intracellular proteins, which are then ionized and separated according to their mass–to-charge ratio (*m*/*z*) and recorded as distinct peaks, generating a complex spectrum, or fingerprint that is characteristic of a bacterial sample. In general, microbial MALDI-TOF MS signatures can be used to identify unknown bacterial samples down to the subspecies level, providing they can be matched against comprehensive databases of empirical spectra obtained for sets of well-characterized (but not necessarily type) strains selected to represent the microbial genera of interest. For example, the enlarged Spectral ARchive And Microbial Identifications System (SARAMIS™ ver. 4.10, system ver. 3.4.1.2; Mabritec, Riehen, Switzerland) database contains characteristic signatures for more than 650 bacterial species, mostly of clinical importance.

MALDI-TOF MS biomarkers do not need to be linked to specific proteins to ensure correct identification of bacterial strains. However, many of the individual peaks that form a MALDI-TOF mass spectrum correspond to ribosomal proteins (RPs), which are encoded by more than 50 genes scattered in mostly chromosomal loci. RPs are ubiquitous and among the most abundant cytosolic proteins, regardless of the cell life stage or growth conditions (Ishihama et al. 2008), and have masses mostly within the 4 to 30 kDa range that is examined with MALDI-TOF MS. Ribosomal multilocus sequence typing (rMLST) has been shown to be consistent with current bacterial nomenclature schemes and to provide accurate phylogenies for diverse taxonomic groups (Jolley et al. 2012; Maiden et al. 2013; Yutin et al. 2012). These characteristics make RPs reliable biomarkers in MALDI-TOF MS analyses to obtain species and subspecies identification of diverse bacteria including bifidobacteria (Sato et al. 2011), lactobacilli (Sato et al. 2012; Teramoto et al. 2007a), rhodococci (Teramoto et al. 2009), members of the *Sphingomonadaceae* (Hotta et al. 2012), and strains of *Staphylococcus aureus* (Josten et al. 2013) and *Neisseria meningitidis* (Suarez et al. 2013).

Additionally, the calculated masses of RPs deduced from partial or whole genome sequences were shown to be sufficient for correctly identifying unknown bacteria, even in the absence of empirical mass spectral fingerprints from a corresponding reference microorganism (Pineda et al. 2003). Together, these results suggested that empirical spectra of strains assembled into a reference database, such as SARAMIS™, could be supplemented by lists of biomarkers corresponding to masses of RPs calculated from the sequenced genomes of reference strains.

Given the potential for MALDI-TOF MS to facilitate and expedite the exploration of rhizobial diversity in agricultural and natural ecosystems (Ziegler et al. 2012), our first objective was to establish a comprehensive empirical

Table 1 List of 114 Rhizobia and two tumorigenic strains used to generate the reference database of MALDI-TOF MS spectra for the identification of nodule bacteria

#	Genus	Species	Strain	Isolated from	GOLD ID	Source
1	*Agrobacterium*	*A. tumefaciens*	C58-UWash	Cherry tree tumor	Gc00073	J.F. Pothier
2		*A. vitis*	S4	*Vitis vinifera*	Gc00943	J.F. Pothier
3	*Azorhizobium*	*A. caulinodans*	ORS 571^T	*Sesbania rostrata*	Gc00669	S. Goormachtig
4		*A. doebereinerae*	UFLA1-100^T	*Sesbania virgate*	Gi08849	W.G. Reeve
5	*Bradyrhizobium*	*B. canariense*	WSM4349	*Syrmatium glabrum*	Gi08886	W.G. Reeve
6		*B. elkanii*	USDA 76^T	*Glycine max*	Gi08850	W.G. Reeve
7		*B. elkanii*	USDA 94	*G. max*	Gi10733	W.G. Reeve
8		*Bradyrhizobium* sp.	WSM2783	*Leobordea carinata*	Gi08855	W.G. Reeve
9		*B. elkanii*	USDA 3259	*Phaseolus lunatus*	Gi22151	W.G. Reeve
10		*B. elkanii*	USDA 3254	*Phaseolus acutifolius*	Gi08852	W.G. Reeve
11		*Bradyrhizobium* sp.	WSM1741	*Rhynchosia minima*	Gi08854	W.G. Reeve
12		*B. japonicum*	USDA 6^T	*G. max*	Gc02045	W.G. Reeve
13		*B. japonicum* (*B. diazoefficiens*)	USDA 110	*G. max*	Gc00116	X. Perret
14		*B. japonicum* (*B. diazoefficiens*)	USDA 122	*G. max*	Gi08856	W.G. Reeve
15		*Bradyrhizobium* sp.	USDA 135	*G. max*	Gi08859	W.G. Reeve
16		*Bradyrhizobium* sp.	USDA 38	*G. max*	Gi08860	W.G. Reeve
17		*Bradyrhizobium* sp.	USDA 4	*G. max*	Gi08861	W.G. Reeve
18		*Bradyrhizobium* sp.	WSM1743	*Indigofera* spp.	Gi08863	W.G. Reeve
19		*Bradyrhizobium* sp.	WSM2793	*Rhynchosia totta*	Gi08908	W.G. Reeve
20		*Bradyrhizobium* sp.	ARR65	*Stylosanthes viscosa*	Gi08868	W.G. Reeve
21		*Bradyrhizobium* sp.	BTAi1	*Aeschynomene indica*	Gc00561	E. Giraud
22		*Bradyrhizobium* sp.	ORS285	*Aeschynomene* spp.	Gi09732	E. Giraud
23		*Bradyrhizobium* sp.	ORS278	*Aeschynomene* spp.	Gc00552	E. Giraud
24		*Bradyrhizobium* sp.	USDA 3384	*Crotalaria paulina*	Gi10733	W.G. Reeve
25		*B. canariense*	WSM1253	*Omithopus compressus*	Gi06492	W.G. Reeve
26		*B. canariense*	WSM1417	*Lupinus* spp.	Gi06490	W.G. Reeve
27		*Bradyrhizobium* sp.	WSM2254	*Acacia* spp.	Gi14039	W.G. Reeve
28		*Bradyrhizobium* sp.	WSM3983	*Kennedia coccinea*	Gi08871	W.G. Reeve
29		*B. canariense*	WSM471	*Lupinus angustifolius*	Gi06491	W.G. Reeve
30	*Burkholderia*	*B. mimosarum*	LMG 23256^T	*Mimosa pigra*	Gi08823	W.G. Reeve
31		*Burkholderia* sp.	WSM2230	*K. coccinea*	Gi08831	W.G. Reeve
32		*B. dilworthii*	WSM3556^T	*Lebeckia ambigua*	Gi08872	W.G. Reeve
33		*B. tuberum*	WSM4176	*L. ambigua*	Gi08873	W.G. Reeve
34		*Burkholderia* sp.	JPY347	*Mimosa cordistipula*	Gi08875	W.G. Reeve
35		*Burkholderia* sp.	JPY251	*Mimosa velloziana*	Gi08874	W.G. Reeve
36		*Burkholderia* sp.	UYPR1.413	*Parapiptadenia rigida*	Gi08829	W.G. Reeve
37		*B. sprentiae*	WSM5005^T	*L. ambigua*	Gi06497	W.G. Reeve
38	*Cupriavidus*	*Cupriavidus* sp.	UYPR2.512	*P. rigida*	Gi08830	W.G. Reeve
39		*C. taiwanensis*	LMG 19424^T	*Mimosa pudica*	Gc00754	M. M. Saad
40		*C. taiwanensis*	STM6070	*M. pudica*	Gi08841	W.G. Reeve
41	*Ensifer*	*E. arboris*	LMG 14919^T	*Prosopis chilensis*	Gi08822	W.G. Reeve
42		*E. fredii*	HH103	*G. max*	Gc02113	J.E. Ruiz-Sainz
43		*E. fredii*	USDA257	*G. max*	Gc02300	W.G. Reeve
44		*E. fredii*	NGR234	*Lablab purpureus*	Gc00990	X. Perret
45		*E. medicae*	WSM419	*Medicago murex*	Gc00590	W.G. Reeve
46		*E. medicae*	Di28	*Medicago arabica*	Gi08905	W.G. Reeve
47		*E. medicae*	WSM1115	*Medicago polymorpha*	Gi08906	W.G. Reeve
48		*E. medicae*	WSM1369	*Medicago sphaerocarpos*	Gi08907	W.G. Reeve

Table 1 (continued)

#	Genus	Species	Strain	Isolated from	GOLD ID	Source
49		*E. medicae*	WSM244	*M. polymorpha*	Gi08908	W.G. Reeve
50		*E. medicae*	WSM4191	*Melilotus siculus*	Gi08903	W.G. Reeve
51		*E. meliloti*	AK83	*Medicago falcata*	Gc01810	A. Mengoni
52		*E. meliloti*	Mlalz-1	*Medicago laciniata*	Gi08913	W.G. Reeve
53		*E. meliloti*	CIAM1775	*Medicago lupulina*	Gi08844	W.G. Reeve
54		*E. meliloti*	WSM1022	*Medicago orbicularis*	Gi08916	W.G. Reeve
55		*E. meliloti*	1021	*Medicago sativa*	Gc00059	X. Perret
56		*E. meliloti*	BL225C	*M. sativa*	Gc01811	A. Mengoni
57		*E. meliloti*	MVII-I	*M. sativa*	Gi08914	W.G. Reeve
58		*E. meliloti*	RRl128	*Medicago truncatula*	Gi08915	W.G. Reeve
59		*E. meliloti*	4H41	*Phaseolus vulgaris*	Gi08911	W.G. Reeve
60		*E. meliloti*	GVPV12	*P. vulgaris*	Gi08912	W.G. Reeve
61		*Ensifer* sp.	TW10	*Tephrosia wallichii*	Gi08835	W.G. Reeve
62		*Ensifer* sp.	WSM1721	*Indigofera* spp.	Gi08904	W.G. Reeve
63	*Mesorhizobium*	*M. australicum*	WSM2073^T	*Biserrula pelecinus*	Gc02468	W.G. Reeve
64		*M. ciceri*	WSM4083	*Bituminaria bituminosa*	Gi08880	W.G. Reeve
65		*M. ciceri*	CMG6	*Cicer arietinum*	Gi08879	W.G. Reeve
66		*M. ciceri* bv. *Biserrulae*	WSM1271	*B. pelecinus*	Gc01578	W.G. Reeve
67		*M. loti*	USDA 3471^T	*Lotus corniculatus*	Gi08881	W.G. Reeve
68		*M. loti*	CJ3sym	*L. corniculatus*	Gi08828	W.G. Reeve
69		*M. loti*	MAFF303099	*Lotus japonicus*	Gc00040	X. Perret
70		*M. loti*	NZP2037	*Lotus divaricatus*	Gi08826	W.G. Reeve
71		*M. loti*	R7A	*L. corniculatus*	Gi08825	W.G. Reeve
72		*M. loti*	R88b	*L. corniculatus*	Gi08827	W.G. Reeve
73		*M. loti*	WSM1293	*Lotus* spp.	Gi08882	W.G. Reeve
74		*M. opportunistum*	WSM2075^T	*B. pelecinus*	Gc01853	W.G. Reeve
75		*Mesorhizobium* sp.	WSM2561	*Lessertia diffusa*	Gi08883	W.G. Reeve
76		*Mesorhizobium* sp.	WSM3626	*L. diffusa*	Gi08885	W.G. Reeve
77		*Mesorhizobium* sp.	WSM3224	*Otholobium candicans*	Gi08884	W.G. Reeve
78	*Methylobacterium*	*M. nodulans*	ORS 2060^T	*Crotalaria podocarpa*	Gc00935	E. Giraud
79		*Methylobacterium* sp.	WSM2598	*Listia bainesii*	Gi08887	W.G. Reeve
80	*Microvirga*	*M. lotononidis*	WSM3557^T	*Listia angolensis*	Gi06493	W.G. Reeve
81		*M. lupini*	Lut6^T	*Lupinus texensis*	Gi06478	W.G. Reeve
82	*Rhizobium*	*R. etli*	CFN 42^T	*P. vulgaris*	Gc00342	S. Brom
83		*R. phaseoli*	CIAT 652	*P. vulgaris*	Gc00823	S. Brom
84		*R. giardinii* bv. *giardinii*	H152^T	*P. vulgaris*	Gi08897	W.G. Reeve
85		*R. leguminosarum* bv. *phaseoli*	4292	*P. vulgaris*	Gi08895	W.G. Reeve
86		*R. leguminosarum* bv. *phaseoli*	FA23	*P. vulgaris*	Gi08821	W.G. Reeve
87		*R. leguminosarum* bv. *trifolii*	TA1	*Trifolium spurmosum*	Gi06488	W.G. Reeve
88		*R. leguminosarum* bv. *trifolii*	WSM2297	*Trifolium africanum*	Gi06477	W.G. Reeve
89		*R. leguminosarum* bv. *trifolii*	CC283b	*Trifolium ambiguum*	Gi06484	W.G. Reeve
90		*R. leguminosarum* bv. *trifolii*	CC278f	*Trifolium nannum*	Gi06479	W.G. Reeve
91		*R. leguminosarum* bv. *trifolii*	WSM597	*Trifolium pallidum*	Gi06486	W.G. Reeve
92		*R. leguminosarum* bv. *trifolii*	WSM2304	*Trifolium polymorpha*	Gc00870	W.G. Reeve
93		*R. leguminosarum* bv. *trifolii*	WSM2012	*Trifolium ruepellianum*	Gi06480	W.G. Reeve
94		*R. leguminosarum* bv. *trifolii*	WSM1325	*Trifolium* spp.	Gc01039	W.G. Reeve
95		*R. leguminosarum* bv. *trifolii*	SRDI565	*Trifolium subterraneum*	Gi08843	W.G. Reeve
96		*R. leguminosarum* bv. *trifolii*	SRDI943	*T. subterraneum*	Gi08842	W.G. Reeve
97		*R. leguminosarum* bv. *trifolii*	WSM1689	*Trifolium uniflorum*	Gi06499	W.G. Reeve

Table 1 (continued)

#	Genus	Species	Strain	Isolated from	GOLD ID	Source
98		*R. leguminosarum* bv. *trifolii*	CB782	*Trifolium semipilosum*	Gi06498	W.G. Reeve
99		*R. leguminosarum* bv. *viciae*	128C53	*Pisum sativum*	Gi08894	W.G. Reeve
100		*R. leguminosarum* bv. *viciae*	3841	*P. sativum*	Gc00385	P.S. Poole
101		*R. leguminosarum* bv. *viciae*	GB30	*P. sativum*	Gi08820	W.G. Reeve
102		*R. leguminosarum* bv. *viciae*	Ps8	*P. sativum*	Gi08888	W.G. Reeve
103		*R. leguminosarum* bv. *viciae*	UPM1131	*P. sativum*	Gi08837	W.G. Reeve
104		*R. leguminosarum* bv. *viciae*	UPM1137	*P. sativum*	Gi08836	W.G. Reeve
105		*R. leguminosarum* bv. *viciae*	TOM	*P. sativum*	Gi08893	W.G. Reeve
106		*R. leguminosarum* bv. *viciae*	Vc2	*Vicia cracca*	Gi08889	W.G. Reeve
107		*R. leguminosarum* bv. *viciae*	248	*Vicia faba*	Gi08896	W.G. Reeve
108		*R. leguminosarum* bv. *viciae*	VF39	*V. faba*	Gi08892	W.G. Reeve
109		*R. leguminosarum* bv. *viciae*	WSM1455	*V. faba*	Gi06482	W.G. Reeve
110		*R. leguminosarum* bv. *viciae*	WSM1481	*V. faba*	Gi08891	W.G. Reeve
111		*R. leguminosarum* bv. *viciae*	Vh3	*Vicia hirsuta*	Gi08890	W.G. Reeve
112		*R. mongolense*	USDA 1844^T	*Medicago ruthenica*	Gi08900	W.G. Reeve
113		*Ensifer*[a] sp.	BR816	*Leucaena leucocephala*	Gi08910	W.G. Reeve
114		*R. mesoamericanum*	STM6155	*M. pudica*	Gi08838	W.G. Reeve
115		*Rhizobium* sp.	OR 191	*M. sativa*	Gi08824	W.G. Reeve
116		*R. sullae*	WSM1592	*Hedysarum coronarium*	Gi08899	W.G. Reeve

^{T} type strains. Reference strains were obtained from Wayne G. Reeve (Center for *Rhizobium* Studies, Murdoch University, Perth, Australia); Maged M. Saad (Microbiology Unit, University of Geneva, Geneva, Switzerland); Eric Giraud (Laboratoire des Symbioses Tropicales et Méditéranéennes, Montpellier, France); Julio E. Ruiz-Sainz (Departamento de Microbiología, Universidad de Sevilla, Spain); Susana Brom (Center for Genomic Sciences, Universidad Nacional Autonoma de Mexico, Cuernavaca, Mexico); Sofie Goormachtig (Department of Plant Biotechnology and Genetics, Ghent University, Ghent, Belgium); Alessio Mengoni (Department of Animal Biology and Genetics, Florence University, Florence, Italy); Philip S. Poole (John Innes Center, Norwich, UK); and Joël F. Pothier (ZHAW Wädenswil, Wädenswil, Switzerland). Sequenced genomes were accessed as genome online database identifiers (GOLD ID; https://gold.jgi-psf.org).

[a] Due to changes in taxonomy, strain BR816 has previously been identified as *Rhizobium*, *Sinorhizobium*, and *Ensifer*. The following two strains were recently reclassified as new type strains: *Bradyrhizobium diazoefficiens* USDA110^T (Delamuta et al. 2013) and *Burkholderia dilworthii* WSM3556^T (De Meyer et al. 2014)

reference database for unambiguously assigning RNB to known rhizobial genera. To this end, two strains of plant-tumorigenic and 114 strains of plant-symbiotic bacteria that were representative of the major rhizobial genera were used to construct the RNB-specific module extension to the SARAMIS™ database. As the genomes of these 116 selected strains had been sequenced, mostly as part of the Genomic Encyclopedia of Bacteria and Archaea-Root Nodule Bacteria (GEBA-RNB) project, we expanded the empirical RNB-specific module with additional SARA MIS™-compatible databases of biomarkers that corresponded to (a) the deduced ribosomal protein masses (DRPM) calculated from genomic data and (b) a subset of putatively ascribed ribosomal protein masses (PARPM) that matched mass peaks in the empirical spectra. The DRPM and PARPM databases could be used to provide accurate identification of rhizobial isolates, suggesting that lists of biomarker masses deduced from genomic data may complement the databases of empirical spectra obtained from pure culture of bacteria.

Materials and methods

Acquisition of root nodules and bacteria

Reference strains used in this study are listed in Table 1 together with some of their major characteristics including the name of the host plant they were isolated from and the genome online database identifier (GOLD ID) number under which sequence data was stored. Root-nodulating and tumorigenic bacteria were grown in/on TY (Beringer 1974) or YM (Vincent 1970) media and incubated at 28 °C as growth temperature. Nodules were collected from field-grown legumes at locations in Western Australia (Brookton and Esperance) and Ivory Coast (Bondoukou and Kossou-Bouafla) and kept desiccated until further use. For analysis of nodule bacteria, desiccated nodules were rehydrated 30 min prior to surface sterilization with 70 % EtOH and 4 % bleach, with respective incubation times adjusted for nodule size. Once thoroughly rinsed with sterile H_2O, surface sterilized nodules were then processed as described by Ziegler et al. (2012).

Determining symbiotic properties of nodule isolates

Nodulation assays were carried out in Magenta jars containing vermiculite (Lewin et al. 1987) with B&D nitrogen-free nutritive solution (Broughton and Dilworth 1971). Once surface-sterilized, seeds of *C. cajan* (pigeon pea) and *Vigna unguiculata* cv. Red Caloona (cowpea) were incubated several days in the dark at 27 °C for germination. Seedlings were then planted in autoclaved Magenta jars, and 2–3 days later each plantlet was inoculated with 200 μl of a solution containing 2×10^8 bacteria. Once inoculated, plants were grown for 6 weeks at a day temperature of 27 °C, a night temperature of 20 °C and a light phase of 12 h prior to harvesting of nodules.

Preparation of samples for MALDI-TOF MS analyses

For generating spectra for strain identification and/or construction of the Rhizobia-specific module of SARAMIS™, free-living bacteria were harvested after 1, 2, 3, 5, and 7 days growth on TY (Beringer 1974) or YM (Vincent 1970) plates incubated at 28 °C. Bacterial samples were spotted in quadruplicate on MALDI steel target plates. Spots were overlaid with 1 μl of 25 % formic acid, air-dried, and again overlaid with 1 μl of matrix solution consisting of saturated solution of alpha-cyano-4 hydroxycinnamic acid (CHCA; Sigma-Aldrich, Buchs, Switzerland) in 33 % acetonitrile (Sigma-Aldrich), 33 % ethanol and supplemented with 3 % trifluoroacetic acid (TFA). Alternatively, sinapinic acid (SA) diluted in 60 % acetonitrile and 0.3 % TFA was used as matrix. Finally, all samples processed in these ways were air-dried for a few minutes at room temperature. For ascribing ribosomal proteins to specific mass peaks in empirical spectra, cell lysates were used instead of whole bacteria. Cell lysates were obtained as follows: after growth on the appropriate medium cells were collected and washed three times in TMA-I buffer (10 mM Tris–HCl pH 7.8, 30 mM NH_4Cl, 10 mM $MgCl_2$, and 6 mM 2-mercaptoethanol). Once resuspended in 500 μl TMA-I buffer, cells were ground three times for 20 s at 600 rpm using 0.1 mm glass beads in a FastPrep FP120 bead beater cell disrupter (Farmingdale, NY, USA). Glass beads and cell debris were removed by centrifugation at 5,800×*g* for 10 min and aliquots of the supernatant spotted for MALDI-TOF MS measurements as described above and using SA as matrix.

Acquisition of MALDI-TOF mass spectra and data processing

Empirical spectra were obtained using a MALDI-TOF Mass Spectrometer Axima™ Confidence machine (Shimadzu-Biotech, Kyoto, Japan), with detection in the linear positive mode at a laser frequency of 50 Hz and within a mass range of 3 to 20 kDa or 4 to 30 kDa when CHCA or SA were used as matrices, respectively. Acceleration voltage was 20 kV, and the extraction delay time was 200 ns. A minimum of ten laser shots per sample was used to generate each ion spectrum. For each bacterial sample, a total of 50 to 100 protein mass fingerprints were averaged and processed using the Launchpad™ 2.8 software (Shimadzu-Biotech). For peak processing of raw spectra, the following settings were selected for Launchpad™ 2.8: (i) the advanced scenario from the Parent Peak Cleanup menu, (ii) a peak width set at 80 chans, (iii) the smoothing filter width set at 50 chans, (iv) a baseline filter width of 500 chans, and (v) a threshold apex method set as dynamic. The threshold offset was set at 0.020 mV with a response factor of 1.2. Each target plate was externally calibrated using spectra of the reference strain *Escherichia coli* DH5α (Invitrogen, Carlsbad, USA).

Constructing the databases of deduced and putatively ascribed ribosomal protein masses (DRPM and PARPM)

Genomes of the 116 selected RNB or tumorigenic bacteria were accessed at the US Department of Energy's (DoE) Joint Genome Institute (JGI) or the National Center for Biotechnology Information (NCBI) using the strain-specific GOLD ID listed in Table 1. DNA sequences annotated as coding for RPs were collected, aligned, and checked for correct start and stop codons. Genes coding for RPs that were either misannotated or missing from a given genome were manually curated using corresponding sequences from closely related strains as in silico probes. DNA sequences verified in this way and coding for a complete set of 54 ribosomal proteins for each of the 116 selected strains were translated into corresponding amino acid sequences. Masses were then deduced taking into account the most frequent posttranslational modifications in bacteria (Arnold and Reilly 1999). Thus, 131 Da corresponding to the loss of the N-terminal methionine was subtracted from the deduced ribosomal protein mass when the penultimate amino acid was found to be either Gly, Ala, Pro, Ser, Thr, Val, or Cys. With the exception of *Burkholderia* strains for which empirical spectra showed no such modifications, ribosomal protein L33 was always considered to be methylated. For each strain, a text file compatible with the SARAMIS™ database and containing the relevant taxonomical information and deduced masses for all 54 RPs together with arbitrary intensity values was created and stored into a separate DRPM database. Using a possible error interval of 800 ppm, this dataset was matched against empirical spectra to ultimately identify among mass peaks those that corresponded to RPs. Mass peaks putatively ascribed to RPs L20, L24, L28, L29, L31, L32, L33, L36, S10, S15, S16, S17, and S20 were found to be consistently present in empirical spectra generated for each of the 116 reference strains. For each strain, mass peaks ascribed to each of these 13 marker RPs and intensity values were compiled into a SARAMIS™-compatible text file that was stored into a

distinct subdatabase called PARPM for putatively ascribed ribosomal proteins masses.

Genetic analyses

Sequences of 16S rRNA, *recA*, *gyrB*, *rpoB*, *dnaJ*, and *atpD* genes or ribosomal protein genes (*rplT*, *rplX*, *rpmB*, *rpmC*, *rpmE*, *rpmF*, *rpmG*, *rpmJ*, *rpsJ*, *rpsO*, *rpsP*, *rpsQ*, and *rpsT*) were accessed using the NCBI (www.ncbi.nlm.nih.gov) and JGI (genome.jgi.doe.gov) databases. As full 16S rRNA gene sequences were missing from the draft genomes of *Rhizobium leguminosarum* bv. *viciae* strains 128C53, Ps8, and Vh3, as well as strain *Rhizobium sullae* WSM1592, this gene was amplified and sequenced. Genomic DNA was extracted using the InstaGene™ kit (Bio-Rad Laboratories, Hercules, CA, USA). Primers nucleotide sequence 5′-ATTCTAGAGT TTGATCATGGCTCA-3′ (UNI-L) and nucleotide sequence 5′-ATGGTACCGTGTGACGGGCGGTGTGTA-3′ (UNI-R) were then used to amplify and sequence a fragment of ca. 1,400 bp (Brosius et al. 1978). Sequencing reactions were carried out using Big DyeTerminator v3.1 Cycle Sequencing Kit and results were analyzed using the 3500 Data Collection software v.1.00 (Applied Biosystems, Foster City, CA, USA).

Phylogenetic analyses

For each locus of interest, DNA sequences were trimmed to the same length and aligned using the MEGA 5 software package (Tamura et al. 2011). Phylogenetic trees were reconstructed using the maximum composite likelihood method and default settings. For hierarchical clustering of protein masses, the presence/absence binary matrices were generated using the SARAMIS Superspectra tool. Results were then imported into the PAleontological STatistics (PAST) software (Natural History Museum, Oslo University, Norway). Multivariate neighbor joining cluster analyses with similarity distances were calculated using Dice algorithm in PAST software (Feltens et al. 2010). Figure 1 was generated and colored using the R software (http://www.r-project.org/).

Results

A comprehensive database for the identification of nodule bacteria via MALDI-TOF MS

In order to establish a database of MALDI-TOF MS spectra that could capture the existing diversity of rhizobia, we sourced representative strains of diverse rhizobial species (see Table 1). Most of the selected RNB were part of the DoE Joint Genome Institute's (JGI) international Genomic Encyclopedia of Bacteria and Archaea-Root Nodule Bacteria (GEBA-RNB) project that aimed to systematically sequence the genomes of diverse rhizobial strains (Reeve et al. 2014). The remaining strains were well-studied bacteria whose genomes have previously been sequenced. Together, these 116 selected strains represented the ten major genera of RNB and tumorigenic bacteria: *Agrobacterium*, *Azorhizobium*, *Bradyrhizobium*, *Burkholderia*, *Cupriavidus*, *Ensifer* (syn. *Sinorhizobium*), *Mesorhizobium*, *Methylobacterium*, *Microvirga*, and *Rhizobium*, isolated from 72 legume species and over 30 countries. This collection of strains included 27 fully described species, of which *R. leguminosarum* and its three biovars *phaseoli*, *trifolii*, and *viciae* comprised the largest number of strains (n=27).

To build up the database of MALDI mass spectra, each strain was grown at 28 °C on YM or TY solid media for 1, 2, 3, 5, and 7 days before being fingerprinted in quadruplicate using suspensions in 25 % formic acid. This procedure was found to be the most reliable for obtaining reproducible MALDI mass spectra for both free-living plate-cultured Rhizobia and bacteria analyzed directly from nodules, independently of the age of colonies or plant nodules (Ziegler et al. 2012). For each of the 116 reference strains, a reference SuperSpectrum™ (SSp) was computed using at least 15 individual spectra of the highest quality and, depending on the range of biomarker masses targeted, the alpha-cyano-4-hydroxycinnamic acid (CHCA; 3–12 kDa) or sinapinic acid (SA; 4–30 kDa) matrices (Pineda et al. 2003). The 232 SSp generated in this way together formed the "rhizobia-specific module" of the SARAMIS™ database.

Using deduced ribosomal proteins as markers for MALDI-TOF MS identification of rhizobia

Given the increasing number of sequenced bacterial genomes and the difficulty of obtaining numerous yet well-characterized and curated reference strains for further developing the rhizobia-specific SARAMIS™ module, we searched for ways to replace the empirically derived MALDI-TOF mass spectra with databases consisting of deduced masses of biomarker ribosomal proteins, which had been calculated from genomic sequence data.

Genomes of the 116 selected RNB and tumorigenic bacteria that were accessed online using the JGI or National Center for Biotechnology Information (NCBI) homepages and the genome online database identifiers (GOLD ID) are listed in Table 1. The search revealed that each of the genomes included genes of 54 RPs that potentially could be used as biomarkers. Masses of the 54 RPs were deduced from genomic sequences and compiled into strain-specific lists of biomarkers. Where genes coding for RPs were either missing or found to be misannotated, ribosomal protein nucleotide sequences of closely related strains were used to identify the correct open-reading frame. Table S1 in the Supplementary Material shows the 163 RPs for which these genome

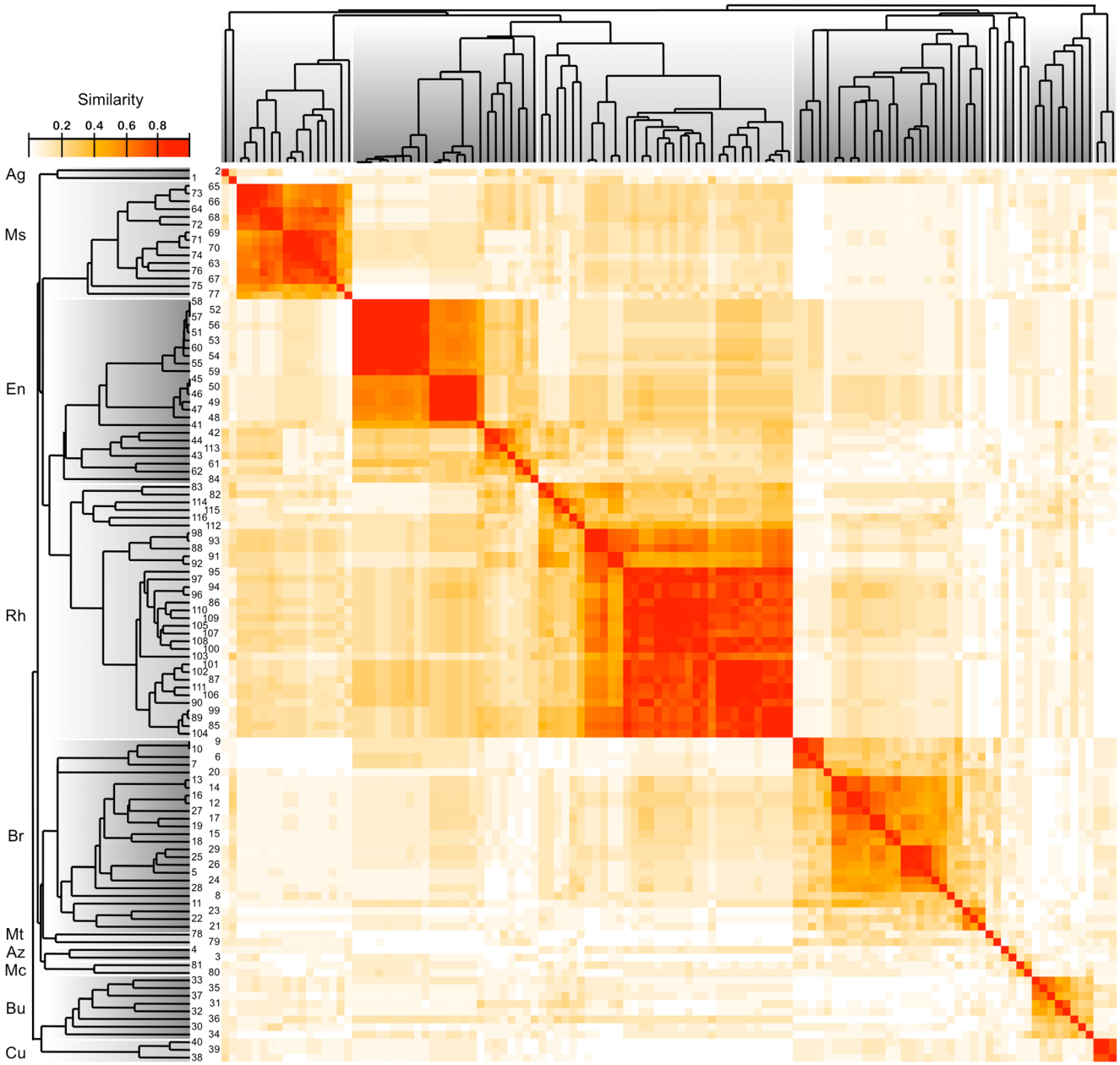

Fig. 1 Hierarchical cluster analysis showing similarities in the number of identical masses calculated for all of the ribosomal proteins (RPs) and in pairwise combinations of strains. Strains are numbered as in Table 1, and alternating shades of *grey* delimit the boundaries (from *left* to *right* and *top* to *bottom*) for the *Agrobacterium* (*Ag*), *Mesorhizobium* (*Ms*), *Ensifer* (*En*), *Rhizobium* (*Rh*), *Bradyrhizobium* (*Br*), *Methylobacterium* (*Mt*), *Azorhizobium* (*Az*), *Microvirga* (*Mc*), *Burkholderia* (*Bu*), and *Cupriavidus* (*Cu*) genera. The degree of homogeneity or heterogeneity of clustering groups is illustrated in the heatmap. Levels of similarity between the theoretical spectra composed of 54 RPs are color-coded, from *white* (no similarity) to *dark red* (full identity)

annotation corrections needed to be made. In order to calculate masses of ribosomal proteins as accurately as possible, the most frequent posttranslational modifications were taken into account. For example, 131 Da corresponding to a cleaved methionine was subtracted from the predicted mass of the ribosomal protein when the amino acid in position 2 was either Gly, Ala, Pro, Ser, Thr, Val, or Cys. In addition, ribosomal protein L33 was always considered to be methylated (Arnold and Reilly 1999), except for *Burkholderia* strains, where our experimental data showed no evidence for such methylation. Masses for RPs calculated in this way ranged from 4.3 kDa (L36 of *Cupriavidus taiwanensis* LMG 19424^T) to 63.2 kDa (S1 of *Bradyrhizobium* sp. ARR65), of which 71 % had a mass below the 30 kDa upper limit for MALDI-TOF spectra acquired using sinapinic acid as matrix. To introduce this dataset into the SARAMIS database, a text file listing all of the 54 RPs masses with arbitrary intensity values was compiled for each of our 116 bacterial strains. This complete

dataset of deduced ribosomal protein masses was labeled as DRPM.

Figure 1 shows two complementary representations of a hierarchical cluster analysis using the Dice algorithm for similarities in the number of identical masses among the 54 deduced RPs in pairwise combinations of strains. In the cluster trees, genera boundaries are delimited using labeled and shaded domains with strains being numbered as in Table 1. In the central diagram, color codes ranging from white (no similarity) to dark red (full identity) show corresponding levels of similarity. Similar and more detailed analyses for strains belonging to each of the genera *Bradyrhizobium*, *Ensifer*, *Mesorhizobium*, and *Rhizobium*, respectively, are presented in Figs. S1–S4.

Except for *Rhizobium giardinii* bv. *giardinii* type strain H152^T (strain #84), strains clustered into distinct and taxonomically correct genera with Rhizobia belonging to the *Beta* subdivision of *Proteobacteria* (*Burkholderia* and *Cupriavidus*) forming a distinct branch (shown at the very bottom or the very right in Fig. 1). Interestingly, the theoretical RPs spectrum of *R. giardinii* bv. *giardinii* H152^T appeared to be more similar to those of *Ensifer* than *Rhizobium* strains, a finding consistent with the unclear phylogenetic position of H152^T reported previously (Laguerre et al. 2001; Mousavi et al. 2014; Ormeño-Orrillo et al. 2015). Closely related strains, such as the *Ensifer medicae* (#45 to 50) and *Ensifer meliloti* (#51 to 60) isolates, formed two compact clusters clearly separated from other *Ensifer* members by *Ensifer arboris* (#41) (see also Supplementary Fig. S2). A similar topology was reported for a 16S rRNA-based phylogenetic tree of *Ensifer* strains when describing the genome sequence of *E. arboris* LMG 14919^T (Reeve et al. 2013). The theoretical RPs spectrum also confirmed differences between *Rhizobium etli* CFN 42^T (#82) and strain CIAT 652 (#83) and supported the classification of CIAT 652 as belonging to *Rhizobium phaseoli* (López-Guerrero et al. 2012). The separate clusters formed by the *Mesorhizobium loti* strains (#67 to 73) indicated that these strains, previously identified on the basis of their ability to nodulate *Lotus* spp., were likely to belong to several different *Mesorhizobium* species. In contrast to the *Ensifer*, *Mesorhizobium* and *Rhizobium* genera, spectra of RPs calculated for bradyrhizobia showed significantly lower levels of similarity. This was not surprising given the heterogeneity of bradyrhizobial strains (Sánchez-Juanes et al. 2013). Figure 1 also showed that in spite of only two isolates of distinct species selected to represent each of the *Agrobacterium*, *Azorhizobium*, *Methylobacterium*, and *Microvirga* genera, the clustering of theoretical RPs spectra was unambiguous. This suggested that the limited number of representative strains currently included in the rhizobia-specific module for these four genera did not prevent accurate strain identification. These results indicated that together, masses of ribosomal proteins were potentially sufficient markers for generating strain-specific MALDI-TOF MS fingerprints.

Ascribing ribosomal proteins to masses found in empirical MALDI-TOF MS spectra

In routine procedures for strain identification, factors such as temperature, growth media, and conditions can influence the quality of MALDI-TOF MS spectra, and the number and intensity of peaks that are recorded (Wieme et al. 2014). Thus, to maximize the chances of recovering as many ribosomal proteins as possible, cell lysate samples were used to obtain at least four high-quality MALDI-TOF MS spectra for each of the 116 bacterial reference strains. All mass peaks between 4 and 30 kDa found in the empirical spectrum of a given strain were matched against the ribosomal protein masses that were deduced from the corresponding genome sequence. To account for possible experimental variations in measurements, an error interval of 800 ppm was tolerated, and only mass peaks successfully ascribed to a ribosomal protein and found in at least half of the empirical spectra produced for each strain were further considered. Depending on the genus to which individual strains belonged, the numbers of reproducible masses successfully ascribed to ribosomal proteins were as follows: *Agrobacterium* 25, *Azorhizobium* 22, *Bradyrhizobium* 17, *Burkholderia* 22, *Cupriavidus* 21, *Ensifer* 22, *Mesorhizobium* 15, *Methylobacterium* 25, *Microvirga* 25, and *Rhizobium* 22. Of these assigned RP masses, 13 were consistently identified in all spectra, regardless of the strain or genus considered (Fig. 2). This subset of reproducible masses corresponded to ribosomal subunits L20, L24, L28, L29, L31, L32, L33, L36, S10, S15, S16, S17, and S20 (see Table S2 in the Supplementary Material). As purified cell lysates were used to generate the empirical spectra used for ascribing ribosomal proteins to detectable masses, rather than isolated ribosomes, this experimental dataset was recorded in the PARPM database.

To assess whether this PARPM database was able to provide accurate identification of strains using standard empirical spectra as queries, four to five of the sinapinic acid-prepared spectra that had initially been collected from each the 116 reference strains were analyzed again for the presence or absence of mass peaks that corresponded to the selected subset of 13 RPs. In 539 of the 571 spectra (94 %) examined in this way, up to 13 RP markers were successfully assigned, with eight markers found to be the minimal threshold for obtaining an accurate identification at the species level.

Congruent topologies for 16S-, MLSA-, rMLSA, or PARPM-based phylogenetic trees

Ribosomal proteins have previously been shown to be reliable mass-spectra markers for phylogenetic classification of various bacteria (Suarez et al. 2013; Teramoto et al. 2007b). To

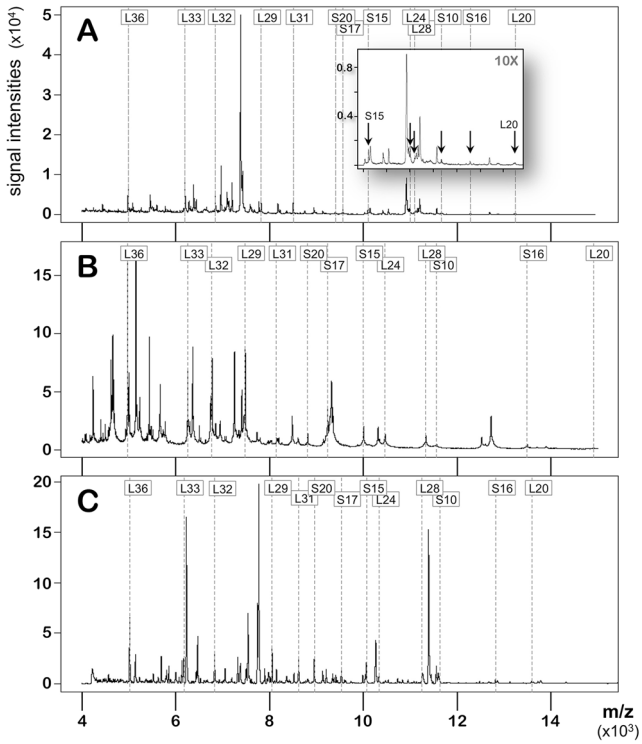

Fig. 2 Empirical spectra of strains *Bradyrhizobium* sp. WSM 2793 (panel **a**), *Ensifer fredii* NGR234 (**b**), and *Mircrovirga lotononidis* WSM3557^T (**c**) onto which the respective positions of the 13 ribosomal subunits (L36–L20 and 4–15 kDa in size) that were selected to construct the PARPM database are shown as annotated *dashed lines*. Note that although the intensities of peaks for several RPs appeared weak, signals were nonetheless recorded as protein hits in the automatic analysis mode. For example, low-intensity peaks matching ribosomal subunits are marked by *arrows* in the superimposed section of panel **a** in which signal intensities were magnified ten times (10×)

test whether the subset of 13 RPs that comprised the PARPM database also provided a sufficient basis for phylogenetic analyses of the 116 selected strains of RNB and tumorigenic bacteria, we compared the topologies of trees reconstructed using the 16S rDNA sequences (16S rRNA); concatenated *recA*, *gyrB*, *rpoB*, *dnaJ*, and *atpD* sequences (MLSA); concatenated sequences coding for the ribosomal proteins *rplT* (L20), *rplX* (L24), *rpmB* (L28), *rpmC* (L29), *rpmE* (L31), *rpmF* (L32), *rpmG* (L33), *rpmJ* (L36), *rpsJ* (S10), *rpsO* (S15), *rpsP* (S16), *rpsQ* (S17), and *rpsT* (S20) (rMLSA); and the PARPM data (PARPM, see Fig. 3). As the 16S rRNA genes were missing from the draft genome sequences of *R. leguminosarum* bv. *viciae* strains 128c53, Ps8, and Vh3, as well as *R. sullae* WSM1592, the 16S rDNA sequences were amplified and deposited in GenBank (accession numbers KF678361–KF678364) before reconstructing the corresponding tree.

As shown in Fig. 3 and except for strain H152^T, the remaining 115 strains included in this analysis grouped into genera that were consistent with current rhizobial taxonomy. Depending on the data used for reconstructing the phylogenetic tree, strain *R. giardinii* bv. *giardinii* strain H152^T clustered with either agrobacteria (16S rRNA) or members of the *Ensifer* genus (MLSA, rMLSA, and PARPM). Data presented in Fig. 3 illustrates the limited resolution obtained at the species level for trees reconstructed using DNA sequences. In contrast, the PARPM dataset offered an in-depth resolution including at the subspecies level and, in particular, for the agronomically important *Bradyrhizobium*, *Ensifer*, and *Rhizobium* genera. However, while the PARPM dataset consistently grouped the strains within genera and accurately reflected diversity and differences at the species and often subspecies level, it was not able to resolve rhizobial evolutionary relationships at higher taxonomic levels. Similarly, as genes that determine host range are part of the accessory genome, related strains of *R. leguminosarum* assigned to distinct symbiovars *phaseoli*, *viciae*, and *trifolii* were not resolved using PARPM (see positions of strains #85, 89, 99, and 104 in Fig. 3d).

Testing the rhizobia-specific module of the SARAMIS™ database

The reliability and resolving power of the rhizobia-specific module was initially assessed on plate cultures of more than 100 unknown bacterial isolates obtained from root nodules of field-grown pigeon pea collected from Ivory Coast sites. According to MALDI-TOF MS results, most of the pigeon pea isolates clustered into two major subgroups, identified as either *Bradyrhizobium elkanii* or *Bradyrhizobium* spp. (data not shown). The identities of representative strains of both subgroups were later confirmed by 16S rRNA sequencing (see Table 2). Additionally, the symbiotic phenotypes of these strains were assayed on either the original host *C. cajan* or the more promiscuous legume *V. unguiculata* (Lewin et al. 1987). A similar procedure was carried out for a few isolates, such as strain CI-34D1, which were not recognized by MALDI-TOF MS as a potential RNB. In fact, strain CI-34D1 was identified by MALDI-TOF MS as a *Brevibacillus* sp. and 16S rDNA sequencing showed it was related to *Brevibacillus reuszeri*. As shown in Table 2, only the *C. cajan* isolates identified by MALDI-TOF MS as members of known rhizobial genera were found to nodulate and fix nitrogen with both pigeon pea and cowpea. As expected, the non-RNB strain CI-34D1 failed to nodulate either of the tested host plants. In further tests, endosymbiotic bacteria from nodules collected from roots of clover (*Trifolium* spp.) or French seradella (*Ornithopus sativus*) growing in Australian fields were analyzed directly by MALDI-TOF MS using the protocol established by Ziegler et al. (2012), which does not require isolation or cultivation of the RNB. Nodules were found to harbor bacteria identified by MALDI-TOF MS as either *R. leguminosarum* (clover symbionts) or *Bradyrhizobium canariense* (French seradella symbionts). To verify the accuracy of MALDI-TOF MS-based results, nodule bacteria were

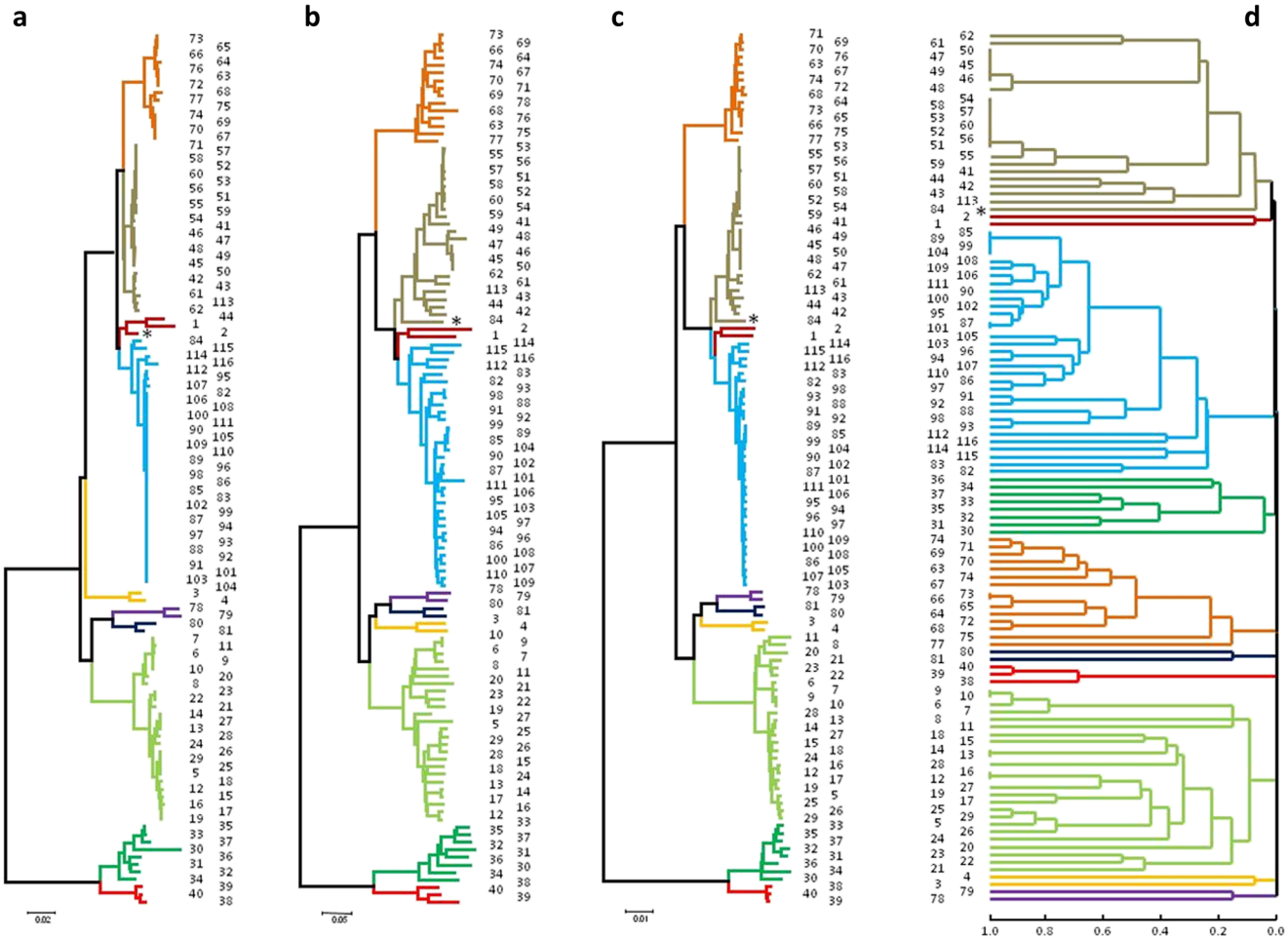

Fig. 3 Topology of trees obtained by cluster analysis of the **a** partial 16S rRNA sequences (1,006 nucleotides), **b** concatenated sequences of *recA*, *gyrB*, *rpoB*, *dnaJ*, and *atpD* genes (in total 9,531 bases), **c** concatenated sequences of the protein-coding *rplT*, *rplX*, *rpmB*, *rpmC*, *rpmE*, *rpmF*, *rpmG*, *rpmJ*, *rpsJ*, *rpsO*, *rpsP*, *rpsQ*, and *rpsT* genes (total of 2,991 nucleotides), or **d** binary matrix derived from PARPM database. Strains were numbered as in Table 1 and were listed next to each of the tree branches. To facilitate comparison of tree topologies, bacterial genera were color-coded as follows: *Agrobacterium* (*wine red*), *Azorhizobium* (*orange*), *Bradyrhizobium* (*light green*), *Burkholderia* (*green*), *Cupriavidus* (*red*), *Ensifer* (*olive*), *Mesorhizobium* (*brown*), *Methylobacterium* (*violet*), *Microvirga* (*dark blue*), and *Rhizobium* (*light blue*). Poorly resolved taxonomic position of *R. giardinii* bv. *giardinii* type strain H152^T marked by as *asterisk*

isolated and the 16S rDNA sequences obtained for five (A-BRK13, A-G150, A-G169, A-Q1.7, and A-WB) out of the 41 isolates identified as *R. leguminosarum* and one (A-GQ1.6) of the four *B. canariense* strains (see Table 2). As all of these results corroborated the initial MALDI-TOF MS identifications, the rhizobia-specific module of SARAMIS™ was found to correctly identify symbiotic bacteria whether grown on solid media or found inside nitrogen-fixing nodules.

Spectra that were generated for identifying bacteria isolated from Ivory Coast pigeon pea nodules were also matched against PARPM. Ultimately, all of the spectra of *C. cajan* isolates previously identified as members of the *B. elkanii* genus were found to best match the PARPM profiles of *B. elkanii* strains USDA76^T, USDA94, USDA3259, and USDA3254. In addition, the SA-prepared spectra of Rhizobia isolated from *Trifolium* sp. nodules were found to share 9 to 12 of the 13 RPs markers included in the PARPM accessions for the reference *R. leguminosarum* bv. *trifolii* strains SRDI943 and WSM1325. These results indicated that identification of rhizobial isolates through a PARPM-based analysis was both feasible and reliable.

Discussion

This work highlights the robustness and accuracy of mass spectrometry for high-throughput identification of RNB and tumorigenic bacteria. In comparison to classical identification procedures, such as biochemical tests, restriction or amplified fragment length polymorphism (RFLP or AFLP), or sequencing of marker genes (e.g., 16S rRNA), MALDI-TOF MS is considerably faster, more cost-effective, and less labor-intensive. These advantages, along with the proven accuracy of MALDI-TOF MS-based identification of bacteria, have

Table 2 Characteristics of strains isolated from field nodules

Strain	Identification result as given by		Isolated from	Symbiotic phenotype on	
	MALDI-TOF MS	16S sequencing (closest related reference)		Pigeon pea	Cowpea
CI-36B	*Bradyrhizobium elkanii*	*B. elkanii* (USDA76^T/100 %)	*Cajanus cajan*	Fix+	Fix+
CI-1B	*Bradyrhizobium* sp.	*B. elkanii* (USDA76^T/99 %)	*C. cajan*	Fix+	Fix+
CI-34D1	*Brevibacillus* sp.	*Brevibacillus reuszeri* (DSM9887^T)	*C. cajan*	Nod−	Nod−
A-WB	*Rhizobium leguminosarum*	*R. leguminosarum* bv. *trifolii* (WSM1325)	*Trifolium spumosum*	nt	Nt
A-G150	*R. leguminosarum*	*R. leguminosarum* bv. *trifolii* (WSM1325)	*T. subterraneum*	nt	Nt
A-G169	*R. leguminosarum*	*R. leguminosarum* bv. *trifolii* (WU95)	*T. subterraneum*	nt	Nt
A-BRK13	*R. leguminosarum*	*R. leguminosarum* bv. *trifolii* (WSM1325)	*T. spumosum*	nt	Nt
A-Q1.7	*R. leguminosarum*	*R. leguminosarum* bv. *trifolii* (WSM1325)	*Trifolium* sp.	nt	Nt
A-Q1.6	*Bradyrhizobium canariense*	*B. canariense* (WSM471)	*Ornithopus sativus*	nt	Nt

Identity of selected nodule isolates was established using the Rhizobia-specific module of the upgraded SARAMIS™ database (MALDI-TOF MS) or by matching the DNA sequence of the amplified 16S rRNA gene against the GenBank database (16S sequencing). Nodulation and nitrogen fixation properties of the isolated strains from Ivory Coast was examined using Magenta jars as described in the "Determining symbiotic properties of nodule isolates". Symbiotic proficiency of bacteria isolated from nodules of *Trifolium* spp. and *O. sativus* collected in Australia was not tested (nt) as no seeds of these particular legumes could be obtained

revolutionized the routine identification of microorganisms to species and subspecies level in clinical diagnostic and food microbiological laboratories (Croxatto et al. 2012; Mazzeo et al. 2006). MALDI-TOF MS has also been shown to be an excellent tool for identification of diverse rhizobial species of *Bradyrhizobium*, *Ensifer*, *Shinella*, and *Rhizobium* applicable to large populations of isolates in ecological and taxonomic studies (Ferreira et al. 2011; Sánchez-Juanes et al. 2013).

However, accurate identification of bacteria via MALDI-TOF MS depends upon the establishment of comprehensive spectral databases for a broad range of type or reference strains, and the currently available rhizobial databases are not representative of the phylogenetic diversity of RNB. We, therefore, developed a rhizobia-specific module for the SARA MIS™ database that consisted of 114 fully sequenced and well-characterized strains belonging to nine of the most common genera of RNB, plus two agrobacterial accessions. It included 27 strains of *R. leguminosarum* bv. *phaseoli*, *trifolii*, or *viciae* and a number of strains of *E. medicae* and *E. meliloti*. Providing multiple strains for these species allowed us to analyze interstrain variability and thereby assess the effectiveness of MALDI-TOF MS for typing rhizobial strains, as has previously been shown for *Neisseria meningitides* isolates (Suarez et al. 2013). Our results showed that MALDI-TOF MS analysis using the rhizobia-specific database grouped RNB strains into genera and species congruent with the phylogenies obtained using marker gene sequences (16S rRNA, MLSA) and consistent with classifications previously obtained for cluster analysis of MALDI-TOF spectra of *Ensifer* and *Rhizobium* spp. (Ferreira et al. 2011). The database also provided an accurate and highly resolved analysis of strain diversity at the species and often subspecies level. This highlights the value of MALDI-TOF MS as a tool not only for identification, but also for microbial systematics (Welker and Moore 2011).

Using either pure plate cultures of bacterial isolates from Ivory Coast pigeon pea nodules or bacteria obtained directly from nodules of *Trifolium* spp. or French seradella, we then confirmed that the rhizobia-specific module could be used to accurately identify symbiotic rhizobia in field studies. When non-RNB were analyzed, the larger SARAMIS™ bacterial database provided an identification that was in agreement with subsequent 16S rRNA analyses. This makes MALDI-TOF MS and the rhizobia-specific module an ideal replacement for the demanding and expensive sequencing of partial 16S rRNA or AFLP/RFLP analyses that have been performed to identify and determine the diversity of microsymbionts of wild legumes in recent large-scale studies (Aserse et al. 2013; Safronova et al. 2014; Yang et al. 2013; Zhao et al. 2014). In the economically important field of symbiotic nitrogen fixation in agricultural systems, knowledge of the biodiversity of rhizobia and of local populations is important for the design of successful inoculation strategies (Lindström et al. 2010). Identification of RNB to strain level offers the possibility of using MALDI-TOF MS in diagnostic tests, for example, in ecological studies to test the ability of inoculant strains to compete with naturalized soil populations of rhizobia (Denton et al. 2002).

As new legume symbionts are isolated, named and described, the rhizobia-specific module will need to be updated with additional reference strains, in particular, for genera that are poorly represented in this study (e.g., *Methylobacterium* and *Microvirga*) or not yet included (e.g., *Phyllobacterium*). Obtaining a pure copy of a reference strain for generating the corresponding MALDI-TOF mass spectra and superspectra is becoming ever more difficult due to increasing legal

constraints that limit the transfer of microbiological material between countries. We, therefore, explored the possibility of using whole or partial rhizobial genome sequences as a basis for calculating protein masses that could be used as strain-specific biomarkers in MALDI-TOF mass spectra. These proteins needed to (i) be representative of the microorganism's core genome, (ii) fall within the mass range detected by available MALDI-TOF mass spectrometers, and (iii) constitute the most frequent mass peaks in empirical spectra. As proteins encoded by MLSA marker genes, such as *dnaJ*, *atpD*, *gyrB*, *recA*, and *rpoB*, had masses larger than 38 kDa (the molecular weight of RecA) and thus were too large for standard mass spectral analyses, we decided instead to focus on the RPs. A recent study found that a subset of biomarkers composed of 15 RPs and named as *S10-spc-alpha* operon gene-encoded ribosomal protein mass spectrum (SP10 GERMS) are highly conserved in eubacterial genomes and can discriminate bacteria at species and strain levels (Tamura et al. 2013).

By using the genome sequences of the 116 reference strains to deduce the masses of their RPs, then using hierarchical cluster analysis to search for similarities in the number of identical RP masses in pairwise combinations of strains (Fig. 1), we showed that RPs could be used as molecular markers to assign RNB to the genera or species to which they belonged. This confirms RPs as accurate markers for bacterial discrimination (Sato et al. 2012; Teramoto et al. 2007b) and, in a major finding, shows that they can be used to construct in silico reference MS spectra for strains whose genomes have been fully sequenced. When mass peaks found in empirical spectra of RNB were matched against genome-deduced masses of RPs, we identified subunits L20, L24, L28, L29, L31, L32, L33, L36, S10, S15, S16, S17, and S20 as the 13 most frequently ascribed ribosomal protein masses that together formed fingerprints that were specific for 89 of the 116 reference strains. This data was translated into a SARA MIS™-compatible format and stored into a specific database called PARPM. Querying the PARPM database with empirical mass spectra obtained using RNB grown on plates, or directly from nodules, resulted in correct identification of unknown isolates to the species level.

The rapid and constant evolution and revision of rhizobial taxonomy presents a challenge to the determination of microbial identity and phylogeny. Advances in genomic sequencing technologies and analyses now allow a much greater precision in the taxonomic placement of strains (Jolley et al. 2012; Kim et al. 2014; Klenk and Göker 2010). The increasing affordability of next-generation sequencing technologies is resulting in a vast increase in the number of sequenced bacterial genomes. However, it is of crucial importance that the sequenced strains are selected with care to reflect bacterial diversity. To this end, sequencing projects, such as the Genomic Encyclopedia of Bacteria and Archaea-type strains, ensure reproducibility of the results as well as provide a strong link between genome data and biologically relevant features of each of the selected strains (Kyrpides et al. 2014b). The integration of sequencing technologies with MS databases has the potential to revolutionize rhizobial, or indeed microbial, diagnostics and systematics. Already more than 130 additional and complementing whole genomes of RNB have been accessed and stored as DRPM in the SARAMIS™ database (data not shown). That type strains of novel species should have their genomes sequenced at the time they are deposited into culture collections as was recently proposed (Kyrpides et al. 2014a), will facilitate the development of this database.

Over the years, genome data has become essential for understanding the molecular basis of plant symbioses (Mora et al. 2014), and some of the mechanisms involved in the evolution and speciation of rhizobia (Tian et al. 2012). Our results now illustrate how such genomic datasets can become crucial elements for establishing new technological developments able to facilitate large-scale studies of rhizobial populations in fields and natural environments.

Acknowledgments This work was funded by the Swiss Commission for Technology and Innovation (CTI project 11225.1 PFLS_LS), the Swiss National Science Foundation (Grant Nos. 31003A-116591 and 31003A_146548), Mabritec, and Meat & Livestock Australia/ Australian Wool Innovation Grant B.PSP.0013. We are grateful to Susana Brom, Sofie Goormachtig, Eric Giraud, Alessio Mengoni, Philip Poole, José-Enrique Ruiz-Sainz, and Maged Saad for providing a number of reference strains, Damiana Ravasi and Paola Decristophoris for assistance with DNA sequencing of a number of marker genes, and Antoine Huyghe for his help in generating Fig. 1.

References

Arnold RJ, Reilly JP (1999) Observation of *Escherichia coli* ribosomal proteins and their posttranslational modifications by mass spectrometry. Anal Biochem 269:105–112. doi:10.1006/abio.1998.3077

Aserse AA, Rasanen LA, Aseffa F, Hailemariam A, Lindström K (2013) Diversity of sporadic symbionts and nonsymbiotic endophytic bacteria isolated from nodules of woody, shrub, and food legumes in Ethiopia. Appl Microbiol Biotechnol 97:10117–10134. doi:10.1007/s00253-013-5248-4

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

Brosius J, Palmer ML, Kennedy PJ, Noller HF (1978) Complete nucleotide sequence of a 16S ribosomal RNA gene from *Escherichia coli*. Proc Natl Acad Sci U S A 75:4801–4805

Broughton WJ (2003) Roses by other names: taxonomy of the *Rhizobiaceae*. J Bacteriol 185:2975–2979

Broughton WJ, Dilworth MJ (1971) Control of leghaemoglobin synthesis in snake beans. Biochem J 125:1075–1080

Croxatto A, Prod'hom G, Greub G (2012) Applications of MALDI-TOF mass spectrometry in clinical diagnostic microbiology. FEMS Microbiol Rev 36:380–407. doi:10.1111/j.1574-6976.2011.00298.x

De Meyer SE, Cnockaert M, Ardley JK, Van Wyk BE, Vandamme PA, Howieson JG (2014) *Burkholderia dilworthii* sp. nov., isolated from *Lebeckia ambigua* root nodules. Int J Syst Evol Microbiol 64:1090–1095. doi:10.1099/ijs. 0.058602-0

Delamuta JR, Ribeiro RA, Ormeno-Orrillo E, Melo IS, Martinez-Romero E, Hungria M (2013) Polyphasic evidence supporting the reclassification of *Bradyrhizobium japonicum* group Ia strains as *Bradyrhizobium diazoefficiens* sp. nov. Int J Syst Evol Microbiol 63:3342–3351. doi:10.1099/ijs. 0.049130-0

Denton MD, Coventry DR, Murphy PJ, Howieson JG, Bellotti WD (2002) Competition between inoculant and naturalised *Rhizobium leguminosarum* bv. *trifolii* for nodulation of annual clovers in alkaline soils. Aust J Agr Res 53:1019–1026. doi:10.1071/Ar01138

Doyle JJ (2011) Phylogenetic perspectives on the origins of nodulation. Mol Plant Microbe Interact 24:1289–1295. doi:10.1094/MPMI-05-11-0114

Feltens R, Gorner R, Kalkhof S, Groger-Arndt H, von Bergen M (2010) Discrimination of different species from the genus *Drosophila* by intact protein profiling using matrix-assisted laser desorption ionization mass spectrometry. BMC Evol Biol 10:95. doi:10.1186/1471-2148-10-95

Ferreira L, Sánchez-Juanes F, García-Fraile P, Rivas R, Mateos PF, Martínez-Molina E, González-Buitrago JM, Velázquez E (2011) MALDI-TOF mass spectrometry is a fast and reliable platform for identification and ecological studies of species from family *Rhizobiaceae*. PLoS ONE 6:e20223. doi:10.1371/journal.pone.0020223

Freiberg C, Fellay R, Bairoch A, Broughton WJ, Rosenthal A, Perret X (1997) Molecular basis of symbiosis between *Rhizobium* and legumes. Nature 387:394–401

Gevers D, Cohan FM, Lawrence JG, Spratt BG, Coenye T, Feil EJ, Stackebrandt E, Van de Peer Y, Vandamme P, Thompson FL, Swings J (2005) Opinion: re-evaluating prokaryotic species. Nat Rev Microbiol 3:733–739. doi:10.1038/nrmicro1236

Gibson KE, Kobayashi H, Walker GC (2008) Molecular determinants of a symbiotic chronic infection. Annu Rev Genet 42:413–441. doi:10.1146/annurev.genet.42.110807.091427

Hanage WP, Fraser C, Spratt BG (2006) Sequences, sequence clusters and bacterial species. Philos Trans R Soc Lond Ser B Biol Sci 361:1917–1927. doi:10.1098/rstb.2006.1917

Herridge DF, Peoples MB, Boddey RM (2008) Global inputs of biological nitrogen fixation in agricultural systems. Plant Soil 311:1–18. doi:10.1007/S11104-008-9668-3

Hotta Y, Sato H, Hosoda A, Tamura H (2012) MALDI-TOF MS analysis of ribosomal proteins coded in S10 and *spc* operons rapidly classified the *Sphingomonadaceae* as alkylphenol polyethoxylate-degrading bacteria from the environment. FEMS Microbiol Lett 330:23–29. doi:10.1111/j.1574-6968.2012.02525.x

Ishihama Y, Schmidt T, Rappsilber J, Mann M, Hartl FU, Kerner MJ, Frishman D (2008) Protein abundance profiling of the *Escherichia coli* cytosol. BMC Genomics 9:102. doi:10.1186/1471-2164-9-102

Jolley KA, Bliss CM, Bennett JS, Bratcher HB, Brehony C, Colles FM, Wimalarathna H, Harrison OB, Sheppard SK, Cody AJ, Maiden MC (2012) Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. Microbiology 158:1005–1015. doi:10.1099/mic. 0.055459-0

Josten M, Reif M, Szekat C, Al-Sabti N, Roemer T, Sparbier K, Kostrzewa M, Rohde H, Sahl HG, Bierbaum G (2013) Analysis of the matrix-assisted laser desorption ionization-time of flight mass spectrum of *Staphylococcus aureus* identifies mutations that allow differentiation of the main clonal lineages. J Clin Microbiol 51:1809–1817. doi:10.1128/JCM. 00518-13

Kim M, Oh HS, Park SC, Chun J (2014) Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. Int J Syst Evol Microbiol 64:346–351. doi:10.1099/ijs. 0.059774-0

Klenk HP, Göker M (2010) En route to a genome-based classification of *Archaea* and *Bacteria*? Syst Appl Microbiol 33:175–182. doi:10.1016/j.syapm.2010.03.003

Kyrpides NC, Hugenholtz P, Eisen JA, Woyke T, Goker M, Parker CT, Amann R, Beck BJ, Chain PS, Chun J, Colwell RR, Danchin A, Dawyndt P, Dedeurwaerdere T, DeLong EF, Detter JC, De Vos P, Donohue TJ, Dong XZ, Ehrlich DS, Fraser C, Gibbs R, Gilbert J, Gilna P, Glockner FO, Jansson JK, Keasling JD, Knight R, Labeda D, Lapidus A, Lee JS, Li WJ, Ma J, Markowitz V, Moore ER, Morrison M, Meyer F, Nelson KE, Ohkuma M, Ouzounis CA, Pace N, Parkhill J, Qin N, Rossello-Mora R, Sikorski J, Smith D, Sogin M, Stevens R, Stingl U, Suzuki K, Taylor D, Tiedje JM, Tindall B, Wagner M, Weinstock G, Weissenbach J, White O, Wang J, Zhang L, Zhou YG, Field D, Whitman WB, Garrity GM, Klenk HP (2014a) Genomic encyclopedia of bacteria and archaea: sequencing a myriad of type strains. PLoS Biol 12:e1001920. doi:10.1371/journal.pbio.1001920

Kyrpides NC, Woyke T, Eisen JA, Garrity G, Lilburn TG, Beck BJ, Whitman WB, Hugenholtz P, Klenk HP (2014b) Genomic encyclopedia of type strains, phase i: the one thousand microbial genomes (KMG-I) project. Stand Genomic Sci 9:1278–1284. doi:10.4056/sigs.5068949

Laguerre G, Nour SM, Macheret V, Sanjuan J, Drouin P, Amarger N (2001) Classification of Rhizobia based on *nodC* and *nifH* gene analysis reveals a close phylogenetic relationship among *Phaseolus vulgaris* symbionts. Microbiology 147:981–993

Lewin A, Rosenberg C, Meyer ZAH, Wong CH, Nelson L, Manen J-F, Stanley J, Dowling DN, Dénarié J, Broughton WJ (1987) Multiple host-specificity loci of the broad host range *Rhizobium* sp. NGR234 selected using the widely compatible legume *Vigna unguiculata*. Plant Mol Biol 8:447–459

Lindström K, Murwira M, Willems A, Altier N (2010) The biodiversity of beneficial microbe-host mutualism: the case of Rhizobia. Res Microbiol 161:453–463. doi:10.1016/j.resmic.2010.05.005

López-Guerrero MG, Ormeño-Orrillo E, Velázquez E, Rogel MA, Acosta JL, González V, Martínez J, Martínez-Romero E (2012) *Rhizobium etli* taxonomy revised with novel genomic data and analyses. Syst Appl Microbiol 35:353–358. doi:10.1016/j.syapm.2012.06.009

Maiden MC, Jansen van Rensburg MJ, Bray JE, Earle SG, Ford SA, Jolley KA, McCarthy ND (2013) MLST revisited: the gene-by-gene approach to bacterial genomics. Nat Rev Microbiol 11:728–736. doi:10.1038/nrmicro3093

Martens M, Dawyndt P, Coopman R, Gillis M, De Vos P, Willems A (2008) Advantages of multilocus sequence analysis for taxonomic studies: a case study using 10 housekeeping genes in the genus *Ensifer* (including former *Sinorhizobium*). Int J Syst Evol Microbiol 58:200–214. doi:10.1099/ijs. 0.65392-0

Masson-Boivin C, Giraud E, Perret X, Batut J (2009) Establishing nitrogen-fixing symbiosis with legumes: how many *Rhizobium* recipes? Trends Microbiol 17:458–466. doi:10.1016/j.tim.2009.07.004

Mazzeo MF, Sorrentino A, Gaita M, Cacace G, Di Stasio M, Facchiano A, Comi G, Malorni A, Siciliano RA (2006) Matrix-assisted laser desorption ionization-time of flight mass spectrometry for the discrimination of food-borne microorganisms. Appl Environ Microbiol 72:1180–1189. doi:10.1128/AEM. 72.2.1180-1189.2006

Mora Y, Díaz R, Vargas-Lagunas C, Peralta H, Guerrero G, Aguilar A, Encarnacion S, Girard L, Mora J (2014) Nitrogen-fixing rhizobial strains isolated from common bean seeds: phylogeny, physiology, and genome analysis. Appl Environ Microbiol 80:5644–5654. doi:10.1128/AEM. 01491-14

Moulin L, Bena G, Boivin-Masson C, Stepkowski T (2004) Phylogenetic analyses of symbiotic nodulation genes support vertical and lateral gene co-transfer within the *Bradyrhizobium* genus. Mol Phylogenet Evol 30:720–732. doi:10.1016/S1055-7903(03)00255-0

Mousavi SA, Osterman J, Wahlberg N, Nesme X, Lavire C, Vial L, Paulin L, de Lajudie P, Lindström K (2014) Phylogeny of the *Rhizobium–Allorhizobium–Agrobacterium* clade supports the delineation of *Neorhizobium* gen. nov. Syst Appl Microbiol 37:208–215. doi:10.1016/j.syapm.2013.12.007

Nandasena KG, O'Hara GW, Tiwari RP, Sezmis E, Howieson JG (2007) In situ lateral transfer of symbiosis islands results in rapid evolution of diverse competitive strains of mesorhizobia suboptimal in symbiotic nitrogen fixation on the pasture legume *Biserrula pelecinus* L. Environ Microbiol 9:2496–2511. doi:10.1111/j.1462-2920.2007.01368.x

Oldroyd GE, Downie JA (2008) Coordinating nodule morphogenesis with rhizobial infection in legumes. Annu Rev Plant Biol 59:519–546. doi:10.1146/annurev.arplant.59.032607.092839

Ormeño-Orrillo E, Servín-Garcidueñas LE, Rogel MA, González V, Martínez-Romero J, Martínez-Romero E (2015) Taxonomy of Rhizobia and agrobacteria from the *Rhizobiaceae* family in light of genomics. Syst Appl Microbiol in press. doi:10.1016/j.syapm.2014.12.002

Perret X, Staehelin C, Broughton WJ (2000) Molecular basis of symbiotic promiscuity. Microbiol Mol Biol Rev 64:180–201

Pineda FJ, Antoine MD, Demirev PA, Feldman AB, Jackman J, Longenecker M, Lin JS (2003) Microorganism identification by matrix-assisted laser/desorption ionization mass spectrometry and model-derived ribosomal protein biomarkers. Anal Chem 75:3817–3822. doi:10.1021/Ac034069b

Reeve W, Tian R, Braü L, Goodwin L, Munk C, Detter C, Tapia R, Han C, Liolios K, Huntmann M, Pati A, Woyke T, Mavrommatis K, Markowitz V, Ivanova N, Kyrpides N, Willems A (2013) Genome sequence of *Ensifer arboris* strain LMG 14919; a microsymbiont of the legume *Prosopis chilensis* growing in Kosti, Sudan. Stand Genomic Sci 9:473–483

Reeve W, Ardley JA, Tian R, Eshragi L, Yoon JW, Ngamwisetkun P, Seshadri R, Ivanova NN, Kyrpides NC (2014) A genomic encyclopedia of the root nodule bacteria: assessing genetic diversity through a systematic biogeographic survey. Stand Genomic Sci 9(39). doi:10.1186/1944-3277-9-39

Rivas R, Martens M, de Lajudie P, Willems A (2009) Multilocus sequence analysis of the genus *Bradyrhizobium*. Syst Appl Microbiol 32:101–110. doi:10.1016/j.syapm.2008.12.005

Safronova VI, Kimeklis AK, Chizhevskaya EP, Belimov AA, Andronov EE, Pinaev AG, Pukhaev AR, Popov KP, Tikhonovich IA (2014) Genetic diversity of Rhizobia isolated from nodules of the relic species *Vavilovia formosa* (Stev.) Fed. Antonie Van Leeuwenhoek 105:389–399. doi:10.1007/s10482-013-0089-9

Sánchez-Juanes F, Ferreira L, Alonso de la Vega P, Valverde A, Barrios ML, Rivas R, Mateos PF, Martínez-Molina E, González-Buitrago JM, Trujillo ME, Velázquez E (2013) MALDI-TOF mass spectrometry as a tool for differentiation of *Bradyrhizobium* species: application to the identification of *Lupinus* nodulating strains. Syst Appl Microbiol 36:565–571. doi:10.1016/j.syapm.2013.09.003

Sato H, Teramoto K, Ishii Y, Watanabe K, Benno Y (2011) Ribosomal protein profiling by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry for phylogenety-based subspecies resolution of *Bifidobacterium longum*. Syst Appl Microbiol 34:76–80. doi:10.1016/j.syapm.2010.07.003

Sato H, Torimura M, Kitahara M, Ohkuma M, Hotta Y, Tamura H (2012) Characterization of the *Lactobacillus casei* group based on the profiling of ribosomal proteins coded in *S10-spc-alpha* operons as observed by MALDI-TOF MS. Syst Appl Microbiol 35:447–454. doi:10.1016/j.syapm.2012.08.008

Schleifer KH (2009) Classification of bacteria and archaea: past, present and future. Syst Appl Microbiol 32:533–542. doi:10.1016/j.syapm.2009.09.002

Seng P, Drancourt M, Gouriet F, La Scola B, Fournier PE, Rolain JM, Raoult D (2009) Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Clin Infect Dis 49:543–551. doi:10.1086/600885

Suarez S, Ferroni A, Lotz A, Jolley KA, Guerin P, Leto J, Dauphin B, Jamet A, Maiden MC, Nassif X, Armengaud J (2013) Ribosomal proteins as biomarkers for bacterial identification by mass spectrometry in the clinical microbiology laboratory. J Microbiol Methods 94:390–396. doi:10.1016/j.mimet.2013.07.021

Sullivan JT, Ronson CW (1998) Evolution of Rhizobia by acquisition of a 500-kb symbiosis island that integrates into a *phe*-tRNA gene. Proc Natl Acad Sci U S A 95:5145–5149

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739. doi:10.1093/molbev/msr121

Tamura H, Hotta Y, Sato H (2013) Novel accurate bacterial discrimination by MALDI-time-of-flight MS based on ribosomal proteins coding in *S10-spc-alpha* operon at strain level S10-GERMS. J Am Soc Mass Spectrom 24:1185–1193. doi:10.1007/s13361-013-0627-8

Teramoto K, Sato H, Sun L, Torimura M, Tao H (2007a) A simple intact protein analysis by MALDI-MS for characterization of ribosomal proteins of two genome-sequenced lactic acid bacteria and verification of their amino acid sequences. J Proteome Res 6:3899–3907. doi:10.1021/pr070218l

Teramoto K, Sato H, Sun L, Torimura M, Tao H, Yoshikawa H, Hotta Y, Hosoda A, Tamura H (2007b) Phylogenetic classification of *Pseudomonas putida* strains by MALDI-MS using ribosomal subunit proteins as biomarkers. Anal Chem 79:8712–8719. doi:10.1021/ac701905r

Teramoto K, Kitagawa W, Sato H, Torimura M, Tamura T, Tao H (2009) Phylogenetic analysis of *Rhodococcus erythropolis* based on the variation of ribosomal proteins as observed by matrix-assisted laser desorption ionization-mass spectrometry without using genome information. J Biosci Bioeng 108:348–353. doi:10.1016/J.Jbiosc.2009.04.010

Tian CF, Zhou YJ, Zhang YM, Li QQ, Zhang YZ, Li DF, Wang S, Wang J, Gilbert LB, Li YR, Chen WX (2012) Comparative genomics of Rhizobia nodulating soybean suggests extensive recruitment of lineage-specific genes in adaptations. Proc Natl Acad Sci U S A 109:8629–8634. doi:10.1073/pnas.1120436109

Udvardi M, Poole PS (2013) Transport and metabolism in legume–Rhizobia symbioses. Annu Rev Plant Biol 64:781–805. doi:10.1146/annurev-arplant-050312-120235

Vincent JM (1970) A manual for the practical study of the root-nodule bacteria. International biological programme handbook, 15th edn. Blackwell Scientific Pubilications, Oxford

Welker M, Moore ER (2011) Applications of whole-cell matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry in systematic microbiology. Syst Appl Microbiol 34:2–11. doi:10.1016/j.syapm.2010.11.013

Werner GD, Cornwell WK, Sprent JI, Kattge J, Kiers ET (2014) A single evolutionary innovation drives the deep evolution of symbiotic N2-fixation in angiosperms. Nat Commun 5:4087. doi:10.1038/ncomms5087

Wieme AD, Spitaels F, Aerts M, De Bruyne K, Van Landschoot A, Vandamme P (2014) Effects of growth medium on matrix-assisted laser desorption–ionization time of flight mass spectra: a case study of acetic acid bacteria. Appl Environ Microbiol 80:1528–1538. doi:10.1128/AEM. 03708-13

Willems A (2006) The taxonomy of Rhizobia: an overview. Plant Soil 287:3–14

Yang W, Kong Z, Chen W, Wei G (2013) Genetic diversity and symbiotic evolution of Rhizobia from root nodules of *Coronilla varia*. Syst Appl Microbiol 36:49–55. doi:10.1016/j.syapm.2012.10.004

Yutin N, Puigbo P, Koonin EV, Wolf YI (2012) Phylogenomics of prokaryotic ribosomal proteins. PLoS ONE 7:e36972. doi:10.1371/journal.pone.0036972

Zhang YM, Tian CF, Sui XH, Chen WF, Chen WX (2012) Robust markers reflecting phylogeny and taxonomy of Rhizobia. PLoS ONE 7:e44936. doi:10.1371/journal.pone.0044936

Zhao L, Fan M, Zhang D, Yang R, Zhang F, Xu L, Wei X, Shen Y, Wei G (2014) Distribution and diversity of Rhizobia associated with wild soybean (*Glycine soja* Sieb. & Zucc.) in Northwest China. Syst Appl Microbiol 37:449–456. doi:10.1016/j.syapm.2014.05.011

Ziegler D, Mariotti A, Pfluger V, Saad M, Vogel G, Tonolla M, Perret X (2012) In situ identification of plant-invasive bacteria with MALDI-TOF mass spectrometry. PLoS ONE 7:e37189. doi:10.1371/journal.pone.0037189