

# In Situ Identification of Plant-Invasive Bacteria with MALDI-TOF Mass Spectrometry

Dominik Ziegler<sup>1,3\*</sup>, Anna Mariotti<sup>1,2\*</sup>, Valentin Pflüger<sup>3</sup>, Maged Saad<sup>1</sup>, Guido Vogel<sup>3</sup>, Mauro Tonolla<sup>1,2</sup>, Xavier Perret<sup>1\*</sup>

**1** Department of Botany and Plant Biology, University of Geneva, Geneva, Switzerland, **2** Institute of Microbiology, Bellinzona, Switzerland, **3** Mabritec AG, Riehen, Switzerland

## Abstract

Rhizobia form a disparate collection of soil bacteria capable of reducing atmospheric nitrogen in symbiosis with legumes. The study of rhizobial populations in nature involves the collection of large numbers of nodules found on roots or stems of legumes, and the subsequent typing of nodule bacteria. To avoid the time-consuming steps of isolating and cultivating nodule bacteria prior to genotyping, a protocol of strain identification based on the comparison of MALDI-TOF MS spectra was established. In this procedure, plant nodules were considered as natural bioreactors that amplify clonal populations of nitrogen-fixing bacteroids. Following a simple isolation procedure, bacteroids were fingerprinted by analysing biomarker cellular proteins of 3 to 13 kDa using Matrix Assisted Laser Desorption/Ionization Time of Flight (MALDI-TOF) mass spectrometry. In total, bacteroids of more than 1,200 nodules collected from roots of three legumes of the *Phaseoleae* tribe (cowpea, soybean or siratro) were examined. Plants were inoculated with pure cultures of a slow-growing *Bradyrhizobium japonicum* strain G49, or either of two closely related and fast-growing *Sinorhizobium fredii* strains NGR234 and USDA257, or with mixed inoculants. In the fully automatic mode, correct identification of bacteroids was obtained for >97% of the nodules, and reached 100% with a minimal manual input in processing of spectra. These results showed that MALDI-TOF MS is a powerful tool for the identification of intracellular bacteria taken directly from plant tissues.

**Citation:** Ziegler D, Mariotti A, Pflüger V, Saad M, Vogel G, et al. (2012) In Situ Identification of Plant-Invasive Bacteria with MALDI-TOF Mass Spectrometry. PLoS ONE 7(5): e37189. doi:10.1371/journal.pone.0037189

**Editor:** Akos Vertes, The George Washington University, United States of America

**Received:** March 2, 2012; **Accepted:** April 18, 2012; **Published:** May 17, 2012

**Copyright:** © 2012 Ziegler et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** XP acknowledges funding from the Swiss National Science Foundation (grant n°3100A0-116591) and the University of Geneva. DZ thanks Mabritec AG for financial support of the research project. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** Co-authors Valentin Pflüger and Guido Vogel are employees of Mabritec AG. Dominik Ziegler is a PhD student in Geneva whose research is partially funded by Mabritec AG. There are no patents or products in development to declare. This does not alter the authors' adherence to all of the PLoS ONE policies on sharing data and materials.

\* E-mail: xavier.perret@unige.ch

† These authors contributed equally to this work.

## Introduction

Of the diverse microorganisms that thrive in the rhizosphere of higher plants, only a few can colonize the inner root tissues. Fewer still can enter plant cells and establish persistent intracellular infections. Amongst them, rhizobia form a heterogeneous group of  $\alpha$ - and  $\beta$ -proteobacteria capable of establishing beneficial nitrogen-fixing associations with legumes [1]. Rhizobia are responsible for as much as half of the nitrogen fixed each year by terrestrial biological systems [2], and are often used in agriculture to supplement plants with the reduced forms of nitrogen absent in many soils around the world. Reduction of atmospheric nitrogen ( $N_2$ ) by rhizobia occurs predominantly inside plant cells found in specialized root (occasionally stem) organs called nodules. To colonize the inner tissues of nodules, rhizobia generally follow a trans-cellular path of infection that crosses the root epidermis as well as several layers of cortical cells. This infection process remains mostly under the control of host plants and rhizobial proliferation is restricted to the tips of infection threads (ITs), which ultimately ramify when reaching the newly formed nodule meristem [3,4]. Plants restrict detrimental infections of root tissues by harmful pathogens or poor nitrogen fixers by using sophisti-

cated screens that involve the coordinated exchange of molecular signals between nodulating rhizobia and host plants [5,6]. Flavonoids released by roots, bacterial surface polysaccharides, nodulation factors (Nod-factors) and type-three secreted proteins released by rhizobia are amongst the most discriminating signals involved in the establishment of a proficient symbiosis. Ultimately, rhizobia are released from ITs into the cytoplasm of nodule cells where they differentiate into nitrogen-fixing bacteroids. In return for reduced nitrogen, plants provide bacteroids with ample carbon supply mostly in the form of dicarboxylates. To protect rhizobial nitrogenase from its irreversible inactivation by traces of free oxygen, plants have evolved a number of mechanisms including a cortical oxygen diffusion barrier that surrounds the infected nodule tissue and the synthesis of leghemoglobin, which together regulate the oxygen tension within infected cells. Leghemoglobin is particularly abundant inside  $N_2$ -fixing nodules where it accounts for as much as 30% of all proteins [7]. As a single rhizobial cell attached at the root-hair tip is sufficient to initiate the formation of an infection thread, a population of bacteroids inside a nodule is mostly clonal. In fact nodules can be regarded as small bioreactors in which each of the infected nodule cells contains several hundred bacteroids. Occasionally infection threads can accommodate

distinct bacteria, and thus may lead to nodules that house a mixed population of rhizobia [3,4].

Matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) mass spectrometry (MS) has become a method of choice for the identification of bacteria [8]. In such procedures, isolated bacteria are first grown as micro-colonies on solid media prior to MALDI-TOF MS analysis. Then, small numbers of cells are lysed to release the intracellular proteins, which in turn will be ionized and separated according to their mass to charge ratio ( $m/z$ ). Protein masses are recorded as distinct peaks that together form a complex spectrum or fingerprint, which is characteristic of a bacterial sample (see Fig. 1). Each spectrum of protein masses is subsequently matched against reference spectra that are stored in dedicated databases, and were initially established using well characterized reference strains and defined growth conditions [9]. The Spectral Archive and Microbial Identification System (SARAMIS<sup>TM</sup>) of database management reduces the complexity of mass spectra by restricting datasets to subsets of peaks from the most abundant proteins or peptides found in the 2 to 20 kDa range. These selected peaks together form the so called SuperSpectra<sup>TM</sup> (SSp) that predominantly include ribosomal proteins as selected biomarkers [10]. Depending on the number of SSp that are stored in a reference database for a given group of bacteria, a sampled strain can be reliably assigned to a known family, genus, species or more rarely a particular subspecies [11,12,13,14,15].

Except for a few protocols in which infectious bacteria from urine or blood samples were concentrated by centrifugation prior to cell typing [16,17,18], the identification of environmental or pathogenic bacteria by MALDI-TOF MS analysis requires preliminary steps of bacterial isolation and purification. Unlike most animal pathogens that form isolated micro-colonies in a matter of hours, plant-interacting bacteria require longer incubation times that, as in the case of bradyrhizobia, may extend over several days. To prevent growth of soil contaminants during such long incubation periods, standard methods for the identification of endosymbiotic rhizobia include a surface-sterilization of root nodules prior to sampling serial dilutions of their contents on plates [19]. Recently, MALDI-TOF MS technology was applied to the identification of fast-growing rhizobia, but only to strains cultivated on solid media [20]. To bypass the initial and time-consuming steps of bacterial isolation and subsequent cultivation, we explored the possibility of applying MALDI-TOF MS analysis directly to populations of endosymbiotic bacteroids found in root nodules. For this study, three symbiotic strains were used in combination with the three host plants: *Vigna unguiculata* (cowpea), *Macroptilium atropurpureum* (siratiro) and *Glycine max* (soybean). With a generation time close to 3 h., the fast-growing *Sinorhizobium fredii* strain NGR234 was selected for its unique ability to induce nodule formation on more than 120 genera of legumes, and fix nitrogen with a subset of at least 150 plants [21]. The other fast-grower *S. fredii* strain USDA257 shares with NGR234 many of its symbiotic genes [22,23], but unlike NGR234 forms N<sub>2</sub>-fixing symbioses with selected soybean cultivars [21]. In contrast, the slow-growing micro-symbiont *Bradyrhizobium japonicum* strain G49 is genetically distant from NGR234 and USDA257 and fixes nitrogen in association with diverse legumes, including *G. max*, *M. atropurpureum* and *V. unguiculata*. Using these distinct strains, we show here that MALDI-TOF MS can allow for the rapid and reliable identification of nodule bacteria, including at the subspecies level, thus paving the way for large-scale population studies of rhizobia isolated from root nodules found on crops of agricultural importance.

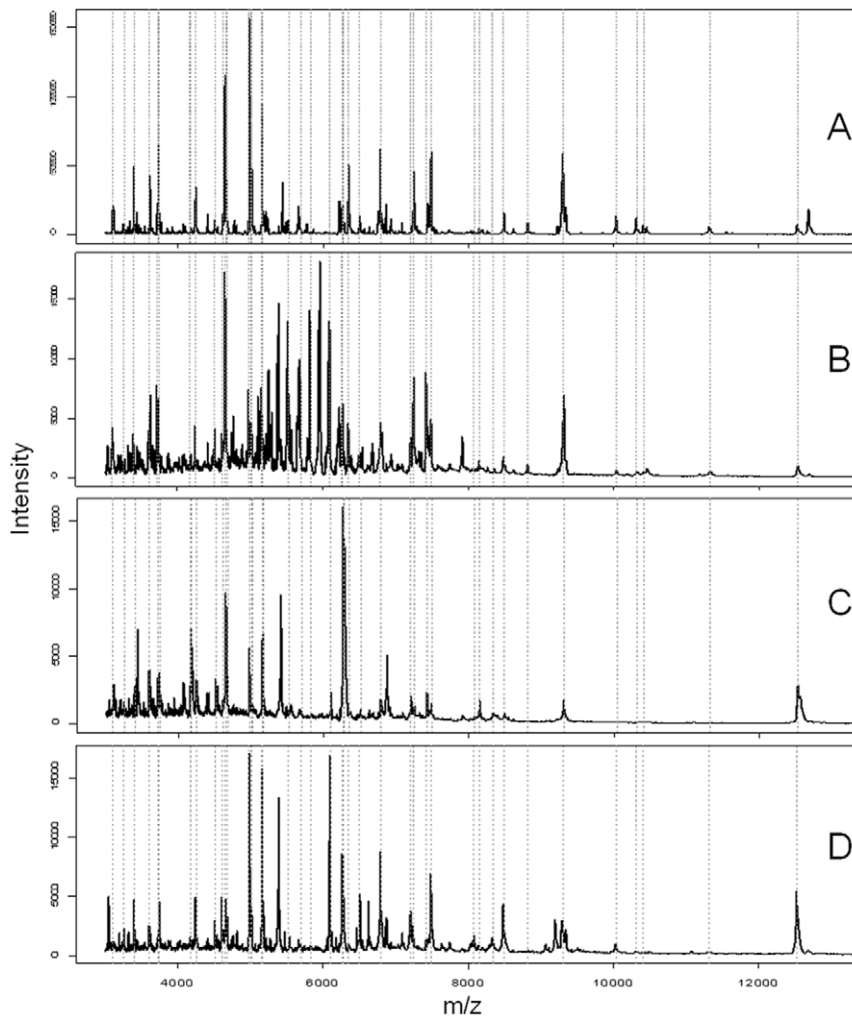
## Results

### MALDI-TOF MS fingerprinting of free-living rhizobia

Reliable identification of bacterial strains requires the selection as biomarkers of protein masses that appear consistently in spectra, irrespective of the growth conditions or age of the cultures. Thus, cells of strains G49, NGR234, and USDA257 were grown at 28°C on solid media for 2, 3, 5, 7, 10, and 14 days. Initially, bacterial samples were subsequently prepared using either (i) the direct smear method, (ii) a suspension in 25% formic acid, or (iii) a pre-purification step for cellular proteins via an extraction with acetonitrile and 70% formic acid. For each strain, 18 averaged spectra of good quality (generated using the 3 methods of sample preparation for each of the 6 incubation periods) were compiled into a single representative SuperSpectrum<sup>TM</sup> (SSp) comprised of 39 to 42 biomarker masses that are listed in Table S1. Because of their abundance, small size and ubiquitous distribution, ribosomal proteins are often selected as biomarkers [10,24] that, in some cases, are sufficiently discriminatory to allow for an accurate typing at the strain level [25,26]. Accordingly, 35 of the 108 biomarkers selected for the SSp of G49, NGR234 and USDA257 were tentatively assigned to ribosomal proteins with corresponding molecular weights that were calculated using the corresponding genomic sequences (see Table S1). As befits strains that belong to the same *Sinorhizobium* genus, the reference spectra of NGR234 and USDA257 shared 12 biomarkers of which 7 were identified as ribosomal proteins. In contrast, none of the 39 biomarkers specific to G49 were found in NGR234 or USDA257, and none of the 108 masses listed in Table S1 were common to all three rhizobial SSp. Ultimately, the sets of biomarker masses were matched against each other and against the whole SARAMIS database (ver. 4.09, system ver. 3.4.1.2, with >70,000 spectra and >2,600 SSp). This procedure showed that an unequivocal strain assignment was reached when at least 60% of the biomarkers of a given spectrum of G49, NGR234 or USDA257 matched those of the reference SSp. Thus, the identification thresholds were set at 24 (from a total of 39), 25 (of 40) and 26 (of 42) matching biomarkers for the spectra of respectively, strains G49, NGR234 and USDA257.

### Generating reproducible spectra of root-nodule bacteria

To establish a reliable protocol of rhizobial identification that was both robust and able to be applied to many symbioses, a number of plant-microsymbiont combinations were tested. Batches of 16 plants of *G. max* cv. Davis, *M. atropurpureum* (DC.) Urb., or *V. unguiculata* cv. Red Caloona were inoculated with either strain G49, NGR234 or USDA257, corresponding to a total of 144 plants in 72 Magenta jars for each legume species. Except for NGR234 that forms uninfected pseudonodules on *G. max* and USDA257 that fixes poorly with *M. atropurpureum*, all remaining plant-bacterium combinations lead to proficient nitrogen-fixing associations. To avoid potential problems linked to senescence of bacteroids, nitrogen-fixing nodules of cowpea, siratiro, and soybean were harvested before flowering time at respectively, 42, 49 and 35 days post-inoculation (dpi) and stored at -60°C until further use. Following surface sterilisation, nodules were processed as described in the materials and methods section to remove contaminating leghemoglobin from the samples. Enriched bacteroid pellets were subsequently treated using either of the three methods described above for free-living bacteria, and two replicates for each nodule. A total of 420 nodules were examined in this way, including 60 nodules for each of the following microsymbiont/host-plant combinations G49/cowpea, G49/soybean, NGR234/cowpea, NGR234/siratiro, USDA257/cowpea and USDA257/soybean, and 30 nodules for the G49/siratiro

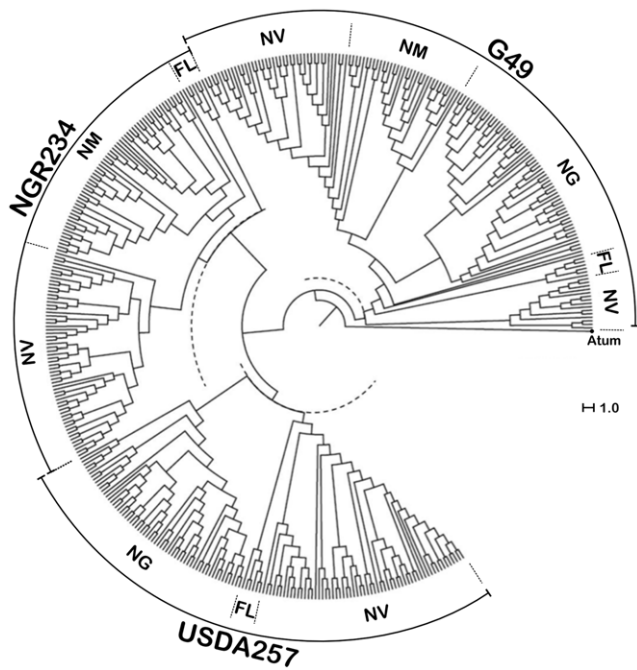


**Figure 1. Typical MALDI mass spectra between  $m/z$  3000 and  $m/z$  13000 of *S. fredii* strain USDA257 grown on TY for 5 days at 28°C (A) or found inside nitrogen-fixing nodules of *G. max* (B), *M. atropurpureum* (C), and *V. unguiculata* (D) collected respectively 35, 49, and 42 days post-inoculation.** Sample preparation included washing steps to remove leghemoglobin, and suspension of bacterial pellets in 25% formic acid. Positions of biomarkers selected for the USDA257 SSp are showed with dashed lines.  
doi:10.1371/journal.pone.0037189.g001

and USDA257/siratro combinations. Representative spectra of free-living or endosymbiotic cells of USDA257 are shown in Figure 1. Using the threshold of 60% biomarkers in common that was established earlier for G49, NGR234 or USDA257 SSp, a positive strain-identification was reached for 389 of the 420 nodules (92%). When the 30 small and poorly infected nodules formed by USDA257 on siratro were excluded from the analysis, the success rate for the identification of nodule bacteria increased to 97% (379 out of 390 nodules). However, nodule bacteria from the remaining 11 unassigned nodules (3%) were identified by matching manually each of the unknown spectra against the representative spectra selected to generate the SSp of G49, NGR234 or USDA257. Thus, with a minimal operator input, bacteroids from all nodules were successfully identified. The circular dendrogram presented in Figure 2 shows the unambiguous clustering in distinct G49, NGR234 or USDA257 groups of 410 individual spectra obtained from free-living bacteria or nitrogen-fixing bacteroids, using *Agrobacterium tumefaciens* as an outgroup. Ultimately, as spectra of optimal quality were produced using the preliminary treatment of isolated bacteria in 25% formic acid, all subsequent samples were prepared in this way.

### A competition assay using MALDI-TOF MS to identify nodule bacteria

In the soil, competition between rhizobia to gain access to the nodule environment is widespread and often limits attempts to use improved strains for field inoculations [27,28]. To assess the reliability of the MALDI-TOF MS protocol for identifying root-nodule bacteria, a competition assay was undertaken. Cowpea was selected as a host plant because of its low selectivity for rhizobia [29]. Sinorhizobia NGR234 (Rif<sup>R</sup>, Km<sup>S</sup>) and USDA257 (Rif<sup>S</sup>, Km<sup>R</sup>) were chosen as microsymbionts because both strains (i) are genetically closely related, (ii) have overlapping host-ranges [21], and (iii) possess distinct antibiotic resistances that facilitate their identification using appropriate selective media. Inoculants of  $2 \times 10^8$  rhizobia per plant were prepared using mixtures of NGR234 and USDA257 cells at ratios of 100/1, 50/50 or 1/100. Plants were harvested 42 dpi and a total of 113 to 120 nodules from up to 11 plants were examined for each of the three inoculants (Table 1). Nodule occupancy was determined using both MALDI-TOF MS spectra of bacteroids and by plating isolated nodule bacteria on selective media. In all conditions tested, NGR234 outcompeted USDA257 for nodulation on *V.*



**Figure 2. Unsupervised hierarchical cluster analysis (Dice algorithm) of G49, NGR234 and USDA257 MALDI-TOF MS mass spectra comprised in the size range of  $m/z$  3000 to 15000.** Bacteria were isolated from free-living cultures (FL), or nodules of *G. max* (NG), *M. atropurpureum* (NM), or *V. unguiculata* (NV). The respective positions of the 60% identity level used as a threshold to assign strains are shown as dotted semi-circles. *Agrobacterium tumefaciens* (Atum) was selected as the outgroup.  
doi:10.1371/journal.pone.0037189.g002

*unguiculata* including when cells of NGR234 were outnumbered by a factor of 100. Although MALDI-TOF MS analysis systematically resulted in the identification of a single nodule strain (most probably the strain with the most abundant bacteroid population), growth on selective media confirmed that 4 to 25% of the nodules were infected with both NGR234 and USDA257. As no bacterial growth was observed on plates on which surface-sterilized nodules were rolled, the sterilization procedure was efficient and cases of nodule occupancy documented as mixed were unlikely to result from contaminating bacteria found on the nodule epidermis. In a similar assay using G49 and NGR234 as competitors, the analysis of an additional 484 nodules showed that G49 outcompeted NGR234 for nodulation on *V. unguiculata* irrespective of the cell ratios used as inoculants (data not shown). These competition assays also showed that a correct identification of the predominant bacteroid population was reached by SARAMIS<sup>TM</sup> in the automatic mode for 753 (90%) of the 837 nodules examined, while strain identification for the remaining 84 nodules required only a minimal operator input.

## Discussion

That NGR234 so efficiently outcompeted USDA257 for nodulation on *V. unguiculata* was surprising since both strains are genetically closely related, secrete Nod-factors of similar structures, and have largely overlapping host-ranges [6]. The absence of functional copies of the *nodSU* genes [30] is probably not the reason for the lower competitiveness of USDA257, as a mutant of NGR234 deleted in *nodSU* nodulates cowpea with the same efficiency as the wild type [31]. It seems likely that competitiveness

**Table 1. Identification of bacteria competing for nodulation on *V. unguiculata*.**

NGR234 vs. USDA257			
Cell ratio	100/1	50/50	1/100
Nodules/plants analyzed	120/10	120/11	113/11
Nodule occupancy	NGR234	115 (95.9)	102 (85.0)
	USDA257	0 (0.0)	5 (4.2)
	mixed – NGR234 spectra	4 (3.3)	4 (3.3)
Numbers of manual inputs in MALDI11 analysis	mixed – USDA257 spectra	1 (0.8)	9 (7.5)
		3 (2.5)	2 (1.8)

% of total nodules are shown in brackets.

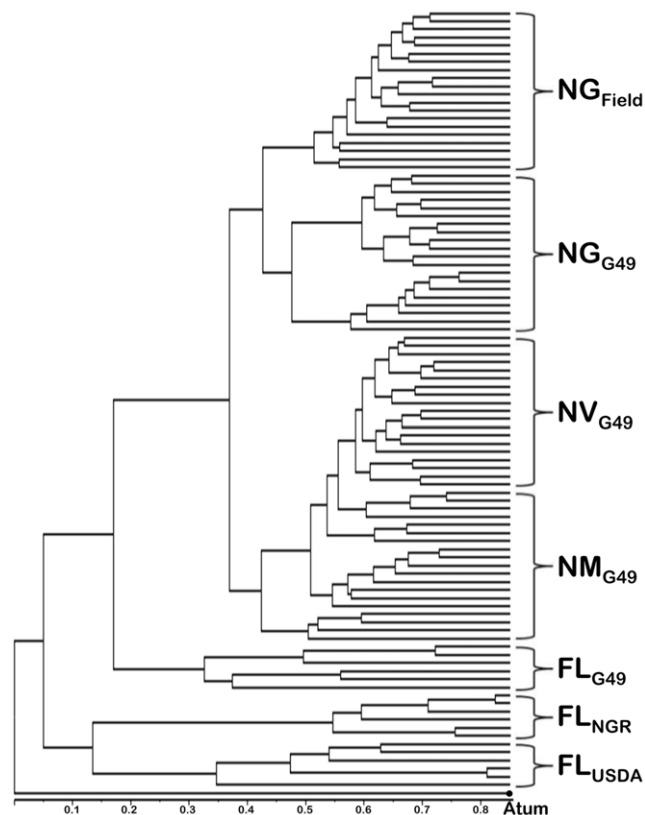
doi:10.1371/journal.pone.0037189.t001

of NGR234 and USDA257 on *V. unguiculata* is not mediated by Nod-factors, but rather due to specific alterations in the composition and structure of their respective cell surface components. Both NGR234 and USDA257 activate type three protein secretion systems (T3SS) in response to flavonoid stimulation [32], yet the presence or absence of a functional T3SS has no measurable effect on the nodulation of cowpea by NGR234 [33]. Instead, flavonoid-inducible modifications in the *O*-antigen portion of lipopolysaccharides were shown to affect the ability of NGR234 to fix nitrogen on *V. unguiculata* [34]. As genes responsible for rhamnan synthesis are absent from USDA257 [23], surface polysaccharides are candidate determinants of the NGR234 competitiveness on cowpea.

More importantly, the competition assays confirmed that our protocol for bacteroid identification via MALDI-TOF MS was capable of discriminating between closely related strains such as NGR234 and USDA257. The possibility to bypass the time-consuming step of cultivating individual colonies on plates to reach a reliable identification of nodule bacteria by MALDI-TOF MS opens up new perspectives. According to our data, bacteroids found within a single mature nitrogen-fixing nodule are in sufficient numbers to generate reproducible spectra of good quality. Except for the abundant leghemoglobin that needs to be washed away to prevent the saturation of spectra, nodule cells or their proteins do not seem to interfere with the process of bacteroid identification. Although spectra of USDA257 isolated from nodules of distinct plants appeared to differ (see Fig. 1), a correct identification was ultimately reached because only the presence or absence (and not the respective signal intensities) of specific biomarker masses were used as discriminating quanta. That one SSP generated from spectra of free-living rhizobia was sufficient to reach a correct strain identification, irrespective of the plant from which the bacteroids were collected, facilitates further developments. The need to improve the SARAMIS database with SSP's of both free-living and endosymbiotic rhizobia would have been impractical, given that some promiscuous strains may nodulate more than one hundred genera of legumes [21]. Currently, a rhizobia-specific database is being established using well-characterized fast- and slow-growing strains belonging to the  $\alpha$ - and  $\beta$ -proteobacteria. Once established, such a database is bound to facilitate the implementation of MALDI-TOF MS as a tool to study the great diversity in populations of symbiotic

rhizobia in agricultural or natural soils. To test whether our protocol for strain identification was efficient for such a purpose, 24 nodules were collected in a field of mature soybeans in the Swiss state of Ticino. MALDI-TOF MS spectra of the bacteroids found within these nodules were matched against those of cells of G49, NGR234 and USDA257 isolated from nodules of plants grown in laboratory conditions. Figure 3 shows that field-nodule bacteria clustered together with G49 bacteroids isolated from nodules of *G. max* cv. Davis while free-living cells and bacteroids of G49 found in cowpea or siratro formed distinct subclusters. This result is consistent with the use of *B. japonicum* strains as commercial inoculants for soybean cultures (including in Switzerland), and confirms that direct identification via MALDI-TOF MS of strains found in environmental samples is possible.

However, nodules come in many forms and shapes in nature, and the maturation of bacteroids and their fate once nodules senesce vary considerably in a host-dependent manner [1]. For example, nodules of cowpea, siratro and soybean, are all of determinate type with populations of bacteroids being regarded as relatively homogenous and to some extent capable of resuming growth once outside of plant cells [35,36]. In contrast, nodules of the indeterminate type are characterized by the presence of a persistent meristem, and several developmental zones within which bacteroids display distinct differentiation states [37]. Legumes of the inverted repeat-lacking clade (IRLC) such as



**Figure 3. Typing of bacteroids found in nodules of a soybean field.** Unsupervised hierarchical cluster analysis (Dice algorithm) of complex MALDI-TOF MS spectra ( $m/z$  3000 to 15000) of rhizobia found inside nodules of *G. max* from a field in the Swiss state of Ticino ( $NG_{Field}$ ), or of G49 isolated from nodules of soybean ( $NG_{G49}$ ), siratro ( $NM_{G49}$ ), and cowpea ( $NV_{G49}$ ). Spectra from free-living cells of G49 ( $FL_{G49}$ ), NGR234 ( $FL_{NGR}$ ), and USDA257 ( $FL_{USDA}$ ), were also included. *A. tumefaciens* (*Atum*) was used as an outgroup. doi:10.1371/journal.pone.0037189.g003

*Medicago sativa*, form indeterminate nodules in which nitrogen-fixing bacteroids are terminally differentiated [38]. Terminal differentiation results from the action of abundant nodule-specific and cystein-rich peptides (NCRs) of 3 to 5 kDa that induce the endoreduplication of the bacteroid genome and a concomitant cell enlargement [39,40]. It is unclear whether the endoreduplicated state of *Sinorhizobium meliloti* bacteroids in indeterminate nodules of *M. sativa* may affect the process of strain identification by MALDI-TOF MS. In this respect, further studies will also confirm whether NCRs that are ultimately targeted to bacteroids, can be identified by comparing the spectra of free-living and endosymbiotic cells of *S. meliloti*. Nevertheless, our current protocol of strain identification suits many agronomically important legume crops such as soybean, bean (*Phaseolus vulgaris*) or pigeon pea (*Cajanus cajan*), all of which develop determinate nodules on their roots.

## Materials and Methods

### Characteristics of bacteria and growth conditions

Derivatives of *S. fredii* strains NGR234 resistant to rifampicin ( $Rif^R$ ) [41] and USDA257 resistant to kanamycin ( $Km^R$ ) [42] were grown at 27°C in/on TY [43], supplemented with antibiotics at final concentrations of 50  $\mu\text{g ml}^{-1}$ . *B. japonicum* strain G49 was isolated from a batch of the commercial inoculum HiStick Soybean (Becker Underwood, Littlehampton, West Sussex, UK) and grown in/on yeast-mannitol (YM) medium. Strain G49 was sensitive to both, rifampicin and kanamycin ( $Rif^S$ ,  $Km^S$ ).

### Plant assays

Surface sterilized seeds of *G. max*, *M. atropurpureum* and *V. unguiculata* were dispersed on agar plates and incubated to germinate in the dark, at 27°C. Once germinated, seedlings were planted in Magenta jars containing vermiculite [31], and watered using nitrogen free B&D solution [44]. Two to three days after their transfer to Magenta jars, each plantlet was inoculated with 200  $\mu\text{l}$  of a solution of  $2 \times 10^8$  bacteria. For competition assays, *V. unguiculata* plantlets were inoculated with either 1/100, 50/50 or 100/1 titters of the NGR234/USDA257 or NGR234/G49 combinations. All plants were grown 6 to 8 weeks post inoculation at a day temperature of 27°C, a night temperature of 20°C and a light phase of 12 hours.

### Sample preparation for MALDI-TOF MS identification of free-living bacteria

Prior to spotting MALDI steel target plates, bacteria of G49, NGR234 or USDA257 were prepared using either of the three protocols: (i) direct smear method using a disposable loop, (ii) suspension in 25% formic acid, and (iii) extraction with acetonitrile (ACN) and formic acid (70%). Then, bacterial samples were spotted twenty times on MALDI steel target plates after growth periods at 28°C of 2, 3, 5, 7, 10 and 14 days on plates. Spots were overlaid with 1  $\mu\text{l}$  of matrix consisting of a saturated solution of alpha-cyano-4-hydroxycinnamic acid (Sigma-Aldrich, Buchs, Switzerland) in 33% acetonitrile (Sigma-Aldrich), 33% ethanol and supplemented with 3% trifluoroacetic acid, and air-dried within minutes at room temperature.

### Preparation of nodule bacteria

Once harvested, nodules were frozen and kept at  $-60^\circ\text{C}$  until further use. Prior to isolation of endosymbiotic bacteria, nodules were surface sterilized in 70% ethanol and crushed into 400  $\mu\text{l}$  of sterile  $\text{ddH}_2\text{O}$ . For replica plating of nodule bacteria, 20  $\mu\text{l}$  of these homogenates were transferred into independent wells of 96 MicroWell Plates Round Bottom of Nunc (Nunc GmbH & Co.

KG, Langensbold, Germany), and aliquots were subsequently plated on selective media. The rest of the crude supernatants were transferred free of nodule debris into new microfuge tubes and were centrifuged at 20,000 g to collect bacteroids. The bacterial pellets were washed three times with 200  $\mu$ l of sterile ddH<sub>2</sub>O to remove plant leghemoglobin. Bacteroids were prepared for MALDI-TOF MS fingerprinting as described above.

### MALDI-TOF fingerprinting

Protein mass fingerprints were obtained using a MALDI-TOF Mass Spectrometry Axima<sup>TM</sup> Confidence machine (Shimadzu-Biotech Corp., Kyoto, Japan), with detection in the linear positive mode at a laser frequency of 50 Hz and within a mass range from 2,000–20,000 Da. Acceleration voltage was 20 kV, and the extraction delay time was 200 ns. A minimum of 10 laser shots per sample was used to generate each ion spectrum. For each bacterial sample, a total of 50 protein mass fingerprints were averaged and processed using the Launchpad<sup>TM</sup> 2.8 software (Shimadzu-Biotech Corp., Kyoto, Japan). For peak processing of raw spectra, the following settings were selected for Launchpad<sup>TM</sup> 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 threshold response factor of 1.2. Each target plate was first externally calibrated using spectra of the reference strain *Escherichia coli* DH5 $\alpha$ . In addition, the spectra that were acquired to generate the Superspectra were internally calibrated using strain-specific sets of marker masses at mass-to-charge-ratio ( $m/z$ ) of 4990, 5090, 6188, 6848, 7035, 7480, 7840, 8511 and 8950 for *B. japonicum* strain G49;  $m/z$  4979, 5153, 7255, 7484, 8484, 8816, 9239, 10007, 10313 and 11333 for *S. fredii* strain NGR234; and  $m/z$  4979, 5153, 7240, 7484, 8144, 8484, 8815, 9305, 10034, 10312, and 11334 for *S. fredii* strain USDA257. The processed spectra were exported as ASCII datasets with each peak being characterized by an  $m/z$  value and a signal intensity.

### References

1. 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.
2. Herridge DF, Peoples MB, Boddey RM (2008) Global inputs of biological nitrogen fixation in agricultural systems. Plant Soil 311: 1–18.
3. Gage DJ (2002) Analysis of infection thread development using Gfp- and DsRed-expressing *Sinorhizobium meliloti*. J Bacteriol 184: 7042–7046.
4. Gage DJ (2004) Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes. Microbiol Mol Biol Rev 68: 280–300.
5. Jones KM, Kobayashi H, Davies BW, Taga ME, Walker GC (2007) How rhizobial symbionts invade plants: the *Sinorhizobium-Medicago* model. Nat Rev Microbiol 5: 619–633.
6. Perret X, Staehelin C, Broughton WJ (2000) Molecular basis of symbiotic promiscuity. Microbiol Mol Biol Rev 64: 180–201.
7. Baulcombe D, Verma DP (1978) Preparation of a complementary DNA for leghaemoglobin and direct demonstration that leghaemoglobin is encoded by the soybean genome. Nucleic Acids Res 11: 4141–4155.
8. Sauer S, Kliem M (2010) Mass spectrometry tools for the classification and identification of bacteria. Nat Rev Microbiol 8: 74–82.
9. Cherkaoui A, Hibbs J, Emonet S, Tangomo M, Girard M, et al. (2010) Comparison of two matrix-assisted laser desorption ionization-time of flight mass spectrometry methods with conventional phenotypic identification for routine identification of bacteria to the species level. J Clin Microbiol 48: 1169–1175.
10. Fenselau C, Demirev PA (2001) Characterization of intact microorganisms by MALDI mass spectrometry. Mass Spectrom Rev 20: 157–171.
11. Stephan R, Ziegler D, Pfluger V, Vogel G, Lehner A (2010) Rapid genus- and species-specific identification of *Cronobacter* spp. by matrix-assisted laser

### Generation of Superspectra and cluster analysis

Fingerprints of protein masses were analyzed with SARAMIS<sup>TM</sup> (Spectral Archive and Microbial Identification System, AnagnosTec, Potsdam-Golm, Germany). For each preparation, lists of peaks were imported and binary matrix was calculated using the SARAMIS<sup>TM</sup> Superspectrum tool with an error window of 800 ppm. Average spectra were generated with Excel macros in which noise masses found in less than half of the twenty replicates were eliminated, resulting in 18 representative averaged spectra (for three preparations and 6 cultivation periods) per strain. For each of the three selected strains, a SuperSpectrum<sup>TM</sup> consisting in a specific pattern of biomarker masses was calculated using the SARAMIS<sup>TM</sup> Superspectrum tool. For cluster analysis of protein mass fingerprints, binary matrices were generated using the SARAMIS<sup>TM</sup> Superspectra tool, and exported as text files. Data was imported into the PAST software (Natural History Museum, Oslo University, Norway) and multivariate neighbor joining cluster analysis was performed using the Dice algorithm to calculate distance matrices. Dendograms were drawn using the FigTree v1.3.1 software (<http://tree.bio.ed.ac.uk/software/figtree/>) and distance matrices formatted as nexus file.

### Supporting Information

**Table S1** Selected biomarkers for SuperSpectra of strains G49, NGR234 and USDA257. (DOCX)

### Acknowledgments

We thank Natalia Giot for her help in many aspects of this work, as well as Gérard Hopfgartner and Franco Widmer for their insightful inputs. We would like to acknowledge Julie Ardley for her critical reading of the manuscript.

### Author Contributions

Conceived and designed the experiments: XP MT AM DZ. Performed the experiments: AM DZ MS. Analyzed the data: AM DZ VP GV XP MT. Contributed reagents/materials/analysis tools: XP MT GV. Wrote the paper: XP DZ AM MT.

- desorption ionization-time of flight mass spectrometry. J Clin Microbiol 48: 2846–2851.
12. Stephan R, Cernela N, Ziegler D, Pfluger V, Tonolla M, et al. (2011) Rapid species specific identification and subtyping of *Yersinia enterocolitica* by MALDI-TOF mass spectrometry. J Microbiol Met 87: 150–153.
13. Freiwald A, Sauer S (2009) Phylogenetic classification and identification of bacteria by mass spectrometry. Nat Protoc 4: 732–742.
14. Hahn D, Mirza B, Benagli C, Vogel G, Tonolla M (2011) Typing of nitrogen-fixing *Frankia* strains by matrix-assisted laser desorption ionization-time-of-flight (MALDI-TOF) mass spectrometry. Syst Appl Microbiol 34: 63–68.
15. Rezzonico F, Vogel G, Duffy B, Tonolla M (2010) Application of whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry for rapid identification and clustering analysis of *Pantoea* species. Appl Environ Microbiol 76: 4497–4509.
16. Ferreira L, Sanchez-Juanes F, Gonzalez-Avila M, Cembrero-Fucinos D, Herrero-Hernandez A, et al. (2009) Direct identification of urinary tract pathogens from urine samples by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol 48: 2110–2115.
17. Prod'hom G, Bizzini A, Durussel C, Bille J, Greub G (2010) Matrix-assisted laser desorption ionization-time of flight mass spectrometry for direct bacterial identification from positive blood culture pellets. J Clin Microbiol 48: 1481–1483.
18. Stevenson LG, Drake SK, Murray PR (2010) Rapid identification of bacteria in positive blood culture broths by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol 48: 444–447.
19. Somasegaran P, Hoben HJ (1985) Methods in Legume-Rhizobium technology. . 510 p.

20. Ferreira L, Sánchez-Juanes F, García-Fraile P, Rivas R, Mateos PF, et al. (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.
21. Pueppke SG, Broughton WJ (1999) *Rhizobium* sp. strain NGR234 and *R. fredii* USDA257 share exceptionally broad, nested host ranges. Mol Plant-Microbe Interact 12: 293–318.
22. Perret X, Broughton WJ (1998) Rapid identification of *Rhizobium* strains by Targeted PCR Fingerprinting. Plant Soil 204: 21–34.
23. Perret X, Fellay R, Bjourson AJ, Cooper JE, Brenner S, et al. (1994) Subtraction hybridisation and shot-gun sequencing: a new approach to identify symbiotic loci. Nucleic Acids Res 22: 1335–1341.
24. Sun LW, Teramoto K, Sato H, Torimura M, Tao H, et al. (2006) Characterization of ribosomal proteins as biomarkers for matrix-assisted laser desorption/ionization mass spectral identification of *Lactobacillus plantarum*. Rapid Commun Mass Spectrom 20: 3789–3798.
25. Hotta Y, Teramoto K, Sato H, Yoshikawa H, Hosoda A, et al. (2010) Classification of genus *Pseudomonas* by MALDI-TOF MS based on ribosomal protein coding in *S10-spe*-alpha operon at strain level. J Proteome Res 9: 6722–6728.
26. Teramoto K, Sato H, Sun L, Torimura M, Tao H, et al. (2007) Phylogenetic classification of *Pseudomonas putida* strains by MALDI-MS using ribosomal subunit proteins as biomarkers. Anal Chem 79: 8712–8719.
27. Denison RF, Kiers ET (2004) Lifestyle alternatives for rhizobia: mutualism, parasitism, and forgoing symbiosis. FEMS Microbiol Lett 237: 187–193.
28. Kuykendall LD, Weber DF (1978) Genetically marked *Rhizobium* identifiable as inoculum strain in nodules of soybean plants grown in fields populated with *Rhizobium japonicum*. Appl Environ Microbiol 36: 915–919.
29. Lewin A, Rosenberg C, Meyer zAH, Wong CH, Nelson L, et al. (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.
30. Krishnan HB, Lewin A, Fellay R, Broughton WJ, Pueppke SG (1992) Differential expression of *nodS* accounts for the varied abilities of *Rhizobium fredii* USDA257 and *Rhizobium* sp. strain NGR234 to nodulate *Leucaena* spp. Mol Microbiol 6: 3321–3330.
31. Lewin A, Cervantes E, Wong C-H, Broughton WJ (1990) *nodSU*, two new *nod* genes of the broad host range *Rhizobium* strain NGR234 encode host-specific nodulation of the tropical tree *Leucaena leucocephala*. Mol Plant-Microbe Interact 3: 317–326.
32. Deakin WJ, Broughton WJ (2009) Symbiotic use of pathogenic strategies: rhizobial protein secretion systems. Nat Rev Microbiol 7: 312–320.
33. Viprey V, Del Greco A, Golinowski W, Broughton WJ, Perret X (1998) Symbiotic implications of type III protein secretion machinery in *Rhizobium*. Mol Microbiol 28: 1381–1389.
34. Broughton WJ, Hanin M, Relić B, Kopcińska J, Golinowski W, et al. (2006) Flavonoid-inducible modifications to rhamnan O antigens are necessary for *Rhizobium* sp. strain NGR234-legume symbioses. J Bacteriol 188: 3654–3663.
35. Oke V, Long SR (1999) Bacteroid formation in the *Rhizobium*-legume symbiosis. Curr Opin Microbiol 2: 641–646.
36. Patriarca EJ, Taté R, Federova E, Riccio A, Defez R, et al. (1996) Down-regulation of the *Rhizobium ntr* system in the determinate nodule of *Phaseolus vulgaris* identifies a specific developmental zone. Mol Plant-Microbe Interact 11: 243–251.
37. Sprent JI, James EK (2007) Legume evolution: where do nodules and mycorrhizas fit in? Plant Physiol 144: 575–581.
38. Kereszt A, Mergaert P, Kondorosi E (2011) Bacteroid development in legume nodules: evolution of mutual benefit or of sacrificial victims? Mol Plant Microbe Interact 24: 1300–1309.
39. Mergaert P, Uchiumi T, Alunni B, Evanno G, Cheron A, et al. (2006) Eukaryotic control on bacterial cell cycle and differentiation in the *Rhizobium*-legume symbiosis. Proc Natl Acad Sci USA 103: 5230–5235.
40. Van de Velde W, Zehirov G, Szatmari A, Debreczeny M, Ishihara H, et al. (2010) Plant peptides govern terminal differentiation of bacteria in symbiosis. Science 327: 1122–1126.
41. Stanley J, Dowling DN, Broughton WJ (1988) Cloning of *hemA* from *Rhizobium* sp. NGR234 and symbiotic phenotype of a gene-directed mutant in diverse legume genera. Mol Gen Genet 215: 32–37.
42. Heron DS, Ersek T, Krishnan HB, Pueppke S (1989) Nodulation mutants of *Rhizobium fredii* USDA257. Mol Plant-Microbe Interact 2: 4–10.
43. Beringer JE (1974) R factor transfer in *Rhizobium leguminosarum*. J Gen Microbiol 84: 188–198.
44. Broughton WJ, Dilworth MJ (1971) Control of leghaemoglobin synthesis in snake beans. Biochem J 125: 1075–1080.