Division of Microbiology and Biotechnology
Department of Food and Environmental Sciences
University of Helsinki
Finland

Genetic and Metabolic Features in Host Specificity and Stress Tolerance of Potential Inoculant Rhizobia

Petri Penttinen

ACADEMIC DISSERTATION
To be presented, with the permission of the Faculty of Agriculture and Forestry of the University of Helsinki, for public examination in Infocenter Korona Lecture hall 2, Viikinkaari 11, on September 19th 2014, at 12 o’clock noon.

Helsinki 2014
"That's Life"
(Kay & Gordon, 1964)
# TABLE OF CONTENTS

List of original papers........................................................................................................... 6
   The contribution of the author to the papers ................................................................. 6
List of abbreviations........................................................................................................... 7
Summary ............................................................................................................................ 8
Tiivistelmä (Summary in Finnish)................................................................................... 10
Acknowledgements ......................................................................................................... 12
1 Introduction .................................................................................................................... 14
   1.1 Biological nitrogen fixation...................................................................................... 14
   1.2 The host plants in this study ................................................................................... 15
   1.3 Rhizobia .................................................................................................................. 16
   1.4 Nodulation ............................................................................................................. 17
   1.5 Symbiont and host specificity .................................................................................. 17
   1.6 Nodulation in the field............................................................................................ 20
2 Aims and outline of the study ....................................................................................... 22
3 Materials and methods................................................................................................. 23
   3.1 Isolation and nitrogen fixation efficiency testing ................................................. 23
   3.2 Grouping by phenotypic and genotypic methods .................................................. 23
   3.3 Detailed characterization of LCOs ......................................................................... 24
   3.4 Comparative genomic hybridization ...................................................................... 26
   3.5 Gene sequence analysis ......................................................................................... 26
4 Results and discussion............................................................................................................. 28
  4.1 The specificity of nodulation.......................................................................................... 28
  4.2 Phenotypic and genetic characteristics......................................................................... 28
  4.3 Rhizobial signaling molecules...................................................................................... 29
  4.4 Nodulation genes........................................................................................................... 30
  4.5 What is the role of all the LCO variants in symbiosis?.............................................. 30
  4.6 Determinants of symbiotic performance..................................................................... 31
  4.7 Stress tolerance and response...................................................................................... 32
  4.8 Accessory gene diversity............................................................................................... 33
5 Conclusions ......................................................................................................................... 34
6 Future prospects .................................................................................................................. 35
7 References .......................................................................................................................... 36
LIST OF ORIGINAL PAPERS


Papers I-IV are reproduced with the kind permission of the original publishers.

The contribution of the author to the papers

I. Petri Penttinen (ex-Nowak) conducted the experiments, analyzed the data, interpreted the results, wrote the paper and was the corresponding author.

II. Together with Xu Kaiwei, Petri Penttinen interpreted the results and wrote the paper.

III. Together with Xu Kaiwei, Petri Penttinen interpreted the results and wrote the paper as a co-first author. Petri Penttinen was one of the corresponding authors.

IV. Petri Penttinen designed and conducted the experiments, analyzed the data, interpreted the results, wrote the paper and was the corresponding author.

V. Petri Penttinen analysed the data, interpreted the results and wrote the manuscript together with Dario Greco and Victoria Muntyan. The manuscript has been submitted with Petri Penttinen as the corresponding author.
**LIST OF ABBREVIATIONS**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CGH</td>
<td>Comparative genomic hybridization</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>ES</td>
<td>Electrospray</td>
</tr>
<tr>
<td>GlcNAc</td>
<td>N-acetyl glucosamine</td>
</tr>
<tr>
<td>HPLC</td>
<td>High performance liquid chromatography</td>
</tr>
<tr>
<td>IGS</td>
<td>Intergenic spacer</td>
</tr>
<tr>
<td>LCO</td>
<td>Lipochitooligosaccharide signalling molecule</td>
</tr>
<tr>
<td>m/z</td>
<td>Mass per charge</td>
</tr>
<tr>
<td>MLSA</td>
<td>Multilocus sequence analysis</td>
</tr>
<tr>
<td>MS</td>
<td>Mass spectrometry</td>
</tr>
<tr>
<td>MS/MS</td>
<td>Tandem mass spectrometry</td>
</tr>
<tr>
<td>N</td>
<td>Nitrogen</td>
</tr>
<tr>
<td>N(_2)</td>
<td>Gaseous di-nitrogen</td>
</tr>
<tr>
<td>nod</td>
<td>Nodulation</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>pSym</td>
<td>Symbiotic plasmid</td>
</tr>
<tr>
<td>RFLP</td>
<td>Restriction fragment length polymorphism</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>rRNA</td>
<td>Ribosomal ribonucleic acid</td>
</tr>
<tr>
<td>sp.</td>
<td>Species (singular)</td>
</tr>
<tr>
<td>spp.</td>
<td>Species (plural)</td>
</tr>
<tr>
<td>sv.</td>
<td>Symbiovar</td>
</tr>
<tr>
<td>Tg</td>
<td>Tera grams, million tons</td>
</tr>
<tr>
<td>TLC</td>
<td>Thin layer chromatography</td>
</tr>
</tbody>
</table>
Nitrogen is an essential nutrient for plants. Biological and industrial nitrogen fixation are the major sources of soil nitrogen. The energy intensive industrial nitrogen fixation results in the release of 300 Tg CO₂ from fossil fuel annually. Therefore, encouraging the use of biological nitrogen fixation (BNF) serves a means to mitigate green-house gas emissions.

This study focuses on rhizobia that are the bacterial partners in the biologically nitrogen fixing symbiosis between leguminous plants and bacteria. In symbiosis, the plant and the rhizobia form a specified organ called nodule. Root nodulation starts from the attachment of rhizobia to the plant root. Rhizobia induce the cells of the inner root cortex to divide, giving rise to the nodule. The cells of the outer cortex are induced to form infection threads. Rhizobia enter the nodule through the infection thread and finally differentiate into nitrogen fixing forms.

In symbioses each rhizobial strain nodulates one or more particular plant species. Likewise, the plant may be nodulated by one or more particular rhizobial strain. The specificity of symbiosis is governed by molecules involved in the interaction between the partners. Phenolic molecules secreted by the plant roots induce the rhizobia to produce lipochitooligosaccharide signalling molecules (LCOs). LCOs are synthesized by proteins encoded by nodulation (nod) genes. LCOs have an amino sugar backbone consisting of N-acetylglucosamine residues (GlcNAc). All the LCOs carry an acyl group on the non-reducing terminal GlcNAc residue. The other LCO decorations include acetyl, carbamoyl, glycosyl, methyl and sulphate groups. Plants perceive the LCO signal by receptor kinases, and ultimately the molecular interaction leads to formation of a nodule. The receptor binding activity depends on the length of the LCO backbone. There is only indirect evidence on the role of the other LCO characteristics in symbiosis. In addition to the LCO signalling, rhizobial surface polysaccharides and plant lectins participate in the recognition of compatible partners.

When growing legumes in the field, inoculating the plants with compatible rhizobia usually results in better growth. The soil may not contain rhizobial strains that nodulate selective legumes. The promiscuous legumes are easily nodulated, yet most of the nodulating strains may be ineffective in nitrogen fixation. Environmental conditions, e.g. soil salinity, affect nodulation.

The general aim of this study was to gain knowledge on the features of potential inoculant rhizobia strains. The specific aims were to assess host specific characteristics and the effect of salt stress on rhizobial signalling molecules, LCOs, that are needed for successful symbiosis; to find potential inoculant strains for a promiscuous legume tree; and to characterize potential inoculant strains isolated from promiscuous legume trees and forage legumes to assess genetic features behind the characteristics of the strains.

To reach the aims, LCOs of Sudanese *Sinorhizobium arboris* HAMBI 2361 and *S. kostiense* HAMBI 2362 nodulating *Prosopis chilensis* and *Acacia senegal*, respectively, were determined using mass spectrometry (Papers I, IV); rhizobial
strains were isolated from *Leucaena leucocephala* growing in China, and characterized with phenotypic and genotypic methods (Papers II, III); and the genetic features behind the desirable characteristics of potential inoculant strains isolated from *Medicago minima*, *M. officinalis* and *Prosopis chilensis* from Algeria, Kazakhstan and Sudan, respectively, were analyzed by comparative genomic hybridization (CGH) using a model rhizobium *S. meliloti* Rm1021 microarray (Paper V).

Altogether *S. arboris* HAMBI 2361 and *S. kostiense* HAMBI 2362 produced 31 and 13 LCO variants, respectively, that were N-methylated, partially carbamoylated and partially sulfated, and acylated with common fatty acids (Papers I, IV). The LCOs were similar to those of other *Acacia*-nodulating rhizobia. The *L. leucocephala* nodulating *Ensifer* strains, representing at least three species, shared a common *nodC*, implying that their LCOs might be similar (Paper II). Similarly, no divergence was detected between the *S. meliloti* model strain and the *S. meliloti* wild type strains nodulation genes (Paper V). Even though rhizobia having a shared host range produced similar LCOs and carried similar *nod* genes, comparison of major LCOs between rhizobia with different host ranges suggested that LCOs have only a limited role in host specificity (Papers I, II, IV, V).

*L. leucocephala* hosted eleven rhizobial strains that were efficient in nitrogen fixation making them good candidates as inoculants (Paper II). Ten isolates had a growth slowing effect on the host (Paper II). The variation in effectiveness between the isolates implied that the ability to compete with other strains is an important characteristic for the inoculant of a promiscuous host; the inoculant is of no use if nodules are mainly occupied by less efficient strains. The species distribution was different from those isolated from *L. leucocephala* in other locations, suggesting that strains adapted to local soil conditions would be preferred when choosing inoculants.

The genotypic and phenotypic characteristics of *Leucaena* isolates did not distinguish the efficient from the inefficient and parasitic isolates (Papers II, III). The divergent transposase coding genes of the *S. arboris* and wild-type *S. meliloti* strains suggested that the osmotic shock response of these strains is considerably different to that of *S. meliloti* Rm1021. CGH revealed that 44 *S. arboris* HAMBI 1552 translation, ribosomal structure and biogenesis genes were duplicated, thus the stress tolerance of *S. arboris* HAMBI 1552 might be partially due to a more efficient protein biosynthesis. The genes related to the induction of LCO synthesis and LCO secretion were duplicated in wild-type *S. meliloti* strains, which is possibly connected to the efficient nodulation capabilities of the strains (Paper V). The LCO production of a salt-stressed *S. arboris* HAMBI 2361 was approximately one tenth of that of the non-stressed one, partially explaining the decrease in nodulation under salt stress (Paper IV). The *Leucaena* isolates were phenotypically diverse (Paper III), and the *S. arboris* and wild-type *S. meliloti* strains diverged from the model rhizobium *S. meliloti* Rm1021 mainly in the accessory genome implying that the divergence was important in shaping the adaptability of the strains (Paper V). However, it is impossible to conclude if the diversity and divergence gave the strains any competitive advantage.
TIIVISTELMÄ (SUMMARY IN FINNISH)

Kasvit eivät kasva ilman typpeää. Maaperän typen tärkeimmät lähteet ovat teollinen ja biologinen typensidonta. Teollinen typensidonta vaatii paljon energiaa ja tuottaa ilmakehään 300 miljoonaa tonnia hiilidioksidia vuodessa, joten biologisen typensidonnan lisääminen auttaa vähentämään kasvihuonekaasupäästöjä.


Tämän työn tarkoituksena oli saada uutta tietoa siirrostamiseen sopivien ritsobikantojen ominaisuuksista. Tarkempina tavoitteina oli i) tutkia ritsobin viestimolekyylein rakenteiden vaikutusta isäntäspesisisyyteen ja suolastressin vaikutusta viestimolekyylein tuottoon, ii) löytää sopiva ritsobi vähän valikoivalle palkokasvipuulle, ja iii) selvittää siirrostamiseen sopivien ritsobikantojen ominaisuuksien taustalla olevia geneettisiä piirteitä.

Työssä määritettiin kahden trooppisia palkokasvipuita (Prosopis chilensis ja Acacia senegal) nystyröivän ritsobikannan, Sinorhizobium arboris HAMBI 2361 and S. kostiens HAMBI 2362, viestimolekyylein rakenteet massaspektrometrisesti. Suolastressin vaikutus viestimolekyyleinen tuottoon selvitettiin analysoimalla typen


ACKNOWLEDGEMENTS

I am grateful for the excellent working facilities at the Department of Food and Environmental Sciences, University of Helsinki, where most of this work was done. Professor Kaarina Sivonen is acknowledged for her extremely efficient manner in helping me through the last stages of this dissertation work. I am thankful for the funding from University of Helsinki, Nessling Foundation, Niemi-säätiö and Academy of Finland.

I sincerely thank my supervisor Professor Kristina Lindström for her guidance and encouragement during all these years. It’s been a long way, indeed…

All the co-authors of the publications deserve my wholehearted thanks. I am especially thankful to Professor Jane Thomas-Oates and Docent Leena Räsänen for showing me how data gives rise to results and conclusions after careful examination. Professor Zhang Xiaoping and Assistant Professor Xu Kaiwei are honestly acknowledged for giving me the opportunity to collaborate with their hardworking team.

Professor Mette Svenning and Docent Aino Smolander, the pre-examiners of this thesis, are acknowledged for their valuable comments and advice that showed their serious attitude towards pre-examination and resulted in improving the quality of this work.

All the former and current members of the nitrogen research group are earnestly thanked. Working with you has made working fun! I envy myself for having such a bunch of excellent colleagues;)

Kiitokset äidilleni Helena Pietariselle sekä veljille ja siskolle siitä ettette ihmetelleet tuleeko tästä työstä ikinä valmistaa.

Suurimmat kiitokset puolisolleni Katja Pellikalle sekä lapsilleni Virnalle ja Väinölle siitä että olette olemassa!
1 INTRODUCTION

1.1 Biological nitrogen fixation

Nitrogen is an essential nutrient for plants. Even though nitrogen is the fifth most abundant element in the universe (Heiserman, 1992), the availability of nitrogen in soil is often a limiting factor for plant growth (Ågren et al., 2012). Nitrogen fixation by reducing N\textsubscript{2} to ammonium is the main source of soil nitrogen. This study focuses on rhizobia that are the bacterial partners in the biologically nitrogen fixing symbiosis between leguminous plants and bacteria. In agricultural soil, the amount of biologically fixed nitrogen is estimated to be 60 Tg N (Herridge et al., 2008). The symbiosis of leguminous plants and rhizobia is responsible for most of the biological nitrogen fixation in agricultural systems (Herridge et al., 2008). The other major source of soil nitrogen, industrial nitrogen fixation for fertilizers, contributes approximately 100 Tg N annually (Fowler et al., 2013).

Nitrogen fixation is energy consuming and results in carbon dioxide (CO\textsubscript{2}) emissions. The production of industrial nitrogen fertilizers results in the release of 300 Tg CO\textsubscript{2} from fossil fuel annually (Jensen et al., 2012). Biological nitrogen fixation by legumes releases even more CO\textsubscript{2} but, since the CO\textsubscript{2} released derives from photosynthesis, it does not make a net contribution to atmospheric CO\textsubscript{2} concentration (Jensen et al., 2012). In addition, emissions of N\textsubscript{2}O, a potent greenhouse gas, are smaller in legume soil than when using industrial N fertilizer (Jensen et al., 2012). Therefore encouraging the use of legumes serves a means to mitigate green-house gas emissions.

With the exception of Parasponia, all the plant species that form nitrogen fixing symbiosis with rhizobia belong to the family Leguminosae (Sprent 2007). The Leguminosae species, legumes, are divided into three subfamilies, Caesalpinioideae, Mimosoideae and Papilionoideae. Symbiotic species are most common in the subfamily Papilionoideae, including the major temperate forage and grain legumes, e.g. herbaceous Medicago spp., symbionts of which were studied in Paper V. The species in the subfamily Mimosoideae are mostly tropical and subtropical trees and shrubs, including the genera Acacia, Prosopis and Leucaena, symbionts of which were studied in Papers I-V.

Leguminous trees are valuable for erosion control and water conservation (Allen & Allen 1976). In an agroforestry system, the nitrogen added by the rhizobium-tree symbiosis ranges from tens to hundreds of kilograms per hectare per year (Nygren et al., 2012). Moreover, in agroforestry the trees shield other crops from sun and wind (Le Houérou, 1989). In addition to the nitrogen, leguminous trees provide fodder, food, fuel and to a lesser extent timber (Fagg & Stewart, 1994; Allen & Allen, 1976). In Brazil, leguminous tree species have been successfully applied in reclamation of degraded land (Chaer et al., 2011).
1.2 The host plants in this study

*Acacia senegal*, *Prosopis chilensis* and *Leucaena leucocephala* are leguminous tree species in the subfamily Mimosoideae. *Medicago* spp. are annual and perennial herbs in the subfamily Papilionoideae. All these species form symbiosis with rhizobial bacteria.

*A. senegal* is one of the dominant trees of the Sahelian proper biogeographical zone (Le Houérou, 1989). *A. senegal* has two common names, gum acacia and gum arabic tree. It is a shrub like tree, usually no more than 5 m in height. *A. senegal* sheds its leaves during the dry season. The resin of the tree, arabic gum, is of commercial value (Allen & Allen, 1976). *A. senegal* is moderately salt tolerant (Rehman *et al.*, 2000).

Evergreen *P. chilensis*, known as algarroba or mesquite, is usually shrub-like. *Prosopis* species are originally from South America, and several of them have been introduced to Sudan. Most of the Sudanese prosopis are *P. juliflora* that was introduced to Sudan in the early 20th century (Laxen 2007). According to Laxen (2007), mistaking the Sudanese prosopis as *P. chilensis* is common. As there is no means to ascertain a correct identification, *P. chilensis* is assumed to be the original host of *S. arboris* HAMBI 1552 (Zhang *et al.*, 1991) studied in Papers I, IV and V. *Prosopis* spp. are commonly utilized in campaigns against desertification since they grow rapidly and spread easily. Because of the rapid spread, the species are considered as invasive weeds (Laxen 2007). *P. chilensis* is highly salt tolerant (Cazebonne *et al.*, 1999).

*Leucaena leucocephala* originates from El Salvador, Guatemala, Honduras and southern Mexico in Central America (Allen & Allen, 1976). The introduction of *L. leucocephala* to non-native areas started at pre-Colombian era. Currently it is cultivated or naturalized in the tropics and subtropics in altitudes up to 1500 m. It tolerates occasional frost and drought, but grows best in areas with 25 to 30 °C, 600 mm to 2500 mm annual rainfall and two to six months dry period. *L. leucocephala* tolerates light salinity. *L. leucocephala* was introduced to Sichuan from Guangdong in the 1980’s and 1990’s. The exact origin of the trees is unclear. *L. leucocephala* had been introduced in 1961 to Hainan Island, southern China, from where they were spread to continental Guangdong in 1976. In the 1970’s, Salvador-type *L. leucocephala* varieties were introduced from Philippines and Hawaii to Guangdong (Xu & Wang, 1994).

*Medicago* spp. are native to Eurasia and Africa, especially around the Mediterranean Sea, and widespread in temperate areas. As the *Medicago* spp. have a deeply penetrating root system, they are drought resistant. Due to the high yield and adaptation to various climatic and soil conditions, *M. sativa* (alfalfa), most likely originating from Iran, is an important forage legume in the arid and semi-arid areas that are susceptible to salinization (Djilianov *et al.*, 2003).
1.3 Rhizobia

The bacterial partners in the legume-bacterium symbiosis are collectively called as rhizobia, in singular rhizobium. Rhizobia may live as symbionts inside plants or as saproxytes in soil. Even though rhizobia are usually considered as symbiotic organisms, the number of nonsymbiotic rhizobia in soil seems to be larger than the number of symbiotic ones (Segovia et al., 1991).

The first discovered symbiotically nitrogen-fixing rhizobium species was named as *Rhizobium leguminosarum*, with the genus name derived from the Greek words rhiza and bios, meaning root and life, respectively. Over the years new rhizobial species were found and assigned to the genus. Hand in hand with the development of molecular methods to characterize the genetic relatedness of bacteria rhizobia were split into several genera. Currently, the term rhizobia refers to the characteristic capability of producing nitrogen fixing nodules on the roots or stems of leguminous plants and *Parasponia*, and does not bear any genetic information, since rhizobial species are found in several genera belonging to *Alphaproteobacteria* and *Betaproteobacteria* (Table 1). Most of the rhizobial species described so far belong to the genera *Bradyrhizobium*, *Ensifer* (formerly called as *Sinorhizobium*), *Mesorhizobium* and *Rhizobium*. To date, betaproteobacterial rhizobia have been found to nodulate species of tribe Mimoseae (Barrett & Parker, 2005; Elliott et al., 2007).

**Table 1. Currently recognized rhizobial genera and their respective nodulation types.**

<table>
<thead>
<tr>
<th>Class</th>
<th>Genus</th>
<th>Nodulation type</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Alphaproteobacteria</em></td>
<td><em>Aminobacter</em></td>
<td>root</td>
</tr>
<tr>
<td></td>
<td><em>Azorhizobium</em></td>
<td>root, stem</td>
</tr>
<tr>
<td></td>
<td><em>Bradyrhizobium</em></td>
<td>root, stem</td>
</tr>
<tr>
<td></td>
<td><em>Devisia</em></td>
<td>root</td>
</tr>
<tr>
<td></td>
<td><em>Ensifer</em> (Sinorhizobium)</td>
<td>root, stem</td>
</tr>
<tr>
<td></td>
<td><em>Mesorhizobium</em></td>
<td>root</td>
</tr>
<tr>
<td></td>
<td><em>Methylbacterium</em></td>
<td>root</td>
</tr>
<tr>
<td></td>
<td><em>Microvirga</em></td>
<td>root</td>
</tr>
<tr>
<td></td>
<td><em>Neorhizobium</em>¹</td>
<td>root</td>
</tr>
<tr>
<td></td>
<td><em>Ochrobactrum</em></td>
<td>root</td>
</tr>
<tr>
<td></td>
<td><em>Phyllobacterium</em></td>
<td>root</td>
</tr>
<tr>
<td></td>
<td><em>Rhizobium</em></td>
<td>root</td>
</tr>
<tr>
<td></td>
<td><em>Shinella</em></td>
<td>root</td>
</tr>
<tr>
<td><em>Betaproteobacteria</em></td>
<td><em>Burkholderia</em></td>
<td>root</td>
</tr>
<tr>
<td></td>
<td><em>Cupriavidus</em></td>
<td>root</td>
</tr>
</tbody>
</table>

Data from [http://edzna.ccg.unam.mx/rhizobial-taxonomy/node/4](http://edzna.ccg.unam.mx/rhizobial-taxonomy/node/4),

¹Mousavi et al. (2014).
1.4 Nodulation

In symbiosis, the plant and the rhizobia form a specified organ called nodule. The nodules are commonly on the root, as is the case for the rhizobia studied in Papers I-V, or on the stem of the plant. Inside the nodule, rhizobia convert atmospheric nitrogen to ammonium that serves as a nitrogen source for the plant. In bulk soil, the number of rhizobial cells is usually lower than $1 \times 10^6$ per gram soil (Hirsch, 1996). Rhizobia are present in soils with no prior cultivation of the host plant, yet their numbers increase if the host is present (Cardoso et al., 2012; Hirsch, 1996).

In root nodulation, the symbiosis starts from the attachment of rhizobia to the plant root. Most commonly, rhizobia enter the root via an infection thread. According to a model proposed by van Brussel et al. (1992), rhizobia induce the cells of the inner root cortex to divide, while the cells of the outer cortex are induced to form infection threads. The inner root cortex cells give rise to the nodule. After the infection thread penetrates the nodule, rhizobia are internalized in plant cells and surrounded by plasma membrane made by the plant. In most of the legumes, the nodules are indeterminate, i.e. the cells in the meristem of the nodule continue to divide through the whole lifetime of the nodule (Sprent 2007). In determinate nodules, the meristem activity ceases soon after the nodule formation.

In the nodules rhizobia differentiate into nitrogen fixing forms called bacteroids. The bacteroids in indeterminate nodules are not able to divide, but nodules may harbour also undifferentiated bacteria that are released to the soil after nodule senescence (Masson-Boivin et al., 2009), partly explaining the increase in rhizobial numbers when host plant is present. The symbiosis between the betarhizobium Cupriavidus taiwanensis and Mimosa pudica provides an exception to the rule: even though the nodules are indeterminate, the bacteroids are not terminally differentiated (Marchetti et al., 2011). The bacteroid differentiation is claimed to be under plant control: when a strain nodulating a plant with indeterminate nodules was modified to nodulate a plant with determinate nodules, the bacteroids in determinate nodules maintained the ability to divide (Mergaert et al., 2006). Likewise, when a strain nodulating a plant with determinate nodules was modified to nodulate a plant with indeterminate nodules, the bacteroids were not able to divide (Mergaert et al., 2006).

1.5 Symbiont and host specificity

In symbioses each rhizobial strain nodulates one or more particular plant species. Depending on the number of possible hosts rhizobia may be divided to strains having narrow or broad host-range. For example, Neorhizobium galegae HAMBI 1207 nodulates only Galega officinalis, while Rhizobium sp. NGR 234 nodulates over 110 different plant genera (Lindström, 1989; Fellay et al., 1998). Likewise, the plants may be nodulated with one or more particular rhizobium: Galega officinalis is nodulated only by Neorhizobium galegae symbiovar officinalis (Lipsanen & Lindström, 1988; Radeva et al., 2001), whereas Phaseolus vulgaris, the common bean, is nodulated by at least 16 rhizobial species that represent five rhizobial genera (Martinez-Romero, 2003). The term symbiovar (sv.) is used to define rhizobial strains that have a similar
host range (Rogel et al., 2011). Within a symbiovar the strains carry similar symbiotic genes, irrespective of the species assignment.

The specificity of symbiosis is governed by the molecules involved in the interaction between the partners. Phenolic molecules, mainly flavonoids, secreted by the plant roots increase the binding of a rhizobial regulator protein NodD to a specific region in rhizobial DNA called the nod-box (Figure 1) (Broughton et al., 2000, Peck et al., 2006). The NodD binding results in the transcription of the nodulation (nod) genes. Different variants of the NodD respond to different flavonoids. The composition of flavonoids secreted differs between plant species, making the flavonoid-NodD compatibility the first stage in symbiotic specificity (Broughton et al., 2000). The second stage in the host specificity is the interaction of plant receptors with the rhizobial signalling molecules synthesized and transported by the proteins coded by the nodulation genes (Figure 1). The rhizobial signalling molecules, known as LCOs (lipochitooligosaccharide signalling molecules) or Nod (nodulation) factors, have an amino sugar backbone consisting of N-acetylglucosamine residues (GlcNAc) (Figure 2). In some LCOs one GlcNAc is replaced with a glucose molecule (Bec-Ferté et al., 1996). All the LCOs carry an acyl group on the non-reducing terminal GlcNAc residue. Depending on the rhizobium, LCOs are acylated either with common fatty acids, major constituents of the membrane lipids, or with specific, polyunsaturated fatty acids (Dénarié et al., 1996). The other LCO decorations include acetyl, carbamoyl, glycosyl, methyl and sulphate groups (Dénarié et al., 1996). The decorations are usually on the terminal residues of the LCO backbone (Dénarié & et al., 1996; Olsthoorn et al., 1998; Yang et al., 1999). Due to their vital role in symbiosis, characterization of LCOs is essential to understand symbiosis.
Figure 2. Schematic representation of the most common LCO structural variants. n: 0 – 4. R^1: fatty acyl. The other R-groups are either hydrogen (H) or R^2: methyl; R^3-5: carbamoyl or acetyl; R^6: glycosyl, methyl or sulphate. R^7 is either on C6 as in this figure or on C3.

The plant receptors that perceive the LCO signal are LysM type receptor kinases. The LCO binding activity of the *Medicago truncatula* receptors depends on the length of the LCO (Fliegman *et al.*, 2013; Radutoiu *et al.*, 2007). There is only indirect evidence on the role of the other LCO characteristics in symbiosis. Through mutation, complementation and transfer studies the nodulation genes have been shown to affect the host range of rhizobia (Djordjevic *et al.*, 1983; Debelle *et al.*, 1986, 1988). When the receptor genes were transferred from *Lotus japonicus* to *Medicago* species, the *Lotus* nodulating rhizobial strains induced the formation of infection threads on *Medicago* (Radutoiu *et al.*, 2007). Even before the discovery of LCOs, transferring the nodulation genes *nod*FE, *nod*G, and *nod*H from *R. meliloti* was shown to extend the host range of *R. trifolii* to alfalfa (Debelle *et al.*, 1988).

The two first stages of symbiotic specificity are governed by diffusible molecules, whereas the third stage requires cell to cell contact. The binding of rhizobial surface polysaccharides to plant lectins, carbohydrate binding proteins, seems to act in concert with the first two host specificity mechanisms. The host ranges of rhizobia can be extended with transformation of structural genes for rhizobial surface polysaccharides or plant lectins, but only among relatively closely related species (De Hoff *et al.*, 2009, Simsek *et al.*, 2007).

Leguminous tree and shrub species in the subfamily Mimosoideae are nodulated both by *Alphaproteobacteria* and *Betaproteobacteria* (Bala & Giller, 2006; Leary *et al.*, 2006; Barrett & Parker, 2006). When looking at the specificity of symbiosis among the leguminous trees, it is noteworthy that many of the tree species are nodulated by several rhizobial species (Bala & Giller, 2006; Leary *et al.*, 2006; Barrett & Parker, 2006). As noticed earlier (Bala *et al.*, 2003; Bala and Giller, 2006), at different sites the species distribution of nodulating strains is different, and that differences in soil pH are associated with the species distribution. This is not to say that there is lack of host specificity. Since the symbiotic genes that are involved in the host specificity may be transferred between species, the phylogeny and symbiotic compatibility of rhizobia do not need to correlate (Laguerre *et al.*, 2001). For example, despite originating from geographically distant sources and representing different rhizobial species, the rhizobia nodulating *Sesbania* produce similar LCOs (Lorquin *et al.*, 1997a; Mergaert *et al.*, 1997; Promé *et al.*, 2002). In contrast to this, the LCOs of the *Sesbania* nodulating *S. terangae* sv. sesbaniae and the *Acacia*-nodulating *S. terangae*
sv. acaciae are different despite the genetic and geographical closeness of the strains (Lorquin et al., 1997a; 1997b).

## 1.6 Nodulation in the field

When growing legumes in the field, inoculating the plants with compatible rhizobia usually results in better growth. The soil may not contain rhizobial strains that nodulate selective legumes. The promiscuous legumes are easily nodulated, yet most of the nodulating strains may be ineffective in nitrogen fixation.

Even though the host legume-rhizobium interaction displays a certain level of host specificity, promiscuous legumes are nodulated even in sites with no prior history of host plant cultivation (Cardoso et al., 2012). Since rhizobia are able to survive on legume seeds (Perez-Ramirez et al., 1998), it has been argued that the nodulating strains have been introduced together with the legumes (Rodríguez-Echeverría et al., 2009; 2011; 2012). The possibly ubiquitous distribution of bacteria is another way to explain the presence of the nodulating strains. If the first paradigm of microbial biogeography “Everything is everywhere but the environment selects” by Baas Becking (1934, from De Wit & Bouvier, 2006) defines the actual distribution of bacteria, the bacterial communities in spatially separated but otherwise similar habitats should be similar. Therefore, legumes in similar habitats should be nodulated by similar rhizobia, whether or not the plants grew there before. A third possible explanation for nodulation of an introduced legume is in the degree of strictness of the requirements the legume poses for the rhizobia. The more promiscuous the host, the more likely it is nodulated in any given habitat.

As expected, the environmental conditions affect nodulation. In general, plants are more sensitive to adverse conditions than rhizobia (Zahran & Sprent, 1986). McKay & Djordjevic (1993) noticed that both LCO production and excretion are affected by the same factors that affect nodulation in field conditions, i.e. by suboptimal phosphate level, low or high pH and temperature. The soils in arid and semi-arid areas are prone to salinization. One tenth of the world’s land area suffers from salinization (FAO & IIASA, 2012). The success of reclaiming saline sites using legumes increases if plants are inoculated with efficiently nitrogen-fixing rhizobial strains (Thrall et al., 2005). Salinity disturbs or even inhibits nodulation and decreases nitrogen fixation and plant growth (Duzan et al., 2004). Especially the initial phases of symbiosis during which LCOs play a vital role is sensitive to stress conditions (Zahran & Sprent, 1986). The effect of salinity on LCO production by different strains varies: 50 mM NaCl salt stress had only a minor effect on LCO production by Sinorhizobium fredii SMH12 that nodulates moderately salt tolerant Glycine max (soybean) (Dardanelli et al., 2010). The production of LCOs by symbionts of salt sensitive Phaseolus vulgaris (common bean) Rhizobium etli ISP42 decreased in 50 mM NaCl; interestingly, that of R. tropici CIAT899 increased (Dardanelli et al., 2008).

The Senegalese and Tunisian Sinorhizobium spp. isolated from A. senegal and A. tortilis produced similar LCOs (Lorquin et al., 1997b; Ba et al., 2002). Since the host range of the Senegalese and Tunisian strains was almost identical to the Sudanese Acacia and Prosopis isolates with the exception that the first ones nodulated L.
leucocephala (Zhang et al., 1991; Lortet et al., 1996; Ba et al., 2002), analysis of the LCOs from the Sudanese strains should tell more about the role of LCOs in host specificity. The L. leucocephala nodulating strains showed variation in nitrogen fixation efficiency and the species distribution was different in different locations (Bala et al., 2003; Bala and Giller, 2001; 2006), making it improbable to find a universal inoculant for the tree. Earlier work on the effect of salinity on LCO production concentrated on strains nodulating herbaceous legumes that were salt sensitive or moderately salt tolerant. Because of the differences regarding the effect of salinity on LCO production by different strains, it was impossible to estimate how a symbiont of a salt tolerant leguminous tree would respond to salt stress. How various stress factors affected gene expression of the model rhizobium S. meliloti Rm1021 at a genome level had been studied (Giuntini et al., 2005; Krol et al., 2004; Rüberg et al., 2003), yet again there were no studies on the genetic basis for stress tolerance in wild-type inoculant strains.
2 AIMS AND OUTLINE OF THE STUDY

The general aim of this study was to gain knowledge on the features of potential inoculant rhizobia strains. The specific aims were (i) to assess host specific characteristics and the effect of salt stress on rhizobial signalling molecules, LCOs, that are needed for successful symbiosis; (ii) to find potential inoculant strains for a promiscuous legume tree; and (iii) to characterize potential inoculant strains isolated from promiscuous legume trees and forage legumes to assess genetic features behind the characteristics of the strains.

To reach the aims, LCOs of Sudanese *Sinorhizobium arboris* HAMBI 2361 and *S. kostiense* HAMBI 2362 nodulating *Prosopis chilensis* and *Acacia senegal*, respectively, were analyzed with mass spectrometry (*Papers I, IV*); rhizobial strains were isolated from *Leucaena leucocephala* growing in China, and characterized with phenotypic and genotypic methods (*Papers II, III*); and the genetic features behind the desirable characteristics of potential inoculant strains isolated from *Medicago minima*, *M. officinalis* and *Prosopis chilensis* from Algeria, Kazakhstan and Sudan, respectively, were analyzed by comparative genomic hybridization (CGH) using a model rhizobium *S. meliloti* Rm1021 microarray (*Paper V*).
3 MATERIALS AND METHODS

Many studies on plant associated bacterial populations follow a common path. First, strains are isolated from plants or rhizosphere, purified and tested for characteristic properties. Second, the collected strains are characterized by relatively simple methods that allow their grouping. The characterization is based on the phenotypic and genotypic properties of the strains. When the goal is to obtain preliminary phylogenetic information, reference strains representing known species are included. In addition, the quick grouping methods offer a view on the diversity of the isolates, an often neglected outcome of these analyses. Further on, representative strains or those with desirable properties from the groups are chosen for more detailed characterization at metabolite, protein, RNA or gene level, with sequencing of genes being the most widely used option.

3.1 Isolation and nitrogen fixation efficiency testing

Rhizobia are commonly isolated from nodules, after which the ability to nodulate the original plant host is assessed (Paper II). The nodules originate either from plants growing in the wild or from so-called trap hosts, i.e. plants that have been grown in soil collected from the wild. The nodulation tests for rhizobia may include tests on the capability to nodulate other legumes, as was done by Zhang et al. (1991) for the strains studied in Papers I, IV and V. Assessing the nitrogen fixation efficiency provides valuable information on the strains. Surface sterilized host plant seeds are germinated and planted, after which they are inoculated with the rhizobia. By harvest time the nodule numbers and plant dry weights are measured, as was done in Paper II. The dry weights of the inoculated plants are compared to those of not inoculated plants. Other growth characteristics, e.g. the heights of the plants may be measured, as was done by Räsänen et al. (2001) for the Acacia and Prosopis isolates studied in Papers I, IV and V.

3.2 Grouping by phenotypic and genotypic methods

The phenotypic analyses include testing the ability to grow on different carbon and nitrogen sources, tolerance to salinity, antibiotic resistance and defining the pH and temperature ranges for growth (Paper III). It has been argued that phenotypic test are of limited or no value at all in taxonomy (Ormeño-Orrillo & Martinez-Romero, 2013), i.e. when the goal is to select representative strains for phylogenetic classification the application of phenotypic tests in grouping the strains is questionable. However, phenotypic tests combined with numerical analysis also provide an indirect estimate of the genetic diversity within the isolates (Paper III). In evaluating the potential of rhizobial strains to be applied as inoculants, assessing the tolerance to stress conditions is desirable (Paper IV).

One of the most common genotypic methods for grouping the strains is restriction fragment length polymorphism (RFLP), where DNA is isolated from the strains and a target fragment is amplified by polymerase chain reaction (PCR). The amplified fragments are restricted to smaller fragments enzymatically which are separated in agarose gel thus creating a unique pattern for strains. The most common target is the
gene coding for the small subunit of the bacterial ribosome, the 16S rRNA gene (Wang et al., 1999). The gene coding the large subunit of the ribosome, the 23S rRNA gene, has more phylogenetic information content and gives better resolution than the 16S rRNA gene. However, the 16S rRNA gene sequence databases are more comprehensive than those of 23S rRNA gene sequences. The non-coding DNA sequence between the 16S and 23S rRNA genes, the 16S–23S rDNA intergenic spacer (IGS), is a good target for investigating the relatedness of closely related rhizobial strains. For some strains IGS PCR results in multiple PCR products (Laguerre et al., 1996), partly explained by insertion of tRNA genes in IGS regions (East et al., 1992). When IGS-PCR results in multiple products of varying size IGS PCR-RFLP is not usually carried out in taxonomic studies (Terefework et al., 1998). However, when the goal is to estimate the diversity of the isolates (Paper III), multiple IGS-PCR products are no hindrance. The three above mentioned PCR-RFLP methods were applied in Papers II and III to analyze rhizobial strains isolated from *Leucaena leucocephala*.

### 3.3 Detailed characterization of LCOs

One of the key characteristics of symbiotic nitrogen fixation is the chemical communication between the partners prior to and during the nodule formation. The analysis of the rhizobial lipochitooligosaccharidic signaling molecules (LCOs) is of uttermost importance for understanding rhizobia. Mass spectrometry (MS) has proven to be an excellent method for studying heterogenous biological samples, such as LCOs. Analytes in solution are delivered to the mass spectrometer either by direct infusion (Paper I) or on interfacing liquid chromatography (LC) on-line (Papers I, IV). Prior the mass spectrometer, the analytes are brought to gas phase and ionized. Closely related structural variants may be detected and identified according to their mass per charge (m/z) values. In tandem-MS, ionized analytes are fragmented using collision induced dissociation, and the structure of the precursor ion is deduced from the m/z values of the generated fragments (Figure 3).

More specifically, the structures of *S. arboris* strain HAMBI 2361 (Papers I, IV) and *S. kostiense* strain HAMBI 2362 (Paper I) LCOs were analyzed with electrospray - quadrupole orthogonal -time of flight mass spectrometry (ES-oQ-TOF-MS) and liquid chromatography-electrospray-ion trap-MS (LC-ES-ion trap-MS). In the off-line LC approach, extracted LCOs were fractionated by high performance liquid chromatography (HPLC). Pooled HPLC fractions were then analyzed using ES-oQ-TOF-MS. In the on-line LC approach, LCO extracts from separate fermentations were analyzed using a reversed phase microbore HPLC system coupled on-line with an ES-ion trap mass spectrometer.

More detailed information on the location of substituents on the non-reducing terminal N-acetyl glucosamine (GlcNAc) residue may be achieved with a specific extension of the tandem-MS approach (Treilhou et al., 2000). The fragmentation patterns and product ion intensity ratios depend on the location of substituents on the residue, and on the presence or absence of an N-methyl group. This approach was applied in Paper I to analyze *S. arboris* and *S. kostiense* LCOs.
Determining the full structure of an LCO requires the use of chemical pretreatment methods in combination with MS. Composition of the oligosaccharide backbone, the types of linkages between backbone monosaccharide residues and the nature of the fatty acyl substituents need to be determined by gas chromatography-MS of chemically derivatized samples, as was done in Paper I.

Quantitating LCOs poses a problem to the analyst. In HPLC with ultraviolet detector, the peak areas in the chromatograms give some indication on the LCO quantity. However, molecules with different structures have different extinction coefficients, i.e. the amount of light absorbed depends on the structure. For reliable quantitation one should have a standard for each of the LCO variants, a requirement practically unfeasible. Thin layer chromatography (TLC) combined with radioactive $^{14}$C-labelling of the LCOs results in more precise estimates. One needs to keep in mind that the longer the LCO backbone is, the higher the spot intensity is as compared to LCOs in equal quantity but with shorter backbones. Both these chromatographic methods suffer from low resolution. For example in the analysis of R. tropici CIAT899 LCOs produced under salt stress, TLC separation resulted in five distinct spots whereas MS analysis revealed 46 LCO variants (Estevez et al., 2009).

To overcome the difficulty of linking the observed quantitative changes to defined LCO structural variants, I devised a method for relative quantification of LCOs (Paper IV). In the method, rhizobium is grown in a defined medium with stable isotopes of nitrogen, either $^{14}$N or $^{15}$N, as the nitrogen source. Combining the
3.4 Comparative genomic hybridization

Comparative genomic hybridization (CGH) is a method to assess differences in genomic composition between the strain under study and a reference strain. In CGH, DNA from the reference strain and DNA from the analyzed strain are differentially labelled with fluorescent dyes and hybridized to a microarray. On the array there are probe spots with oligonucleotides or short PCR fragments corresponding to the genes of the reference strain. Fluorescence from each spot for both dyes is measured and the ratio of the fluorescence intensities from analyzed DNA versus the reference DNA is calculated to estimate the genomic composition of the analyzed DNA. For statistical analyses, the ratio may be log-transformed or calculated as fold change. When the fluorescence intensity for a particular gene is higher in the analyzed DNA than in the reference DNA, it is considered to indicate multiplication of that gene. When the intensity is lower for the analyzed DNA than for the reference DNA, it indicates that the particular gene is absent from the analyzed DNA or that the sequence of gene is divergent from the reference. CGH using the model rhizobium *Sinorhizobium meliloti* Rm1021 microarray was applied to search for genetic determinants of stress tolerance of three potential inoculant strains (Paper V).

CGH provides information on the genomic differences between closely related strains (Murray *et al.*, 2001). The main limitation of CGH is the dependence on the reference: genes absent from the reference strain cannot be detected. In addition, due to the ever decreasing prices of sequencing the prime time of CGH is seemingly in the past. However, the results obtained with CGH are still valid.

3.5 Gene sequence analysis

Sequencing the 16S rRNA gene of the isolates and comparing the sequence to those found in databases provides a good basis for assigning the isolates to defined species. The 16S rRNA gene is relatively conserved, thus limiting its applicability in species assignment. To overcome this, sequencing of protein coding genes have been included in species assignments during the last ten years (Martens *et al.*, 2007, 2008). The current gold standard for rhizobial phylogeny and species assignment is the multilocus sequence analysis (MLSA), where at least three protein encoding genes are
sequenced, usually partially, to create a concatenated sequence that carries more information than a single gene sequence. 16S rRNA gene sequencing and MLSA were applied in Papers II and III to analyze rhizobial strains isolated from *Leucaena leucocephala*.

In principle, comparing the concatenated sequence to sequences retrieved from databases allows a more precise species assignment. However, since the microbial diversity is still far from fully characterized, the amount of information in the databases is limited. Furthermore, an average nucleotide identity of 95%-96% is considered as the interspecies boundary (Martens et al., 2008; Richter & Rosselló-Móra, 2009). When the similarity of an isolate to described species is on that range, precise assignment is impossible, as was the case for the *Ensifer (Sinorhizobium)* isolates in Paper III.

Further insight into the rhizobia may be gained by analyzing the genes related to nodulation and nitrogen fixation characteristics. The nodulation gene *nodC* is unique and symbiosis-specific to rhizobia and essential for the synthesis of the LCOs (Debellé et al., 2001). The nitrogen fixation gene *nifH* is essential for nitrogen fixation by rhizobia inside the nodule but also for diazotrophs that fix nitrogen in the free-living state (Raymond et al., 2004). The nodulation and nitrogen fixation genes of rhizobia are often located on plasmids or genomic islands that may transfer between rhizobial species (Laguerre et al., 2001). Strains with different species assignment as revealed by MLSA may carry similar nodulation and nitrogen fixation genes, thus reflecting the gene flow between rhizobial species and the governing role of the plant host in selecting the nodule occupants from the soil population. Sequencing *nodC* and *nifH* were applied in Paper II to analyze rhizobial strains isolated from *L. leucocephala*. 
4 RESULTS AND DISCUSSION

4.1 The specificity of nodulation

When the goal is to grow promiscuous legumes for e.g. sustainable agriculture and as a means to fight erosion, the multitude of nodulating strains poses a problem for plant growers. The problem is due to the variation in the efficiency of symbiotic nitrogen fixation between the strains (Mnasri et al., 2009). The rhizobial isolates from A. senegal and P. chilensis in Sudan included both efficient and inefficient strains as determined by the acetylene reduction assay (Zhang et al., 1991). When analysed with the plant shoot dry weight assay, only 11 out of 41 rhizobial strains isolated from L. leucocephala in Panxi, China, were efficient in nitrogen fixation, making them good candidates to be applied as inoculants (Paper II). Surprisingly, ten isolates were considered as parasitic since they had a growth slowing effect on the host (Paper II), giving further motivation to search for effective inoculant strains. Analyzing the nodulating strains further aimed at providing insight on the characteristics of potential inoculants.

4.2 Phenotypic and genetic characteristics

The phenotypic differences between bacterial strains reflect their genetic differences. At least for rhizobia, phenotypic characterization does not result in phylogenetically meaningful grouping (Ormeño-Orrillo & Martínez-Romero, 2013). Only two out of forty Leucaena isolates had identical phenotypic profiles (Paper III). Numerical analysis based on the phenotypic analysis did not correlate with differences in nitrogen fixation efficiency. Ormeño-Orrillo and Martínez-Romero (2013) suggested that phenotypic diversity is a driver of speciation. The large diversity among the Leucaena isolates could imply a potential to occupy multiple niches, e.g. to nodulate several hosts.

For species assignment, the Leucaena isolates were grouped by RFLP, followed by choosing representative isolates and sequencing four genes from the selected ones (Papers II, III). Based on the RFLP analyses most of the isolates belonged to the genus Ensifer (Sinorhizobium) (32 isolates), and the remaining isolates belonged to the genera Rhizobium, Mesorhizobium, and Bradyrhizobium. Both the 16S rRNA gene (Paper II) and the concatenated gene (Paper III) sequence analyses assigned the representative isolates to the genera Ensifer (Sinorhizobium), Mesorhizobium, Rhizobium, and Bradyrhizobium.

The genus range was similar to those obtained from L. leucocephala elsewhere whereas the species distribution was different (Wang et al., 2006; Romdhane et al., 2006; Benata et al., 2008; Gehlot et al., 2012). As the nodulating species distribution is affected by soil conditions (Bala & Giller 2006; Liu et al., 2012), the differences in species distribution might result from the differences in soil properties in the sampling areas. This suggests that strains adapted to local soil conditions would be preferred when choosing inoculants. With the exception of bradyrhizobia, the species range for L. leucocephala isolates was similar to those from other promiscuous hosts A. senegal and P. chilensis (Zhang et al., 1991). Like the phenotypic analysis, IGS RFLP
indicated a relatively high degree of genetic variability among *L. leucocephala* isolates (Paper III). As in earlier studies on rhizobia isolated from various legumes (Cardoso et al., 2012; Mnasri et al., 2009), the genetic background of the isolates was diverse and not connected to their symbiotic performance.

In its site of origin, Mexico, *L. leucocephala* is most effectively nodulated by *Ensifer* (*Sinorhizobium*) strains (Bala & Giller, 2006). In Panxi, ten out of eleven effectively nitrogen fixing isolates belonged to the genus *Ensifer*. However, the two *Ensifer* RFLP groups included both effective and ineffective isolates. Li et al. (2012) proposed the division of nodulating strains into true and sporadic symbionts. Within a true symbiont species, almost all the strains should nodulate effectively; within a sporadic symbiont species, effectiveness and nodulation ability of the strains varies considerably. Interestingly, in Panxi *L. leucocephala* seems to be nodulated only by sporadic symbionts. Similar conclusion may be drawn from the symbiosis between *Acacia seyal* and mesorhizobia (Diouf et al., 2010). Should these trees be called as truly promiscuous hosts? Is this a characteristic feature of other promiscuous tropical leguminous trees? Irrespective of what the answers are, the variation in effectiveness between the isolates implied that the ability to compete with other strains is an important characteristic for the inoculant of a promiscuous host; the inoculant is of no use if nodules are mainly occupied by less efficient strains. If the previous studies on *L. leucocephala* symbionts had included nitrogen fixation efficiency analyses, it would have been possible to estimate whether the true symbionts are geographically limited, e.g. whether there were true symbionts among the isolates from the center of origin.

**4.3 Rhizobial signaling molecules**

The rhizobial signaling molecules, LCOs, are vital for successful symbiosis. In agreement with the similarity of the nodulation genes within a symbiovar, rhizobial strains nodulating the same tropical leguminous tree produce similar LCOs even if they represent different rhizobial species or originate from distant areas. Like the *Sesbania*-nodulating *Rhizobium* sp. mus10, *Azorhizobium caulinodans* and *Sinorhizobium terrangae* sv. sesbaniae from India and Senegal (Lorquin et al., 1997a; Mergaert et al., 1997; Promé et al., 2002), *S. arboris* HAMBI 2361 and *S. kostiense* HAMBI 2362 that share the same tropical trees as hosts produced similar LCOs (Paper I).

As for the other *Acacia*-nodulating rhizobia analyzed to date, the LCOs of *S. arboris* HAMBI 2361 and *S. kostiense* HAMBI 2362 were N-methylated, partially carbamoylated and partially sulfated, and acylated with common fatty acids (López-Lara et al., 1995; Lorquin et al., 1997b; Price et al., 1992; Bec-Ferté et al., 1994; Ferro et al., 2000; Ba et al., 2002, Paper I). Altogether *S. arboris* HAMBI 2361 and *S. kostiense* HAMBI 2362 produced 31 and 13 LCO variants, respectively (Papers I, IV). As the amount of detected LCO variants for *S. arboris* HAMBI 2361 in Paper IV was double to that of in Paper I, the difference between *S. arboris* HAMBI 2361 and *S. kostiense* HAMBI 2362 plausibly results from differences in analytical techniques.
Since the culture conditions can affect LCO production (Demont et al., 1993), it cannot be concluded that a certain rhizobium produces a certain number of LCO variants. The fatty acyl composition of LCOs acylated with common fatty acids reflect the fatty acid pool, which is known to vary in response to growth conditions and growth phase (Cedergren et al., 1995). It is evident that some variants are missed due to the detection limit of the analytical technique applied.

4.4 Nodulation genes

Looking at the signaling molecule linked characteristics on gene level using the comparative genomic hybridization revealed two major differences between the model rhizobium S. meliloti Rm1021 and S. arboris (Paper V). The Rm1021 microarray includes both structural nod genes and nodD genes which are involved in the activation of LCO synthesis (Krol et al., 2004; Luyten & Vanderleyden, 2000). In agreement with the different host ranges, the nodD2 that perceives the plant signals and regulates the expression of other nodulation genes was divergent in S. arboris 1552. In agreement with the structural differences of the S. meliloti and S. arboris LCOs, the nodulation genes nodL, connected to acetylation of the Nod factor non-reducing terminal, and nodE, linked to Nod factor acylation by polyunsaturated fatty acids, were divergent (Luyten & Vanderleyden, 2000; Papers I, V).

The L. leucocephala nodulating Ensifer (Sinorhizobium) strains representing at least three species shared common nodC, implying that their LCOs might be similar as well (Paper II). Similarly, no divergence was detected between the S. meliloti model strain Rm1021 and the wild type strains nodulation genes (Paper V). The common nodulation genes nodABC are linked to the biosynthesis of the basic structure of LCO. Dissimilarity of the nodABC does not necessarily result in differences in LCOs or host ranges. For example, even though S. arboris, S. kostiensc and Acacia tortilis nodulating Sinorhizobium strains produce almost identical set of LCOs, their nodA genes are considerably different (Paper I, Ba et al., 2002).

4.5 What is the role of all the LCO variants in symbiosis?

As noted by Estevez et al. (2009), defining the biological role of each LCO variant is beyond current analytical techniques. In the initiation of symbiosis, only a few of the LCO structural features have been confirmed to be essential (Roche et al., 1991; Stokkermans et al., 1995). Also, plant responses depend on the combination of LCOs present (Minami et al., 1996). When a certain structural variant is detected, it is plausible that intermediate structures thereof are present, e.g. because some of the LCOs are biosynthetically incomplete by the time of extraction or there is “leakage” from the LCO synthesizing enzymes. As an example of such a series of “complete” LCO and its intermediates are the S. arboris HAMBI 2361 LCO V(18:1, NMe, Cb, S) and LCOs V(18:1), V(18:1, NMe) and V(18:1, NMe, Cb) (Paper IV). It may be asked whether it is expected that each variant in such a series has a biological role. Recognizing the major LCOs might help in clarifying the issue. Before Paper IV, if an LCO was referred to as the major LCO, justifications were missing. As a first attempt, the major LCOs of S. arboris HAMBI 2361 LCOs were defined as LCOs that met the relative quantitation criteria in each of the nine analyses done for
evaluating the applicability of the quantitation method (Paper IV). Major LCOs might also be defined as the variants produced under each of the different growth conditions tested. It must be admitted that in neither one of these approaches the ionization efficiency issue can be eliminated.

When defined as LCOs that were produced under all tested growth conditions, the major LCOs of *R. tropici* CIAT899, a symbiont of *Phaseolus vulgaris* (common bean) and *L. leucocephala*, were N-methylated and partially sulphated tetra- and pentamers acylated with common fatty acids (Estevez et al., 2009). Concerning the major LCOs, the carbamoyl group on *S. arboris* LCOs was the only difference between the *S. arboris* and *R. tropici*, rhizobia with different host ranges (Paper IV). Furthermore, *A. senegal* is nodulated by rhizobia that do not produce carbamoylated LCOs (Ba et al., 2002). For root nodulation of *Sesbania rostrata* none of the LCO structural modifications was indispensable (D’Haeze et al. (2000). In an *in vitro* assay with purified LCOs, only one LCO variant, an N-methylated pentamer acylated with C18, was present in all of the biologically active and in none of the inactive fractions (Guasch-Vidal et al. (2013). All these findings are in line with Broughton et al. (2000) who stated that LCOs may be regarded as general necessary components for the initiation of symbiosis that have only a limited role in host specificity. For *Acacia* nodulating rhizobia, the role of the carbamoyl group is probably not in defining the host specificity of nodulation, but to protect the LCO against degradation or to increase nodulation efficiency, as proposed for acetyl groups in *S. meliloti* and *R. etli* LCOs (Staehelin et al., 1994; Corvera et al., 1999).

However, the host plant seems to pose limits on the complexity of the LCO pool produced by the nodulating rhizobia. The LCOs produced by rhizobia that nodulate tropical leguminous trees are partially sulfated or glycosylated or both. *Leucaena leucocephala* does not pose strict requirements: it is nodulated by strains producing sulfated, glycosylated or glycosylated and sulfated LCOs (Lorquin et al., 1997b; Poupot et al., 1993; 1995; Hernandez-Lucas et al., 1995; Folch-Mallol et al., 1996; Pueppke & Broughton, 1999; Snoeck et al., 2001). *Sesbania* nodulating rhizobia produce glycosylated LCOs (Lorquin et al., 1997a; Prome et al., 2002). *A. senegal* nodulating strains produce mainly carbamoylated and partially sulfated LCOs (Lorquin et al., 1997b; Ba et al., 2002; Paper I). The only difference between the *Acacia*-nodulating *S. terangae* sv. *acaciae* and the *Acacia*-nonnodulating *S. terangae* sv. *sesbaniae* LCOs is the glycosylation of sv. *sesbaniae* LCOs (Lorquin et al., 1997a; 1997b). For *A. senegal* nodulating rhizobia, the lower complexity limit for successful nodulation seems to be an N-methylated and common fatty acid acetylated LCO, whereas reducing terminal glycosylations result in unsuccessful nodulation. The complexity limits are a two way street. *A. senegal* seems to narrow the range of rhizobia that nodulate it, and those requirements effectively narrow the host range of the rhizobia nodulating *A. senegal*: rhizobia that produce LCOs suitable for *A. senegal* do not nodulate *Sesbania* (Ba et al., 2002).

### 4.6 Determinants of symbiotic performance

In searching strains for practical applications, e.g. potential inoculants, it is the response of the plant to the inoculant that is sought after. Thus, evaluating the desired
phenotypic characteristic might be considered to fulfill the requirements of application driven analyst. However, in most of the cases one needs to assess more than one characteristic to make the right choice. Besides having the nodulation and nitrogen fixation abilities the selected strain must be competitive. Competitiveness results from several characteristics. A good inoculant must tolerate environmental stress (Triplett, 1990), grow well and attach effectively to the host root to colonize it (Smit et al., 1992; Hartwig et al., 1991), and survive in the soil after nodule senescence (Murphy et al., 1995).

In Paper V, the genomes of three potential inoculant Sinorhizobium strains were compared to the genome of S. meliloti Rm1021 using comparative genomic hybridization. Rm1021 carries three nodD genes that code for proteins regulating the expression of other nodulation genes: NodD1 and NodD2 in Rm1021 are responsive to different compounds in plant root secretions whereas NodD3 is plant inducer independent (Luyten &Vanderleyden, 2000). Both wild-type S. meliloti strains AK23 and STM 1064 had two variable nodulation genes. The multiplication of nodD3 and LCO secretion linked nodI in AK23 and STM 1064, respectively, is possibly connected to the efficient nodulation capabilities of the strains.

The CGH analysis did not reveal any genetic determinants directly related to nitrogen fixation efficiency. It was also noted in Papers II and III that the genotypic and phenotypic characteristics of Leucaena isolates from from Panxi, China, did not distinguish efficient from the nonefficient and parasitic isolates. In symbiosis, nitrogen fixation might be considered as the most complex process requiring sophisticated cooperation between the partners. While assessing the efficiency requires only straight forward classical techniques, an understanding of the phenomena behind the efficiency differences is yet to be achieved.

4.7 Stress tolerance and response

The host plants of S. arboris and the wild-type S. meliloti strains grow in areas with salinization potential (Schofield & Kirkby, 2003). Since salinity affects nodulation severely (Fall et al., 2009), assessing the effects of salt on potential inoculant strains is important. The isolates from L. leucocephala tolerated up to 1% NaCl (171 mM) on YM agar medium (Paper III), a salt tolerance level equal to that of the host plant (Tomar & Gupta, 1985). The salt tolerant P. chilensis nodulating S. arboris HAMBI 2361 tolerated up to 750 mM NaCl (4.4%) (Paper IV). The S. meliloti strains isolated from Medicago spp. tolerated 600-700 mM NaCl (Paper V), a tolerance level superior to that of the host plants (Abdelly et al., 1995; Humphries & Auricht, 2001).

Depending on the rhizobial strain, mild salt stress causes either a decrease or increase in LCO production (Dardanelli et al., 2008; Estevez et al., 2009). Since the effect of salt stress on LCO production had been assessed for few strains, no general response could be expected. The LCO production by S. arboris HAMBI 2361 decreased under salt stress (Paper IV). Under salt stress S. arboris HAMBI 2361 produced thirteen LCO variants in detectable amounts (Paper IV). Six LCOs were produced in quantities that could be reliably quantitated. Even though S. arboris is stress tolerant, the LCO production decreased to approximately one tenth of the non-stressed ones. Is
this decrease enough to have any biological significance? Not necessarily, since the effect of NaCl stress on soybean root hair curling was the same at LCO concentrations ranging from $1 \times 10^{-6}$ to $1 \times 10^{-10}$ M (Duzan et al., 2004). However, the root exudates produced under salt stress do not induce the LCO production as effectively as the flavonoids applied in vitro (Dardanelli et al., 2008; 2010; 2012). Thus, in soil the joint effect of a decrease in both the inducing activity of the root exudates and the ability of *S. arboris* to produce LCOs probably results in a major decrease in LCO production, which could lead to a decrease in nodule numbers.

CGH analysis revealed some genetic characteristics that might be linked to the stress tolerance of *S. arboris* and *S. meliloti* strains (Paper V). Genomic rearrangement by transposition is related to adaptation to different conditions (Zhong et al., 2004). In *S. meliloti* Rm1021, out of the osmotic shock affected genes ten code for transposases (Domínguez-Ferreras et al., 2006). Eight and nine of these were divergent in *S. arboris* and *S. meliloti* strains, respectively (Paper V), suggesting that the osmotic shock response of the wild-type strains is considerably different to that of Rm1021.

COG J genes are linked to translation, ribosomal structure and biogenesis. Osmotic shock represses altogether 40 Rm1021 COG J genes, probably due to the less intense protein biosynthesis at the lower growth rate under stress (Domínguez-Ferreras et al., 2006). CGH revealed that 44 *S. arboris* HAMBI 1552 COG J genes, including twenty osmotic shock affected genes, were duplicated (Paper V), thus the stress tolerance of *S. arboris* HAMBI 1552 might be partially due to a more efficient protein biosynthesis.

Rhizobactin is involved in the ability of Rm1021 to survive in soil under iron deficiency (Lynch et al., 2001). Seven rhizobactin genes are repressed by salinity stress. All the ten rhizobactin operon genes of Rm1021 were divergent in the wild-type *S. meliloti* strains (Paper V), implying that the strains have alternative systems for iron uptake. Whether the divergence is connected to better growth and stress tolerance in soil remains to be elucidated.

### 4.8 Accessory gene diversity

The *S. meliloti* genes are on three replicons. In general, the housekeeping genes are on the chromosome, the symbiotic genes are on symbiosis plasmid A (pSymA) whereas symbiosis plasmid B (pSymB) carries the genes related to survival in soil (Galibert et al., 2001). For all the wild-type strains the relative proportion of divergent genes was highest on pSymA (Paper V) and in COG category L, both of which include more accessory than core genes (Galardini et al., 2011). The accessory genes include genes related to environmental adaptation. The phenotypic variation of rhizobia strains can be be explained by accessory genes (Galardini et al., 2011; Biondi et al., 2009). As shown in Paper III, phenotypically the *Leucaena* isolates were diverse, suggesting that the isolates had substantial differences in their accessory genes. However, based on the results of Papers III and V it is impossible to conclude if the divergence and diversity gave the strains any competitive advantage.
5 CONCLUSIONS

Even though rhizobia having a shared host range produced similar LCOs and carried similar nod genes, comparison of major LCOs between rhizobia with different host ranges suggested that LCOs have only a limited role in host specificity (Papers I, II, IV, V).

The variation in effectiveness between the L. leucocephala isolates implied that the ability to compete with other strains is an important characteristic for the inoculant of a promiscuous host (Paper II). Nodulation ability and stress tolerance are important in competitiveness. The genes related to the induction of LCO synthesis and LCO secretion were duplicated in wild-type S. meliloti strains, which is possibly connected to the efficient nodulation capabilities of the strains (Paper V). The stress tolerance of S. arboris HAMBI 1552 might be partially due to a more efficient protein biosynthesis. Despite the stress tolerance, the LCO production of a salt-stressed S. arboris HAMBI 2361 was approximately one tenth of the non-stressed one, partially explaining the decrease in nodulation under salt stress (Paper IV).

The species distribution of the L. leucocephala isolates suggested that strains adapted to local soil conditions would be preferred when choosing inoculants (Paper II). The Leucaena isolates were phenotypically diverse (Paper III) and the S. arboris and wild-type S. meliloti strains diverged from the model rhizobium S. meliloti Rm1021 mainly in the accessory genome implying that the divergence was important in shaping the adaptability of the strains (Paper V).
6 FUTURE PROSPECTS

Despite all the sophisticated analysis methods, traditional plant growth tests seem irreplaceable in screening effective inoculants. However, the modern methods are needed for better understanding of the legume-rhizobium symbiosis.

The roles of the structural characteristics of LCOs could be addressed with strains carrying mutations in the structural nod genes. Assessing the nodulation ability of the mutants and the plant responses of purified LCOs should give more needed knowledge on the significance of the characteristics. The study should include both qualitative and quantitative analysis of the LCOs produced by the mutants.

Whole genome sequencing of *L. leucocephala* isolates with different nitrogen fixation abilities might shed light on the genetic determinants of nitrogen fixation efficiency. For studying the competiveness and stress tolerance further, a whole genome analysis including both the potential inoculant strains analysed in this study and closely related symbiotically less efficient strains is needed.

The effect of salt stress on LCO production *in vivo* and how it affects the early stages of symbiosis and the formation of nodules could be assessed in a hydroponic assay combined with $^{14}$C-labelling of the LCOs.
7 REFERENCES


USA 109, 13859-13864


**Cardoso JD, Hungria M, Andrade DS** (2012) Polyphasic approach for the characterization of rhizobial symbionts effective in fixing N₂ with common bean (*Phaseolus vulgaris* L.). *Appl Microbiol Biotechnol* 93, 2035-2049


de Wit R, Bouvier T (2006) 'Everything is everywhere, but, the environment selects'; what did Baas Becking and Beijerinck really say? Environ Microbiol 8, 755-758


Estévez J, Soria-Díaz ME, de Córdoba FF, Moron B, Manyani H, Gil A, Thomas-


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. *Microbiol* 147, 981–993


Genetic organization of the region encoding regulation, biosynthesis, and transport of rhizobactin 1021, a siderophore produced by Sinorhizobium meliloti. J Bacteriol 183, 2576-2585


Mckay I & Djordjevic M (1993) Production and excretion of nod metabolites by Rhizobium leguminosarum bv. trifolii are disrupted by the same environmental-factors that reduce nodulation in the field. Appl Environ Microbiol 59, 3385-3392


Poupot R, Martínez-Romero E, Promé J (1993) Nodulation factors from Rhizobium tropici are sulfated or nonsulfated chitopentasaccharides containing an N-methyl-N-acylglucosaminyl terminus. Biochemistry (NY) 32, 10430-10435


Radutoiu S, Madsen LH, Madsen EB, Jurkiewicz A, Fukai E, Quistgaard EMH,


