Diversity and Phylogeny of Root Nodule Bacteria Isolated from Tree, Shrub and Food Legumes of Ethiopia

Aregu Amsalu Aserse

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Agriculture and Forestry of the University of Helsinki, for public examination in Raisio Oyj Research Foundation Hall (LS 2, A109) at B-building, Latokartanonkaari 7, on November 21st 2013, at 12 o’clock noon.

Helsinki 2013
Supervisors: Professor Kristina Lindström
Department of Environmental Sciences and
Department of Food and Environmental Sciences
University of Helsinki, Finland

Dr. Leena A. Räsänen
Department of Food and Environmental Sciences
University of Helsinki, Finland

Pre-examiners: Dr. Pablo Vinuesa
Programa de Ecologia Genómica
Centro de Ciencias Genómicas-UNAM
Avda. Universidad s/n, col. Chamilpa
Cuernavaca, Mexico

Dr. Julie Ardley
Centre for Rhizobium Studies
School of Veterinary and Life Sciences
Murdoch University, Australia

Opponent: Dr. Philippe de Lajudie
Laboratoire des Symbioses Tropicales et Méditerranéennes
IRD/INRA/CIRAD/UM2/Agro-M (UMR 113)
Campus International de Baillarguet
Montpellier, France

Custos: Professor Per Saris
Department of Food and Environmental Sciences
University of Helsinki, Finland

ISBN 978-952-10-9395-1 (PDF)
ISSN 1799-7372

Unigrafia
Helsinki, Finland

Front cover photos by Aregu Amsalu Aserse
# TABLE OF CONTENTS

ACKNOWLEDGMENTS ........................................................................................................ 5  
SUMMARY .......................................................................................................................... 8  
LISTS OF ORIGINAL PAPERS ........................................................................................... 10  
ABBREVIATIONS .............................................................................................................. 11  
1. INTRODUCTION .......................................................................................................... 12  
1.1. BNF: Implication on sustainable ecosystem services .............................................. 12  
1.2. The plant family Leguminosae ................................................................................. 13  
1.2.1. Woody and shrub legumes: An overview of use in the tropics and subtropics ...... 14  
1.2.2. Food legumes: Focus on cultivation and use in Ethiopian ................................. 16  
1.3. The symbiotic association of legumes with rhizobia .............................................. 17  
1.3.1. Definition of symbiosis ...................................................................................... 17  
1.3.2. Mechanisms of rhizobia-legume symbiotic interactions .................................. 18  
1.4. Nodule endophytic bacteria .................................................................................... 20  
1.4.1. General overview .............................................................................................. 20  
1.4.2. Plant growth promoting properties of nodule endophytic bacteria ................. 20  
1.5. Taxonomic overview of rhizobia .......................................................................... 21  
1.6. Taxonomic criteria used in classification of the rhizobia: An overview ............. 24  
1.7. Molecular methods used in rhizobial diversity and phylogenetic studies ............. 25  
1.8. Genetic methods and markers commonly used in rhizobial diversity and taxonomic studies ................................................................. 26  
1.8.1. 16S rRNA gene sequences .............................................................................. 26  
1.8.2. DNA-DNA hybridization ............................................................................... 27  
1.8.3. Housekeeping protein-coding gene sequences .............................................. 28  
1.8.4. Amplified fragment length polymorphism (AFLP) fingerprinting .................. 30  
1.8.5. Whole genome sequencing ............................................................................ 31  
1.8.6. Accessory genetic elements: Symbiotic genes ............................................. 32  
1.8.7. Genetic markers and methods used in this thesis ......................................... 34  
1.9. Rationale of the study ............................................................................................. 34  
2. OBJECTIVES OF THE STUDY .................................................................................... 35  
3. MATERIAL AND METHODS ....................................................................................... 36  
4. MAIN RESULTS AND DISCUSSIONS ....................................................................... 38
4.1. Bacterial identification

4.2. Phylogeny and diversity of the test bacterial strains: MLSA and AFLP-fingerprinting analyses

4.2.1. The *Rhizobium leguminosarum* complex: True symbionts of common bean in Ethiopia

4.2.2. Ethiopian soils harbor phylogenetically diverse groups of *Bradyrhizobium*.

4.2.3. Diverse non-symbiotic endophytic bacteria obtained from nodules of woody, shrub and food legumes

4.3. Plant growth promoting activities of the sporadic symbionts and endophytic bacteria

4.4. Phylogeny of symbiotic genes

4.4.1. Genes *nodC* and *nifH* of Ethiopian *Rhizobium* species

4.4.2. *Bradyrhizobium nodA* phylogeny: Implications of geographic origin or host legumes

4.4.3. Phylogenetic congruence of *Bradyrhizobium nodA*, *nodY/K* and *nifH* gene sequences: Inference for the monophyletic origin of the symbiotic genes

4.5. Comparative analysis of the phylogeny based on symbiotic and housekeeping genes

5. CONCLUSIONS

6. FUTURE PROSPECTS

7. REFERENCES
ACKNOWLEDGMENTS

This work was done at the Department of Food and Environmental Sciences, University of Helsinki. The work was financially supported by funds and grants from the Academy of Finland, University of Helsinki and the Ella and Georg Ehrnrooth Foundation. I am indebted to all funding agents and institutions for supporting my PhD studies and for providing brilliant working facilities. I am grateful to Professor Per Saris for facilitating the final stages of the dissertation process and supporting my grant applications during the last phase of the study.

I would like to express my warmest and deepest gratitude to my supervisors, Professor Kristina Lindström and Dr. Leena Räsänen, for providing wonderful guidance, inspiration, useful discussions, comments and criticisms all the way through my study. Certainly, without your overall help, my achievements would not have been possible. Stina, your trust in me and your inspiring and encouraging words during the good and bad times helped me to be successful. You are also appreciated for arranging my stipends during the course of my study. Leena, you are unforgettable for showing me the way to lab work and introducing me to molecular techniques and for your tremendous help in reading and editing my articles and for your valuable comments on the thesis.

I extend my special gratitude to Dr. Fassil Assefa and Dr. Asfaw Hailemariam, who were part of the project from the very beginning and coauthors of the articles included in this thesis. I thank you both for having confidence in me and nominating me as PhD student in the first place.

My sincere appreciations go to pre-examiners Dr. Pablo Vinuesa, programa de Ecología Genómica Avda. Universidad s/n, col. Chamilpa, Centro de Ciencias Genómicas-UNAM, Cuernavaca, Mexico and Dr. Julie Ardley, at Centre for Rhizobium Studies in Murdoch University, Australia for providing productive and essential comments for my thesis.

Many thanks go to my colleagues and friends: Petri Penttinen, Leena Suominen, Seyed Abdollah Mousavi, Leone Montonen, Aneta Dresler-Nurmi and Janina Österman as well as other former and present workers in the nitrogen fixation group for helping in the lab and for sharing ideas during our special group seminars and for the friendly guidance you provided me over the years. I also appreciate Leena Suominen for support in the search for grants and for all positive advice provided during the course of my study. Special thanks go to Petri for reading and commenting on my thesis. I would especially like to thank the past and present lab technicians at the Division of Microbiology and Biotechnology for providing consistent technical support.

It is a great pleasure to recognize my fellow Ethiopian colleagues and friends: Abayneh Essayas, Anteneh Argaw, Asnake Demena, Tiru Ejigu, Daniel Muleta, Ashenafi Ali, Genet Mogese and Aberash Assefa for providing professional and technical support during my study time at the National Soil Testing Center, Addid Ababa, Ethiopia. I express my appreciation to Degefie Tibebe from Ethiopian Agricultural Research Institute, Addis Ababa, for organizing meteorological data of the sampling sites.
I am delighted to thank my family in Ethiopia, parents, sisters, brothers and my in-laws for showing love and absolute support over my entire study period. No words can actually express how grateful I am to my beloved mother, Felegush Gedif and my late father, Amsalu Asressa. My father, I know you would be so proud to see my progress as a person and a professional. Unfortunately you are not here with us; though I believe you are somewhere out there always seeing us closely. I also would like to honor my wife, Yayneabeba Dubale. Ayneye, it was because of your love, patience, consistent support and encouragement that you provided me throughout all up and downs what made this dissertation possible. I thank you for everything you did for me. My six years old son Gedion deserves also special thanks for showing tolerance of the separation from me and his mother during my study time in Finland. Finally, my mother-in-law, Haregua Tiruneh and father-in-law, Dubale Worke, you both deserve special thanks for the incredible care you have been offering Gedion, when I and Ayneye have been staying in Finland.

Helsinki, November 2013
Areugu Amsalu Aserse
Dedicated to my Family
SUMMARY

Nitrogen is one of the major essential nutrients for plant growth along with phosphorus and potassium. Soil nutrient depletion, including the depletion of soil nitrogen (N) is a common phenomenon in the sub-Saharan countries including in Ethiopia. The nutrient depletion decreases the productivity of crops including food legumes such as soybean and common bean. Even though the soil fertility can be improved by application of chemical fertilizer, it is very expensive for small-holder farmers. Thus, a cheaper and environmentally friendly nitrogen source, such as biological nitrogen fixation (BNF), is needed to mitigate the fertility problems of the farmlands.

Some specialized bacterial and archaeal species are able to fix atmospheric N\textsubscript{2} into NH\textsubscript{3} and that is subsequently converted into plant usable form of nitrogen, NH\textsubscript{4}\textsuperscript{+} or NO\textsubscript{3}\textsuperscript{-}. The BNF process that occurs by the symbiotic interaction of leguminous plants and certain bacterial species (commonly known as rhizobia) is the main source of biological nitrogen input into the soil and therefore plays an important role in maintaining the sustainability of ecosystem services. Due to the fixed N they get from symbiosis, legume species grow better than other plants in nutrient poor, degraded soils. Thereby leguminous trees and shrubs restore degraded farmland and soil fertility by increasing the nitrogen and organic carbon contents in the soil. The versatile leguminous trees and shrubs, such as *Erythrina brucei*, *Crotalaria* spp., and *Indigofera* spp. can be used as forage for cattle and applied as intercrops or fallow crops in low-input agriculture. The usefulness of these legumes can be boosted by inoculating them with effective nitrogen-fixing rhizobia. The yield of food legumes, such as common bean and soybean can also partly be increased through the use of efficient rhizobial inoculants. Thus, detailed information about the indigenous rhizobia nodulating local food and woody legumes is essential for selecting good inoculant strains. Therefore, this thesis deals with diversity and phylogeny of root nodule bacteria isolated from *E. brucei*, *Crotalaria* spp., and *Indigofera* spp., common bean and soybean growing in different sites in Ethiopia. This study is an initial step in developing bio-fertilizers and utilizing rhizobial inoculants for increasing the productivity of food legumes and for better seedling establishment of woody legumes to rehabilitate degraded farmlands.

Overall 143 bacterial strains were obtained from root nodules of *Crotalaria* spp., *E. brucei*, *Indigofera* spp., common bean and soybean. The taxonomy of the root nodule bacteria was studied using multilocus sequence analyses (MLSA) of the core genes 16S rRNA, *recA*, *rpoB*, and *glnII*. Phylogeny of nodulation (*nodA*, *nodC*, *nodK/Y*) and nitrogen-fixation (*nifH*) genes of the rhizobia were also studied. The whole genome based AFLP fingerprinting technique was used to study the diversity of the strains within the species. Based on MLSA and AFLP fingerprinting analyses combined with nodulation test results, twenty-five strains belonging to the *Rhizobium leguminosarum* complex (*Rhizobium phaseoli*, *Rhizobium etli*, *Rhizobium leguminosarum* and a novel *Rhizobium* taxa) were found to be true symbionts of common bean (*Phaseolus vulgaris*) in Ethiopia (Paper I). Fifty-six strains isolated from root nodules of *E. brucei*, *Crotalaria* spp., and *Indigofera* spp. and soybean (*Glycine max*) were mainly identified as genetically very diverse slow-growing *Bradyrhizobium* species, being distributed into fifteen phylogenetic groups under *Bradyrhizobium japonicum* and *Bradyrhizobium elkanii* super clades. The majority of these strains represented undescribed
Bradyrhizobium genospecies. Two unique lineages which most likely represent novel Ethiopian Bradyrhizobium species were discovered among the collections (Paper II). In addition to the Bradyrhizobium species, a few Rhizobium species (six strains) were found to sporadically nodulate Indigofera spp., E. brucei and common bean. Fifty-six non-symbiotic, endophytic bacterial strains representing diverse Gram-negative and Gram-positive bacterial genera were also isolated from nodules of E. brucei, Crotalaria spp., and Indigofera spp., soybean and common bean (Paper I, III). Among the non-symbiotic bacteria, five strains obtained from the nodules of Crotalaria spp. and E. brucei represented a putative novel Rhizobium species (Paper III).

Phylogenetically the nodA genes of all Ethiopian Bradyrhizobium species belonged to the cosmopolitan nodA clade III.3, which includes nodA genes from Bradyrhizobium species nodulating diverse legume hosts in sub-Saharan Africa. The nifH and nodY/K gene phylogenies of the Ethiopian Bradyrhizobium strains were generally consistent with the nodA phylogeny, supporting the monophyletic origin of the symbiotic genes in Bradyrhizobium (Paper II). The symbiotic gene phylogenies of Bradyrhizobium species were partially consistent to their housekeeping gene phylogenies, indicating that symbiotic genes in Bradyrhizobium species might be maintained by both vertical and horizontal gene transfer. Nevertheless, the symbiotic gene phylogenies of different Rhizobium species (Paper I and III) were fairly similar regardless of their taxonomic background, suggesting that, in contrast to the core genome of the species, the symbiotic genes required for nodulation and nitrogen fixation might have a common origin in Rhizobium, indicative of horizontal gene transfer among these rhizobia.

The nodulation tests showed that most rhizobial species were effective in nitrogen fixation on their respective host plants. Non-nodulating, endophytic bacterial strains representing seven genera, namely Agrobacterium, Burkholderia, Paenibacillus, Pantoea, Pseudomonas, Rhizobium and Serratia, were found to colonize nodules of Crotalaria incana and common bean when co-inoculated with symbiotic rhizobia. In addition, the majority of nodule endophytic bacterial strains and the sporadic symbionts showed plant growth promoting activities, such as production of indole-3- acetic acid (IAA), synthesis of siderophore, phosphate solubilization and protease, lipase and cellulase enzymatic activities, which indicate their potential role in improving plant growth.
LIST OF ORIGINAL PAPERS

Paper I


Paper II


Paper III


The contribution of the author to the papers

**Paper I.** Aregu Amsalu conducted the experiments, analyzed the data, interpreted the results and wrote the manuscript under supervision of Kristina Lindström and Leena Räsänen. Aregu Amsalu was the corresponding author. Fassil Assefa and Asfaw Hailemariam provided expertise on nodule sampling in the field, bacterial isolation in the laboratory and greenhouse experiments.

**Papers II and III.** Aregu Amsalu conducted the laboratory experiments, analyzed the data, interpreted the results and wrote the manuscript under supervision of Kristina Lindström and Leena Räsänen. Aregu Amsalu was the corresponding author. Fassil Aseffa and Asfaw Hailemariam provided expertise on nodule sampling in the field and bacterial isolation in the laboratory.
# ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>Alpha</td>
</tr>
<tr>
<td>ACC</td>
<td>1-aminocyclopropane-1-carboxylate</td>
</tr>
<tr>
<td>AFLP</td>
<td>Amplified fragment length polymorphism</td>
</tr>
<tr>
<td>ADP</td>
<td>Adenosine diphosphate</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>β</td>
<td>Beta</td>
</tr>
<tr>
<td>BNF</td>
<td>Biological nitrogen fixation</td>
</tr>
<tr>
<td>bp</td>
<td>Base pair</td>
</tr>
<tr>
<td>C</td>
<td>Carbon</td>
</tr>
<tr>
<td>CO₂</td>
<td>Carbon dioxide</td>
</tr>
<tr>
<td>CSA</td>
<td>Central statistics authority</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DDH</td>
<td>DNA-DNA hybridization</td>
</tr>
<tr>
<td>Ha</td>
<td>Hectare</td>
</tr>
<tr>
<td>H₂</td>
<td>Dihydrogen</td>
</tr>
<tr>
<td>H</td>
<td>Hydrogen</td>
</tr>
<tr>
<td>hsn</td>
<td>host specific genes</td>
</tr>
<tr>
<td>FAME</td>
<td>Fatty acid methyl ester analyses</td>
</tr>
<tr>
<td>EPP</td>
<td>Ethiopian Pulses Profile</td>
</tr>
<tr>
<td>IAA</td>
<td>Indoleacetic acid</td>
</tr>
<tr>
<td>ITS</td>
<td>Internal transcribed spacer</td>
</tr>
<tr>
<td>Kb</td>
<td>Kilobit</td>
</tr>
<tr>
<td>Kg</td>
<td>Kilogram</td>
</tr>
<tr>
<td>μg</td>
<td>Microgram</td>
</tr>
<tr>
<td>MLEE</td>
<td>Multilocus enzyme electrophoresis</td>
</tr>
<tr>
<td>MLSA</td>
<td>Multilocus locus sequence analysis</td>
</tr>
<tr>
<td>NH₃</td>
<td>Ammonia</td>
</tr>
<tr>
<td>NH₄⁺</td>
<td>Ammonium</td>
</tr>
<tr>
<td>N₂</td>
<td>Di-nitrogen</td>
</tr>
<tr>
<td>nif</td>
<td>nitrogen-fixation</td>
</tr>
<tr>
<td>NO₃</td>
<td>Nitrate</td>
</tr>
<tr>
<td>nod</td>
<td>nodulation</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PGP</td>
<td>Plant growth promoting</td>
</tr>
<tr>
<td>Pᵢ</td>
<td>Inorganic phosphate</td>
</tr>
<tr>
<td>RAPD</td>
<td>Randomly Amplified Polymorphic DNA</td>
</tr>
<tr>
<td>rDNA</td>
<td>Gene encoding ribosomal RNA</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>pSym</td>
<td>Symbiotic plasmid</td>
</tr>
<tr>
<td>sv</td>
<td>symbiovar</td>
</tr>
<tr>
<td>Tg</td>
<td>Million tons</td>
</tr>
<tr>
<td>GUS</td>
<td>β-glucuronidase</td>
</tr>
<tr>
<td>YMA</td>
<td>Yeast manitol agar</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

Nitrogen is one of the major primary essential elements for plant growth along with phosphorus and potassium. It is an important constituent of plant and animal proteins and nucleic acids and an essential part of the metabolic process in transfer of energy. It is also a fundamental component of chlorophyll and hence plays a role in photosynthesis. Molecular nitrogen ($\text{N}_2$) covers approximately 78% of the earth's atmosphere; nevertheless $\text{N}_2$ is chemically inert and cannot be directly assimilated by plants (Frank et al., 2003). Therefore, $\text{N}_2$ needs to be converted into plant-accessible form, such as ammonium ($\text{NH}_4^+$) or nitrate ($\text{NO}_3^-$). However, most living organisms cannot break the strong triple bond within $\text{N}_2$. Only limited numbers of prokaryotes are able to convert the $\text{N}_2$ molecule into a usable form of nitrogen through a process known as biological nitrogen fixation (BNF).

Specialized microorganisms that belong to the domains of Bacteria and Archaea are able to reduce atmospheric $\text{N}_2$ into ammonia ($\text{NH}_3$). Some of these organisms are free-living, nitrogen-fixing soil microorganisms such as $\text{Azotobacter}$, $\text{Azospirillum}$ and some methanogenic archaea. Others form nitrogen-fixing symbiotic associations with plants, for example the associations of rhizobial Proteobacteria with legumes, the Actinobacteria $\text{Frankia}$ with actinorhizal plants, and cyanobacteria with ferns, fungi, Gunnera or cycads (Franche et al., 2009; Raymond et al., 2004; Shridhar, 2012).

In BNF the conversion of atmospheric $\text{N}_2$ into $\text{NH}_3$ is mediated by the bacterial nitrogenase enzyme (Frank et al., 2003). This nitrogen-fixing machinery is an exclusive property of nitrogen-fixing prokaryotes known as diazotrophs. In terms of metal content this enzyme is divided into three structurally, mechanistically and phylogenetically related sub-types: Mo-nitrogenase, Fe-nitrogenase and V-nitrogenase depending on the metal cluster in the active site. Most diazotrophs contain the conventional Mo-nitrogenase, which contains an iron-molybdenum cofactor (FeMoCo). Nitrogenase that contains vanadium (V-nitrogenase) and/or only iron (Fe-nitrogenase) in its cofactor is also obtained from some free-living diazotrophs, such as $\text{Azotobacter}$, $\text{Clostridium}$, $\text{Rhodopseudomonas}$, $\text{Methanosarcina}$, $\text{Azomonas}$, $\text{Azospirillum}$, $\text{Helio bacterium}$ and cyanobacteria (for review see, Bothe et al., 2010; Rubio and Ludden, 2005). The reaction of BNF ($\text{N}_2 + 8\text{H}^+ + 8\text{e}^- + 16\text{Mg ATP} \rightarrow 2\text{NH}_3 + \text{H}_2 + 16\text{ Mg ADP} + 16\text{Pi}$) is an energy expensive process requiring 16 moles of ATP to convert one mole of $\text{N}_2$ into two mole of $\text{NH}_3$ (Cheng, 2008).

1.1. BNF: Implications for sustainable ecosystem services

Before the invention of the Haber–Bosch process, BNF was almost the only global sources of fixed nitrogen with just small amounts of $\text{N}_2$ reduced by lightning, combustion and volcanic activity (Smil, 2001). Nowadays, the Haber–Bosch process is estimated to produce 120 million tons (Tg) of N per annum in the form of $\text{NH}_3$ and this is mainly used for the purpose of agriculture and food production (Galloway et al., 2008). Previously, the annual global BNF was expected to provide about 175 Tg N (Burns and Hardy, 1975). This figure has been revised since then; it was for instance estimated at 122 Tg N by Burris (1980) and at 195 Tg N by Cleveland et al. (1999), and Galloway et al. (2004) estimated that on a global
scale BNF delivers 110 Tg N on land per year. Over 50 % of the biological source of fixed N is obtained from the legume-rhizobia symbiosis (Zahran, 1999). Herridge et al. (2008) estimated that food legume–rhizobia symbioses annually delivered about 21 Tg N and the corresponding fixed N input by forage and fodder legumes–rhizobia symbioses was reported to be 12–25 Tg N. The fixed N contribution by other symbiotic, associative and free-living bacteria in cereal, oilseed and other crop lands was estimated to amount to less than 18 Tg N. These data show that the BNF process, particularly the legume-rhizobia symbiosis, contributes ample fixed N sources in the soil. The contribution of fixed carbon (C) by legumes was reported to be directly proportional to the fixed N (Macedo et al., 2008). Thus, BNF also plays a very important role in sequestering atmospheric carbon dioxide (CO₂) and in increasing the soil C stock. Thereby, BNF increases soil organic content and fertility by providing fixed N and C nutrients that can be utilized subsequently by other plants that are not nitrogen fixing.

Because of the fixed N they get from symbiosis, legume species grow better than other plants in poor degraded soils and they tolerate marginal lands better than other plants. Thus, legume trees have been used successfully in afforestation programs in order to rescue degraded lands and to restore ecosystem biodiversity (reviewed by Chaer et al., 2011).

It is obvious that the nitrogen fertilizers produced industrially are the main source of N in intensive crop production systems. Nevertheless, application of chemical fertilizers leaves negative impacts to the environment. During fertilizer production the chemical industries release high amounts of CO₂ (about 275 Tg per year). Most of the applied N fertilizer on farms is not efficiently utilized by the plants and subsequently the unutilized N fertilizers pollute terrestrial and aquatic systems. Thus the chemical fertilizers become a threat for the quality of climate, water, soil and biodiversity (Olivares et al., 2013). On the other hand, BNF has significantly lower economic and environmental costs compared to the chemical fertilizer industries. It is the main sources of accessible N for living organisms in natural ecosystems and is an essential component of sustainable ecosystem services. As a result of this, BNF in general and legume-rhizobia symbioses in particular has been a subject of research for over one hundred years, since the first nitrogen-fixing bacteria isolated by Beijerinck (1888).

1.2. The plant family Leguminosae

Among the flowering plant families the Leguminosae is the third largest after Orchidaceae and Asteraceae, represented by about 730 genera with more than 19, 320 species (Lewis et al., 2005). The Leguminosae family is widely distributed all over the world (Rundel, 1989). However, their diversity is greater in the tropics and subtropics than in temperate regions. The family Leguminosae is categorized into three subfamilies, mainly based on morphological differences of their flowers (Allen and Allen, 1981): Caesalpinioideae, Mimosoideae and Papilionoideae (Sprent, 1995). The subfamily Caesalpinioideae contains approximately 2250 species within four tribes, which mainly represent tropical woody and forest legumes of Africa, South America and South East Asia. The subfamily Mimosoideae is also divided into four tribes and includes about 3270 species with mainly small trees and shrubs of subtropical regions of Africa, Americas, Asia and Australia. The subfamily
Papilionoideae is the largest of the three subfamilies containing 28 tribes. Approximately 13,800 species of the Papilionoideae subfamily represent diverse groups of herbaceous and woody legumes and are distributed widely in the temperate and tropical regions of the globe (Al-Taweil et al., 2009; Lewis et al., 2005).

In general legume plants show a vast diversity in morphology, habitat and ecology. However, all legume species can be distinguished from non-leguminous plants as they produce pods as a common feature (NAS, 1979). Most species of the Leguminosae form symbiotic associations with nitrogen-fixing root nodule forming bacteria. About 23% of the Caesalpinioideae, 90% of the Mimosoideae and 97% of the Papilionoideae legume species were found to have symbiotic associations with nitrogen-fixing rhizobia (Sprent, 1995). Many legumes form symbiotic associations with arbuscular mycorrhizal fungi and ectomycorrhizal fungi, which might help them in taking up inorganic nitrogen and phosphorus compounds from the soil (Sprent, 2001). Therefore, the legumes are an essential part of the terrestrial nitrogen cycle and used to sustain ecosystem functioning (Sprent, 2001).

With regard to economic and agricultural significance the family Leguminosae is second next to Poaceae (grasses). Legumes are grown for production of food and oil and for fiber, fuel, timber, medicines, forages, biodiesel fuel, and chemicals (http://www.ildis.org/Leguminosae/).

### 1.2.1. Woody and shrub legumes: An overview of use in the tropics and subtropics

Most of the tropical and subtropical woody legumes are able to make symbiotic associations with root nodule forming bacteria. The occurrence of nodulation and nitrogen fixation in tree legumes was reviewed by Sprent and Parsons (2000). Because of the nitrogen-fixing ability the woody and shrub legumes grow fast and consequently provide a range of economically and agriculturally important services. As a result of the diverse utility they provide, many trees and shrub legumes have been recognized as multipurpose, being used as shade, forage, fodder, fuel wood, for soil enrichments, and timber production (Graham and Vance, 2003; Lewis et al., 2005).

In the tropics and subtropics, leguminous trees and shrubs are used in rehabilitating degraded farmlands as well as over-exploited and marginal soils, where the legumes improve soil fertility by increasing the content of nitrogen and organic carbon in the soil (Sileshi et al., 2008). They are used in inter-cropping or as fallow to crop in low-input agriculture, and in agroforestry and plantation forestry systems (Diabate et al., 2005; Sileshi et al., 2011). Among multipurpose legume trees and shrubs, for example Butea, Dalbergia and Millettia have been used intensively in India. In Central America, species belonging to the genera of Calliandra, Gliricidia, Inga and Leucaena have been used for different purposes. The legume tree species Faidherbia and Acacia are among the most valued legume trees in Africa (Lewis et al., 2005). In Ethiopia 13 genera, namely Acacia, Aeschynomene, Albizia, Baphia, Cajanus, Dalbergia, Enterolobium, Erythrina, Faidherbia albida, Mimosa, Pithecellobium, Pterocarpus and Sesbania are among the outstanding nitrogen-fixing trees and shrubs recommended by the Nitrogen Fixation Tree Association (Hunde and Thulin, 1989). However, the proper utilization of legume trees and shrubs in Ethiopia is minimal,
even though some legume trees and shrubs have been used in traditional agroforestry systems, such as home-gardens.

The leguminous tree *Erythrina brucei* is an endemic in Ethiopia and a popular in traditional agroforestry systems in central southern Ethiopia, where the tree can be integrated with crops, such as barley, maize and sorghum or used as shade for coffee plantations (Fig. 1) (Thulin, 1989). *Erythrina* species are also used for feed, fuel wood, and as construction material (Teketay, 1990). Several *Indigofera* species are naturally growing in low moisture areas and show resistance to salinity (Skerman, 1982). Among the species growing in Ethiopia, *I. caerulea*, *I. vicioides* and *I. arrecta* were found to tolerate shortage of water and grow well in low moisture conditions in the greenhouse. Thus, they were suggested as potential good sources of fodder for livestock in the moisture low regions and during long dry seasons (Hassen et al., 2007). In addition, *I. arrecta* and *I. caerulea* can be used for indigo pigment production (Liogier, 1990). *Crotalaria* species are used as green manure and some for feed production (Diabate et al., 2005; Fischler and Wortmann, 1999). As Gathumbi et al. (2004) reported, *Crotalaria* established well under relay cropping with maize. A case study in Uganda showed that the use of *C. ochroleuca* as green manure improved maize grain yield by 39% over the control (Fischler and Wortmann, 1999). Similarly, the production of maize was also improved in Ethiopia when *Crotalaria* species were used as a fallow or in intercropping (Bogale et al., 2001). In addition, *E. brucei* and *I. arrecta* as well as some other species of *Crotalaria*, *Erythrina* and *Indigofera* are reported to be used as traditional medicines in Ethiopia for curing different diseases of humans as well as cattle (Bekalo et al., 2009).

![Fig.1. Coffee growing under the shade of endemic legumes *Erythrina brucei* (tree without leaves) and *Acacia abyssinica* at Wondogenet College in Ethiopia (photo taken by Kristina Lindström).](image-url)
1.2.2. Food legumes: Focus on cultivation and use in Ethiopia

Generally, food legumes are an integral part of African farming systems, covering usually large parts of the farmlands in the region. For example, in the years 2006-2008 in Sub-Saharan Africa more than 20 million hectares of the farmlands (which represent about 28% of the global pulse areas) were used for food legume crop production (Akibode, 2011). Ethiopia has a long tradition of cultivation of food legumes and the country is considered as a center of genetic diversity for several cool season legume crops, such as field pea (*Pisum sativum*), faba bean (*Vicia faba*), lentil (*Lens culinaris*), chickpea (*Cicer arietinum* L.), fenugreek (*Trigonella foenum-graecum* L.) and grass pea (*Lathyrus sativus*) (Tilaye et al., 1994). These food legumes are extensively grown in the cooler highlands of the country, whereas the warm season pulse crops, such as common bean (*Phaseolus vulgaris*), cowpea (*Vigna unguiculata*) and groundnut (*Arachis hypogaea*) are common in warmer and lowland parts of Ethiopia.

Soybeans and common beans are the first and third (after groundnut) most important food legumes cultivated worldwide, respectively (Singh, 1999). Common bean is one of the most widely distributed crop legumes, grown in many parts of the tropics, sub-tropics and temperate regions. In Ethiopia, common bean is commonly cultivated in the southern and eastern regions of Ethiopia at altitudes between 1400 and 2000 meter, as a sole crop or intercropped with sorghum, maize or other crops. The Rift Valley zones and the Hararghe highlands are the major producing areas of common bean (Assefa et al., 2006). Soybean is not an indigenous plant in Ethiopia and this crop was first introduced to Ethiopia as a trial in 1950, but it was soon abandoned due to low yields. The real production was started later in the 1970s with the introduction of high yielding soybean varieties from Europe and the USA (Alema, 1981 cited in http://www.soyinfocenter.com/books/134). Currently, soybean is cultivated mainly in the southern and western parts of the country in Awassa, Jimma, Bako, Pawe and Assosa.

Food legumes are a common source of food in developing countries, including in the Sub-Saharan African countries. In Ethiopia, they are an important part of the national diet and provide a protein source and essential nutritional additives to cereals. The seeds of food legumes can be used whole split or as flour to make a sauce for "wot" (an Ethiopian stew) and served as part of the main dish. The legume seeds can also be consumed as a snack in boiled or roasted forms. Food legumes are also used as income generating commodities for farmers. Common bean and soybean together with faba beans are essential export commodities of the country. In the years 2006-2008, Ethiopia was the leading exporter of common bean in Africa, mainly exporting it to Kenya, Yemen, United Arab Emirates, USA, England, Italy and Germany (Akibode, 2011; Katungi et al., 2011). The pulses are often used as a rotation crop with cereals, by which they increase soil nitrogen content and improve soil nutrients for subsequent cereal plantation.

Even though food legumes have several benefits, their productivity is very low in Ethiopia. For example, data reported by CSA (2009) shows that the yield of soybean was below 1 ton per hectare. In 2010, in Ethiopia a total of 8700 tons of soybean was harvested from 5679 ha,
which is equivalent to 1.5 tons/ha, being less than half of the yield (3.2 tons/ha) registered by the highest world producer, USA (90,605,500 tons) of the same year (http://faostat.fao.org/site/567/DesktopDefault.aspx?PageID=567#ancor). Although the productivity of common bean has increased due to improved bean varieties, the average yield in smallholder farmers land is still very low (0.5-0.8 ton/ha) compared to the yields from research plots (2.5-3.0 tons/ha) (EPP, 2004). This yield gap is mainly attributed to shortage of improved seeds for farmers, poor farm management practices and low soil fertility of the farmlands, mainly due to low nitrogen content of the soils (Beyene and Tsige, 1986; EPP, 2004).

1.3. The symbiotic association of legumes with rhizobia

1.3.1. Definition of symbiosis

The term symbiosis was originally used by de Bary (1879) to refer “the living together of differently named organisms”. The word was used as synonymous to mutualism, in which both interacting organisms live together for mutual benefit in contrast to commensalism, where one organism benefits and the other partner is unaffected or parasitism, in cases when one of the interacting organism profits at the expense of the other partner. From the beginning, de Bary’s definition of symbiosis was criticized since it was broad and lacking inference of benefits or harms that might be caused by the association. The confusion, disagreements and turbulence of the use of the term “living together” have been continued among the biologists for more than 130 years (for more detail, see the review by Martin and Schwab, 2013). Nevertheless, the “living together” usage has been accepted as a suitable concept to define the term “symbiosis” and recently the broader definition of symbiosis (i.e. mutualism, commensalism, and parasitism) is increasingly used in the textbooks (Martin and Schwab, 2013).

The nitrogen-fixing bacteria that form symbiotic associations with leguminous plants are commonly known as rhizobia. The rhizobia-legume symbiotic interaction induces specialized organs known as nodules on roots or stems of host legumes. Inside nodules rhizobia reduce atmospheric N\(_2\) to NH\(_3\). The symbioses between rhizobia and legume plants are mainly a mutualistic interaction (Lindström and Mousavi, 2010; Saffo, 2001). However, it seems that there are cases where these partnerships can also be considered as parasitic. The rhizobia can have two lifestyles: as endo-symbiont inside nodules or as free-living saprophytes in the soil or rhizosphere. As endo-symbiont, the rhizobia promote growth of the host plant by supplying fixed nitrogen and in turn rhizobia get carbohydrates and energy from the host legumes. This phenomenon indicates the existence of mutualism between the two organisms. But the rhizobia can also be considered as parasitic when they form ineffective symbiosis with legumes, in which the rhizobia get a continuous nutrient supply while they fix little or do not fix nitrogen for the host plant. These types of situations may occur when multiple rhizobial strains compete for the same plant and when the strains infect non-specific hosts promiscuously. These rhizobia can also form effective symbioses (mutualistic) when they interact with their own specific host legumes (Denison and Kiers, 2004). For example, *Rhizobium leguminosarum* strains isolated from nodules of native legumes in New Zealand were found to form ineffective nodules, while they still reserved
the ability to form an effective symbiosis with classical host plants (pea, bean and clover) of *R. leguminosarum* (Weir, 2006). *Rhizobium galegae* strains isolated from *Galega orientalis* or *Galega officinalis* form effective symbioses with their original hosts, whereas strains isolated from *G. orientalis* form ineffective nodules on *G. officinalis* and vice versa (Andronov et al., 2003; Lindström, 1989).

In addition, as Soto et al. (2006) stated in their review, the rhizobia and plant pathogenic bacteria use similar mechanisms and genetic traits to colonize and infect the host plant. Like pathogenic bacteria, the rhizobia cause hypersensitive reactions on host legumes during the infection process (Djordjevic et al., 1987). This means they can be considered possible pathogens by the host legumes before the formation of symbiotic nitrogen-fixing nodules. In connection with this, the rhizobia were also referred to as sympathogens (Spaink, 1995).

### 1.3.2. Mechanisms of rhizobia-legume symbiotic interactions

Symbiotic establishment between legumes and most rhizobial species is dependent on the complex molecular interactions between plant root exudates, and rhizobial lipochito-oligosaccharide signal molecules (Nod factors, LCOs) as it is schematically depicted in Fig. 2. However, in a few exceptions, the Nod factors are not required for symbiosis between some nodulating bacteria and legumes, a process known as Nod factor independent interaction (Giraud et al., 2007).

In the case of Nod factor dependent interaction, the formation of nodules on legume roots and stems by soil rhizobia are initiated when secondary metabolites, such as flavonoids, are secreted from the roots of host plants (Perret et al., 2000; Peters et al., 1986; Subramanian et al., 2007). The bacterial receptor NodD protein recognizes the flavonoids and the complex acts as an activator of expression of the rhizobial nodulation genes. The so called common nodulation genes *nodA*, *nodB*, *nodC* that code for acyl transferase, a chitin oligosaccharide deacetylase and N-acetylglucosaminyl transferase enzymes, respectively are highly conserved (Atkinson et al., 1994; reviewed in Lindström and Mousavi, 2010). These enzymes catalyze the production of the backbone of the Nod factors (Jones et al., 2007). The basic chemical structure of the Nod-factor that is common among the rhizobia has chitin backbone containing two to six β–1, 4 linked N–acetyl–D–glucosamine residues with the N-linked fatty acid chain attached at the non-reducing sugar terminal (Fig. 2.). The substitution groups, length, and fatty acid saturation levels of the Nod factors may differ among different rhizobial species. The various substitution groups decorating the backbone of the Nod factors are the main determinants of the host plant ranges of the rhizobia strains (Carlson et al., 1993; López-Lara et al., 1996; Perret et al., 2000). Thus, the host specific nodulation genes (*hsn*), such as *noe*, *nod* and *nol* that code for different decorations may also vary between rhizobial species (Moulin et al., 2004; van Rhijn and Vanderleyden, 1995; Cooper, 2007).

Nod factors are recognized by the plant through receptor-like kinases on the epidermal root hair cells. The formation of tubular structures known as infection threads near the root hair cells and the subsequent introduction of the rhizobia into the host is facilitated by cell wall degrading pectinases and cellulases that can be synthesized either by the rhizobia (Robledo
et al., 2012, 2011) or legumes (Xie et al., 2012). As a result, the rhizobial cells can easily pass the cell boundaries along the infection thread and enter to the cortical cells of the plants. A mitotic cell division at the cortical cells around the infection zone leads to the formation of a nodule primordium. The rhizobia are released into the inner cells of the nodules with the help of the infection thread network through the cortical cells that grow towards the nodule primordial. The rhizobial cells then differentiate into nitrogen-fixing bacteroids inside the nodule compartment known as the symbiosome (Broughton et al., 2000; Oldroyd and Downie, 2008). In addition to plant receptor-kinase, the Nod-factor also prompts a calcium dependent signal transduction pathway in legumes. This brings fluctuation in the calcium concentrations in the nucleoplasm and nuclear-associated cytoplasm (known as calcium spiking) and this initiates expression of other plant specific genes called early nodulins (Oldroyd and Downie, 2008). The nodulin proteins assist the development and the process of nodule organogenesis (Verma et al., 1986).

Fig. 2. Signal exchange in rhizobium-plant symbiosis. Flavonoids produced by the host plant induce rhizobial nod genes. This leads to the production of nodule-inducing (Nod) factors, lipochitooligosaccharides (LCOs), that are differently modified depending on the Rhizobium species. The insert shows an infection thread passing the root cortex toward a cluster of dividing cells that will become a root primordium (adapted and reprinted from Schultze and Kondorosi (1998), with permission). The Nod factor structure part of the Schematic figure was reprinted from Dresler-Nurmi et al., 2009, with permission. Abbreviations for substitutions in Nod factor: Ac, acetyl; Ara, arabinosyl; Et, ethyl; Cb, carbamoyl; Fuc, fucosyl; G, N-glycolyl; Gro, glyceryl; H, hydrogen; Man, mannosyl; Me, methyl and S, sulfate.
1.4. Nodule endophytic bacteria

1.4.1. General overview

In addition to symbiotic rhizobia, legume nodules may also harbor non-symbiotic bacteria. These bacteria are not able to induce nodules by themselves but they coexist with nodulating rhizobia inside nodules without causing any recognized harm to the host plant. These are known as endophytic bacteria (Hallman et al., 1997). According to Hardoim et al. (2008) plant rhizosphere bacteria can colonize internal plant tissues and become opportunistic endophytes. The process of plant colonization by rhizosphere bacteria and subsequent endosphere colonization was reviewed by Compant et al. (2010). Even though it is not very clear how they enter nodules, endophytic bacteria have been isolated from the nodules of various legume species, for example from *A. hypogaea* (Ibáñez et al., 2009), *Glycyrrhiza* spp. (Li et al., 2012), *Mimosa pudica* (Pandey et al., 2005), *V. faba* (Lei et al., 2008), *C. arietinum*, *P. vulgaris* (Mhamdi et al., 2005; Saïdi et al., 2011) and *Glycine max* (Bai et al., 2002; Li et al., 2008). These studies showed that taxonomically the endophytic bacteria mainly belong to the genera *Agrobacterium*, *Bacillus*, *Burkholderia*, *Enterobacter*, *Pantoea*, *Paenibacillus*, *Pseudomonas* and *Serratia*.

1.4.2. Plant growth promoting properties of nodule endophytic bacteria

The role of endophytic bacteria in the nodules is not yet well known; however it is thought that the endophytes find an ecological niche inside the nodules to shelter themselves from environmental stress and microbial competition in the rhizosphere (Kobayashi and Palumbo, 2000). In reverse, endophytic bacteria might benefit the growth of the host plant by suppressing plant disease causing bacteria via competition for nutrients, production of plant growth promoting phytohormones, such as indoleacetic acid (IAA), or provision of solubilized mineral phosphate (Kuklinsky-Sobral et al., 2004; Li et al., 2008). Some endophytic or rhizosphere plant growth promoting bacteria have the ability to synthesize lytic enzymes, such as chitinases, cellulases, glucanases, proteases and lipases that can degrade fungal cell walls and thus found to have biocontrol activity against a number of pathogenic fungi (reviewed by Glick, 2012, Lodewyckx et al., 2002). In addition, with the help of cellulase plant growth promoting bacteria penetrate and form localized lesions on plant cell walls and this phenomenon is thought to enhance the infection process of symbiotic bacteria in legumes (Sindhu and Dادرвал, 2001). Endophytic bacteria capable of producing the enzyme 1-aminocyclopropane-1-carboxylate (ACC) deaminase promote plant growth by cleaving plant-produced ACC (a precursor of ethylene) and lowering the concentration of the plant hormone ethylene. This plant hormone acts as plant growth regulator at lower concentration and is also required to break seed dormancy but, after germination, sustained high level of ethylene may inhibit root elongation (Glick, 2005).

Inoculation of nodule endophytic bacteria or plant growth promoting endophytic or rhizosphere bacteria (PGPR) in general has been considered as a significant practice to improve plant growth and yields (Bai et al., 2003, 2002; Fox et al., 2011; Zhang et al., 1996). Shoot and root dry matter of the fodder legume galega were increased when this legume
inoculated with a mixture of *R. galegae* HAMBI540 and root colonizing *Pseudomonas trivialis* 3Re27 or soil rhizosphere *Pseudomonas extremorientalis* TSAU20 as compared with legume inoculated with *R. galegae* HAMBI 540 alone (Egamberdieva et al., 2009). Nodulation and symbiotic effectiveness of *Medicago truncatula* were improved when the plant co-inoculated with PGPR *Pseudomonas fluorescens* WSM3457 and *Ensifer* (syn. *Sinorhizobium) medicae* WSM419 (Fox et al., 2011). The shoot dry weight and nitrogen and phosphorus contents in common bean plants were improved after co-inoculation of *Rhizobium phaseoli* with *Pseudomonas* sp. LG obtained from alfalfa root nodules or with soil rhizosphere *Bacillus* sp. Bx (Stajkovic et al., 2011). An endophytic *Bacillus* sp. strain isolated from a soybean nodule was reported to increase nodulation and growth of the plant and provide better soybean yield when inoculated alone or when co-inoculated with *Bradyrhizobium japonicum* (Bai et al., 2003, 2002). The nodule formation by *Sinorhizobium meliloti* on *Wisteria sinensis* was enhanced when this strain was co-inoculated with nodule endophytic *Agrobacterium* sp. II CCBAU 21244 (Liu et al., 2010).

### 1.5. Taxonomic overview of rhizobia

Symbiotic nitrogen fixation in legumes was first thought to be restricted to a limited number of bacterial species that belong to the alpha-proteobacteria in the genera *Rhizobium* (Frank, 1889), *Azorhizobium* (Dreyfus et al., 1988), *Allorhizobium* (de Lajudie et al., 1998), *Mesorhizobium* (Jarvis et al., 1997), *Ensifer* (syn. *Sinorhizobium*) (de Lajudie et al., 1994) and *Bradyrhizobium* (Jordan, 1982). However, with continuous explorations of legumes in new biogeographic regions and with the development of modern taxonomic tools, new rhizobial species were also described from other genera of the alpha-proteobacteria (Vandamme et al., 1996). These included root nodule nitrogen-fixing bacterial species in the genera *Devisia* (Rivas et al., 2002), *Methyllobacterium* (Sy et al., 2001), *Ochrobactrum* (Trujillo et al., 2005), *Phyllobacterium* (Valverde et al., 2005) and *Shinella* (Lin et al., 2008). Nodulation of legumes by bacterial species belonging to the genera *Burkholderia* (Moulin et al., 2001) and *Cupriavidus* (Chen et al., 2001) were also reported in the beta-proteobacteria. The discovery of new bacterial species that are capable of making nitrogen-fixing symbiotic associations with legumes is in progress and new species are discovered continuously. More recently, Maynaud et al. (2012) identified *Aminobacter anthyllidis* from the nodules of the trap host *Anthyllis vulneraria* in zinc–lead mine tailings as the first new nitrogen-fixing bacterial symbionts in the genus *Aminobacter*, which is a close relative of the genus *Mesorhizobium* (alpha-proteobacteria). Other newly identified rhizobia from alpha-proteobacteria, such as *Microvirga lupini* and *Microvirga lotononidis* are nodulating *Lupinus texensis* and *Microvirga zambiensis* is a symbiont of *Listia angolensis* (Ardley et al., 2012).

Although all of the above mentioned rhizobia were isolated from the nodules of leguminous plants, certain *Bradyrhizobium* species (Lafay et al., 2006) were found to form nitrogen-fixing nodules on the roots of non-leguminous species *Parasponia* (Akkermans et al., 1978). This suggests that plant nodule forming nitrogen-fixing bacteria are diverse in respect to their host ranges and in the future novel rhizobial species are expected to be discovered from nodules of unexplored plant species.
Generally, the rhizobia are taxonomically diverse: phylogenetic trees show symbiotic species, are taxonomically entwined with non-symbiotic close relatives. Recently, we have studied 16S rRNA gene phylogeny of 160 proteobacterial species, including most symbiotic nitrogen-fixing rhizobia and related taxa (Lindström et al., 2013). Based on the phylogenetic tree in Fig.3, some rhizobial species were grouped closely or distantly with non-symbiotic bacteria (for detail review, Lindström et al., 2013). For example, the non-symbiotic *Rhizobium tarimense*, *Rhizobium rhizogenes*, *Rhizobium soli* were intermixed with symbiotic *Rhizobium* spp. Non-nodulating *Agrobacterium* species formed a sister clade to the genus *Rhizobium*.

*Agrobacterium* spp. were first classified into three species based on their interaction with plants as tumor forming *Agrobacterium tumefaciens*, hairy root producing *Agrobacterium rhizogenes* and non-pathogenic *Agrobacterium radiobacter* (Kersters, 1984). The genus *Agrobacterium* has close relationship with the genus *Rhizobium* in respect to several phenotypic as well as molecularly characteristics. The paraphyly of the two genera has brought a long-lasting debate about their taxonomy (Farrand et al., 2003; Young et al., 2006, 2003, 2001). In brief, according to Young et al. (2001), all *Agrobacterium* spp. were included in the genus *Rhizobium* since neither *Rhizobium* nor *Agrobacterium* was monophyletic based on 16S rRNA gene phylogeny. Nevertheless, Farrand et al. (2003) proposed again to retain the genus *Agrobacterium* by presenting phenotypic and molecular differences between *Agrobacterium* and *Rhizobium* species. This presented confusion among scientific community and the reclassification of *Agrobacterium* as an independent genus was not accepted by all scientist (Young et al., 2003). Consequently, *A. tumefaciens* was sometimes called *A. radiobacter*, *A. rhizogenes* or *Rhizobium rhizogenes*. The later species usually form a well-supported clade together with closely related species *Rhizobium tropici*, *Rhizobium leucaenae*, *Rhizobium multihospitium*, *Rhizobium miluonense* and *Rhizobium lusitanum* (e.g. Costechareyre et al., 2010; Ribeiro et al., 2012, Velázquez et al., 2010). To resolve the confusions concerning the nomenclature, the subcommittee on the taxonomy of *Agrobacterium* and *Rhizobium* suggested transferring of the biovar 2 of *Agrobacterium* (Keane et al., 1970) were called as *A. radiobacter*, *A. rhizogenes* or *Rhizobium rhizogenes*. The later species usually form a well-supported clade together with closely related species *Rhizobium tropici*, *Rhizobium leucaenae*, *Rhizobium multihospitium*, *Rhizobium miluonense* and *Rhizobium lusitanum* (e.g. Costechareyre et al., 2010; Ribeiro et al., 2012, Velázquez et al., 2010). To resolve the confusions concerning the nomenclature, the subcommittee on the taxonomy of *Agrobacterium* and *Rhizobium* suggested transferring of the biovar 2 of *Agrobacterium* into the genus *Rhizobium* (Lindström and Young, 2011). As a result, these species was proposed be called *Rhizobium rhizogenes* instead of *A. radiobacter* or *A. rhizogenes*. However, the genus *Agrobacterium* was maintained by the subcommittee for the species that were reported to be monophyletic, such as *Agrobacterium vitis* (i.e. biovar 3), *Agrobacterium rubi*, *Agrobacterium larrymoorei* and *A. tumefaciens* complex (biovar 1 genomovars, G1 to G9) (Costechareyre et al., 2010). In support of this, accurate and rapid identification of the species or genomic species of *Agrobacterium* was possible using *Agrobacterium*-specific designed recA primers (Shams et al., 2013). In this work as well as in Fig. 3, some species such as *R. aggregatum*, *R. rosettiformans*, *R. pusense*, *R. nepotum* and *R. skierniewicense* were identified in *Agrobacterium* clade. However, the genus name *Rhizobium* was used for them since they were originally described as *Rhizobium* and henceforth these may need to be transferred officially into the emended genus *Agrobacterium*. In addition, phylogenetic tree constructed from six concatenated core gene sequences in our laboratory showed that *A. rubi* and strains from *Agrobacterium* genomic species G1-G9 and G13 were monophyletic. However, *A. vitis* and *Rhizobium taibaishanense* formed a separate clade together with *Allorhizobium undicola*.
(Unpublished data by Mousavi et al.), suggesting that all of these species may need to be included into genus *Allorhizobium*. Therefore, in this thesis including in the published papers, we used the genus *Agrobacterium* for strains belonging to biovar 1 of the *Agrobacterium* (i.e. *A. tumefaciens* complex) and *Rhizobium* for strains belonging to the former *A. rhizogenes*.

Fig. 3. The phylogenetic tree based on 16S rRNA gene sequences, illustrating the relationships among 160 proteobacterial species (reprinted from Lindström et al., 2013, with permission of John Wiley & Sons, Inc.). The neighbour-joining tree was constructed with 10,000 bootstrap replicates, and the cut-off is 50%. The type strains are shown by a “T” at the end of each strain code. The genus names are abbreviated as follows: *A.*, *Agrobacterium*, *B.*, *Bradyrhizobium*, *M.*, *Mesorhizobium*, *R.*, *Rhizobium*, and *S.*, *Sinorhizobium*. Other genera of *Alphaproteobacteria* are in blue, and the *Betaproteobacteria* are in grey. Alternative names: * Rhizobium leguminosarum*, ** *Pseudorhodobacter ferrugineus*, *** *Agrobacterium atlanticum*. 

23
1.6. Taxonomic criteria used in classification of the rhizobia: An overview

The first root nodule bacteria called *Bacillus radicicola* were isolated from root nodules of pea plants by Beijerinck (1888) and these were renamed as *Rhizobium leguminosarum* by Frank (1889). Since then, a considerable number of rhizobial species have been uncovered and the taxonomy of rhizobia has been changing from time to time due to the finding of new rhizobial genera and species. In the 1920’s phenotypic properties of bacteria were used as criteria for bacterial classification (Bergey et al., 1923). In view of that, the conventional morphological characteristics, such as rod-shape, Gram-negative, aerobic, non-spore forming and motile nature of the rhizobia were considered as criteria for their classification; however, these could not always distinguish them from other soil bacteria. Consequently, these methods turned out to be less valid in rhizobial taxonomy. Later, Fred et al. (1932) used the cross inoculation group concept, that is the host range of the rhizobia, as criteria for rhizobial classification, and they could then distinguish six taxonomically different root nodule bacteria. This method also helped to distinguish rhizobia from related agrobacteria since the genus *Rhizobium* was found to contain only species involved in nodule formation on legume roots, whereas, *Agrobacterium* comprised plant pathogenic species. As was published in Bergey’s Manual of Bacteriology published in 1974, all the six rhizobial species were recorded under one genus *Rhizobium* (Jordan and Allen, 1974). These were classified into fast- and slow-growing rhizobial species based on their growth rate on yeast extract–mannitol-mineral salts medium (YM). In addition flagellar position, DNA base composition and the genera of host plants were also recognized as distinguishing features between fast- and slow-growing rhizobia (Jordan and Allen, 1974). Thus, the rhizobia that nodulate *Pisum, Trifolium, Phaseolus* and *Medicago* were identified as fast-growing *R. leguminosarum, Rhizobium trifolii, R. phaseoli* and *Rhizobium meliloti*, respectively, while two slow-growing bacterial species that nodulated *Lupinus* and *Glycine* were classified as *Rhizobium lupini* and *Rhizobium japonicum*, respectively. The slow-growing root-nodulating bacteria were later included into a new genus *Bradyrhizobium* (Jordan, 1982). The cross inoculation group concept was considered as a significant criterion in defining rhizobial species for long time, because it was believed that all rhizobia could be clearly classified into different species according to their host ranges and legumes could also be identified based on their symbiotic partners. However, again the cross inoculation concept was not maintained as an independent criterion for rhizobial classification since several rhizobial species were found to infect a specific legume species promiscuously. For example legume species such as *G. max* or *P. vulgaris* can be infected by more than one genetically different rhizobial species. On the contrary, a single rhizobial species can nodulate more than one host legume. For example, *Rhizobium* sp. NGR234 was reported to nodulate more than 120 genera of legumes and the non-legume *Parasponia andersonii* (Pueppke and Broughton, 1999). These studies indicated that the rhizobia are highly diverse with respect to their host ranges as well as in their physiological properties.

In general, numerical taxonomy comprising biochemical, nutritional and serological characterization of the rhizobia together with the host ranges was used as the main method for classification and identification of root-nodulating bacteria until the molecular era. The rhizobial taxonomy has been improved considerably after the development of molecular
techniques, such as sequencing of the 16S rRNA gene (Woese et al., 1987). Polyphasic approaches, including classical phenotypic and morphological characteristics, whole cell protein analysis or multi-locus enzyme electrophoresis (MLEE), analysis of cellular fatty acids (FAME), DNA: DNA hybridization, RFLP of rRNA genes, 16S rRNA gene sequencing and symbiotic characteristics were for a long time considered as standard guidelines to describe new rhizobial species (de Lajudie et al., 1994; Jarvis et al., 1997; Graham et al., 1991).

1.7. Molecular methods used in rhizobial diversity and phylogenetic studies

In recent past years, efforts have been made to study the diversity and population structure of rhizobia nodulating cultivated herbaceous legumes, as well as wild legume trees, shrubs and medicinal plants in various regions of the world (for example Ando and Yokoyama, 1999; Carelli et al., 2000; Khbaya et al., 1998; Laguerre et al., 1996; Li et al., 2012). For these and similar works, molecular techniques and methods have been used. These include plasmid profile analysis (Selenska-Trajkova et al., 1990), RFLP of rRNA genes and intergenic spacer regions profiling (Gürtler et al., 1991), whole genome fingerprinting using AFLP (Amplified Fragment Length Polymorphism) (Vos et al., 1995), repetitive elements in the genome (rep-PCR) (Versalovic et al., 1991), RAPD (Randomly Amplified Polymorphic DNA) (Williams et al., 1990) and multilocus sequence analysis of core genes (Zeigler, 2003).

Molecular genetic markers in general provide better information, are more sensitive and relatively more accurate to study the relationship of closely related bacterial strains and in detection of higher rhizobial diversity than the phenotypic techniques. Phenotypic features such as colony morphology, and physiological or biochemical responses of a bacterium may vary depending on the media and conditions used for growing it. Nevertheless, it seems that each of the molecular markers and techniques has also its own limitation in determining the diversity and phylogeny of bacteria (for a review see Pontes et al., 2007). In addition, the study of soil rhizobial diversity can be affected by sampling errors and soil factors (Coutinho, 1996). Usually trap host legumes are used to isolate the rhizobia from soils, but only specific, efficient and competent rhizobia can form nodules on a specific host. The diversity of rhizobia that nodulate a specific host plants can also vary depending on the type of land use system and on soil edaphic and biotic factors (Sadowsky and Graham, 1998). Chromosome types and plasmid replication gene (repC) profiles results showed that the diversity of *R. leguminosarum* sv. *viciae* were higher in arable soils than in grassland soils (Palmer and Young, 2000). The authors also indicted that high potential nitrogen and phosphate levels in soils or acidity can affect rhizobial diversity negatively.
1.8. Genetic methods and markers commonly used in rhizobial diversity and taxonomic studies

Below, some significant genetic methods are described in more detail.

1.8.1. 16S rRNA gene sequence

The ribosomal ribonucleic acid (rRNA) gene was considered as a model genetic marker for studying the evolutionary history of life. Comparison of the rRNA sequences was used for construction of the “universal tree of life”, the tree which divided all organisms on earth into three main domains; *Eukarya, Bacteria* and *Archaea* (Woese et al., 1990). Bacteria contain genes coding for 5S, 16S, and 23S rRNAs and 16S–23S rRNA internal transcribed spacer (ITS) regions, in which all are typically structured into a segment of genome, called rRNA operon. The rRNAs are vital components of ribosomes that are involved in translation of messenger RNA and protein synthesis (Acinas et al., 2004; Jensen et al., 1993). Predominantly, the 16S rRNA gene sequence has been used as a standard genetic marker in identification and taxonomic classification of prokaryotes including rhizobia (Harris and Hartley, 2003; Garrity et al., 2001; Vandamme et al., 1996; Woese et al., 1990). This marker has several properties that make it suitable for phylogenetic inference. I) 16S rRNA is universally found in all living organisms, this enable the comparison of the phylogenetic relationship of all organisms and allows the construction of a tree of life. II) It is relatively highly conserved, thus the function of this gene over time has not changed, indicating that random sequence variations that may occur among organisms are more exact measure of evolution. III) The bacterial 16S rRNA gene sequence is a long stretch (about 1500 bp) that includes both conserved and variable regions, which can offer enough information for taxonomic purpose. Nine “hyper-variable” regions (labeled as V1 to V9) that are flanked by conserved stretches exhibit significant sequence variability among different bacterial species and can be used for bacterial phylogenetic studies (Van de Peer et al., 1996). Conserved regions are used in designing universal primers for PCR amplification of 16S rRNA gene and these also allow alignment of sequences of distantly related organisms (Baker et al., 2003; Chakravorty et al., 2007; Gutell et al., 1994).

Because of the above-mentioned significance, 16S rRNA gene sequence has been used extensively as a main criterion for phylogenetic classification of prokaryotes. In general, 97% 16S rRNA gene sequence similarity has been considered as a threshold for species delineation (Stackebrandt and Goebel, 1994). Subsequently, large numbers of 16S rRNA gene sequences have been deposited in the nucleotide databases. This facilitates the comparisons between different species and makes the 16S rRNA sequences further a choice of marker for identification and in building bacterial phylogenies (Woese et al., 1990; Weisburg et al., 1991). However, nowadays using 16S rRNA gene sequence for microbial diversity, phylogenetic and population studies must be regarded with caution as it was recognized that a bacterial genome can have multiple and heterogeneous ribosomal RNA operons (Acinas et al., 2004) and these can be also affected by horizontal gene transfer and genetic recombination (Eardly et al., 2005; van Berkum et al., 2003; Tuorova et al., 2001).
Depending on the species, a bacterial genome contains one up to fifteen rRNA operons. For example, *Bacillus cereus* strain ATCC 14579 and a number of strains of *Escherichia coli* were reported to contain 13 and 7 rRNA operons, respectively. Several species in the genus *Rhizobium*, *Mesorhizobium* and *Ensifer* contain three 16S rRNA copies (http://rrndb.mmg.msu.edu/search.php). Although, the operon number among closely related species was found to be conserved, moderate variations have also been found among strains in the same species (Haukka et al., 1996). For example, different strains of *B. japonicum* may have one or two rRNA operons. The majority of the 16S rRNA sequences derived from different operons of the same genome appeared to have high similarity (only about 1% nucleotide difference) (Acinas et al., 2004), but the variation between 16S rRNA sequences belonging to different operons of the same genome can also be large in some other organisms (Tuorova et al., 2001). Thus, this sequence heterogeneity within a genome may lead to overestimation of bacterial diversity in community study. In addition, the horizontal transfer of 16S rRNA gene among different organisms makes this marker uncertain to be used for reconstruction of phylogeny of bacteria.

In rhizobia, the variation in 16S rRNA gene sequences was shown to be unreliable to resolve species below the genus level (van Berkum et al., 2003). In many of the rhizobial groups, this region has slight divergence between species. Particularly in the genus *Bradyrhizobium*, phylogenetic classification at species level using this gene known to be difficult because the 16S rRNA gene is highly conserved in the genus *Bradyrhizobium* and the sequence similarity is relatively high among closely related species (van Berkum et al., 2003; Vinuesa et al., 2005; Willems et al., 2003). These studies indicated that the rRNA gene sequence phylogeny might not always accurately reflect the rhizobial taxonomy. The sequence analysis of 16S–23SrRNA internal transcribed spacer (ITS) is believed provide greater resolution than the 16S rRNA gene. Nevertheless, ITS analyses sometimes also complicated due to its high variation in size and sequences in some *Bradyrhizobium* strains (Willems et al., 2003).

### 1.8.2. DNA-DNA hybridization

The DNA-DNA hybridization technique is a common genetic method used to study bacterial heterogeneity, speciation and taxonomy (e.g. Degefù et al., 2013; Martens et al., 2008; Rosselló-Mora, 2006; Stackebrandt, 2003). The DNA-DNA hybridization (DDH) method has been used in determining specific differences between closely related bacterial species. After Krieg (1988) this technique has been used as a regular criterion for description of new bacterial species. It is a technique by which the entire sequence similarity of different organisms calculated from the pairwise whole genome comparisons (Rosselló-Mora, 2006). In practice DDH has three main steps: 1) Sheering of the genomic DNAs of test organism and reference strain, 2) mixing the DNA of both strains and heating to dissociate the DNA double strands, and 3) cooling the temperature down until fragments reanneal. The melting temperature value varies according to the base pair matching of the two strands and thereby gives a clue for genetic relatedness of the two strains (Auch et al., 2010). As an ad hoc committee of systematics recommended (Wayne et al., 1987), test bacterial strains to be considered as different species should show 70% or less DDH relatedness value from the references.
Methods used for measuring the DDH values can be varied in different laboratories (Rosselló-Mora, 2006). Practically all the methods are based on the same principle but the DNA quality, concentration or tagging and washings steps may vary from one laboratory to the other. These differences can lead to prominent errors or give conflicting result (Rosselló-Mora, 2006). For example, according to Peng et al. (2002) Sinorhizobium xinjiangensis and Sinorhizobium fredii had 39% DDH relatedness and they were considered as different species. Nonetheless, 74-89% similarity in DDH values was later reported for them by Martens et al. (2008). This result indicated that they were two different strains of the species, S. fredii. An additional main weakness of this technique is that it is not possible to build comparative database since the method gives non-cumulative relative DNA similarity values (Stackebrandt, 2003; Rosselló-Mora, 2006). The DDH technique needs also a large amount and high quality DNA, is technically challenging, time consuming and labor intensive. This technique is therefore limited to a few specialized laboratories and mostly applied if the bacteria under study are known to have closely related 16S rRNA gene sequences (Tindall et al., 2010). Though it has several limitations, the DDH technique is still considered as the gold standard in delineating bacteria at species level. However, the rapid progress of sequencing techniques and with the ever decreasing of its costs, most likely the DDH will be swapped by the more reproducible and accurate whole genome sequencing technique in the near future.

1. 8.3. Housekeeping protein coding gene sequences

Housekeeping protein coding genes are chromosomal in origin and are constitutively expressed in all cells under normal and patho-physiological conditions in order to maintain basic cellular functions of an organism. To overcome the limitations of rRNA genes and DNA–DNA hybridization techniques, the sequence analysis of multiple protein coding genes, known as multilocus sequence analysis (MLSA), has been recently considered as a preferred method to study closely related species and to discriminate strains of the same species. The housekeeping protein coding genes seem to be the most appropriate markers for phylogenetic analysis of bacteria. In general, they have higher level of sequence divergence compared to the 16S rRNA gene but are conserved enough to retain genetic information, and therefore their sequences show better discrimination than 16S rRNA gene sequences. The protein coding genes that are widely distributed, unique within a given genome, long enough and phylogenetically informative but short enough to be sequenced economically, that has acceptable accuracy in predicting the whole-genome relationships and that located separately in the chromosome of the genome are recommended in studying bacterial taxonomy (Zeigler, 2003).

A combination of two up to several housekeeping genes were extensively used in taxonomic studies of root nodulating bacteria, including in the phylogenetic analysis of Rhizobium spp., Sinorhizobium spp. and Mesorhizobium spp. (Degefu et al., 2012, 2011; Martens et al., 2008, 2007), and Bradyrhizobium spp. (Nzoué et al., 2009; Rivas et al., 2009; Steenkamp et al., 2008; Stepkowski et al., 2012, 2007, 2005; Vinuesa et al., 2008, 2005). Martens et al. (2008) analyzed ten housekeeping gene sequences (atpD, gap, gltA, pnp recA, gyrB, rpoB, thrC, glnA and dnaK) from several rhizobial species, mainly from genera of Ensifer but also including species from other genera of rhizobia. Based on this study, protein coding genes
recA, gyrB, glnA and thrC were found to be best markers for resolving the taxonomic position of rhizobial species. In other study, concatenated sequence analysis of three housekeeping genes; dnaK, recA and glnII were reported to give a tree topology which represented the species tree of the genus Bradyrhizobium (Nzoué et al., 2009). Well-resolved monophyletic Bradyrhizobium species groups were also obtained when concatenated sequences of atpD, glnII, recA and rpoB were analyzed (Vinuesa et al., 2008). Furthermore, Rivas et al. (2009) studied concatenated sequence phylogenies of atpD, recA, gyrB, rpoB and dnaK and found them useful for classification of Bradyrhizobium species. These results indicate that combinations of different housekeeping gene sequences seem to provide good phylogenetic classification in the case of Bradyrhizobium spp. Nevertheless, generally it was suggested that the set of best discriminating housekeeping gene markers used for bacterial phylogenetic analysis may vary, depending on the bacterial genus or species group (Nzoué et al., 2009).

Tree topologies and rhizobial groupings resulting from a combination of housekeeping gene sequences were compared with genomic relationship of the rhizobial species obtained from DNA-DNA hybridization data (Martens et al., 2008; Nzoué et al., 2009; Rivas et al., 2009). According to these studies, concatenated sequence phylogenies of 3 up to 8 genes found to accurately reflect the entire genomic similarities of the rhizobial species previously defined using DNA-DNA hybridization techniques. Thus, the use of MLSA was suggested to be used at least as a complement to DNA-DNA hybridization for species description of bacteria (Martens et al., 2008). Even a single best housekeeping gene could be enough to predict the entire genomic relatedness of the bacteria and to provide a robust species delineation, which is equivalent or even superior to DNA-DNA hybridization (Martens et al., 2008; Rivas et al., 2009; Zeigler, 2003). However, most of these studies also showed evidence of variable grouping of some strains depending on the gene type. This variation might be attributed to the differences in evolutionary histories between the genes or horizontal gene transfer as it sometimes affects the sequence structure of protein coding genes (Nichols, 2001; Rivas et al., 2009). For this reason, it is advisable to use more than one protein-coding gene for phylogenetic inference and bacterial identification purposes.

Although analyzing small sets of carefully selected protein coding gene sequences was recommended as an alternative to DNA-DNA hybridization when describing novel bacterial species (Stackebrandt et al., 2002, Zeigler, 2003), the DNA-DNA hybridization technique is still a benchmark technique. Therefore, the MLSA of different housekeeping genes are used along with DNA-DNA hybridization for describing a new bacterial species. For example, in a recent study three novel Mesorhizobium species, M. shonense, M. hawassense, and M. abyssinicae, were defined by analyzing concatenated sequences of recA, glnII, atpD and 16S rRNA genes and using DNA-DNA hybridization techniques (Degefu et al., 2013). In that study, each of the new rhizobial species were distinctly classified on the phylogenetic tree made with concatenated sequences in the same manner as their DNA-DNA relatedness.
AFLP is among the most widely used DNA fingerprinting methods since its publication (Vos et al., 1995) and has been used in several applications, such as in genetic diversity, phylogeny, and ecological studies of plants, animals, fungi and bacteria (Bensch and Kesson, 2005). AFLP fingerprints are produced by restriction of intact genomic DNA with restriction endonuclease and ligation of them with double strand oligonucleotide adapters. The restricted fragments are then selectively amplified and separated either using polyacrylamide gel electrophoresis or more commonly with capillary electrophoresis using automatic sequencer.

Usually one or two of the PCR primers need to be fluorescently labeled in order to produce labeled AFLP-fragments that can be detected and separated according to their size by automatic sequencer. In this case, once the fragment passes in the capillary electrophoresis of the sequencer, the primer that is labeled with fluorescent dye provides signal and the differently sized fragments are then converted to peaks on the computer software. The peaks are equivalent to the bands on silver stained polyacrylamide gel, which can also be employed to separate the AFLP-fragments based on their size in an electrical field. Compared to the polyacrylamide gel electrophoresis, the capillary electrophoresis provides more resolute fingerprint pictures and the data are better manageable. In addition, the silver staining in the polyacrylamide gel electrophoresis is time consuming and demanding. Therefore, the capillary electrophoresis has been used as the preferred method for separation of AFLP-fragments (Dresler-Nurmi et al., 2000; Rahman et al., 1998; Terefework et al., 2001).

With the AFLP technique, multilocus fingerprints representing the whole genome of an organism can be produced with a limited number of primer combinations. As a result, this method can be used to assess both core and accessory genes that are biologically and ecologically important for the organisms. This indicates that the technique helps in studying the species from ecological species points of view and to identify bacterial strains that may be adapted to the same ecological niches or geographic origins (Cohan, 2001). The great number of polymorphic bands that are produced with the AFLP method shows different genetic fingerprints and provide excellent power in resolving differences between very closely related strains compared to MLSA method (Boudon et al., 2005). In comparison to the RAPD fingerprinting method, AFLP was reported to have higher discriminatory power and to show better intraspecific genetic diversity between *Pseudomonas* species (Clerc et al., 1998). AFLP clustering was also found to present much more diversity between *Bradyrhizobium* species compared to 16S-23S rRNA intergenic sequence (ITS) and *recA* phylogenies, though all the three methods showed consistent grouping between the species (Gueye et al., 2009). Generally, the AFLP method has been comprehensively used in genetic diversity study of root nodule bacteria or to know if rhizobial strains in the same species are genetically identical or different (e.g. Li et al., 2012; Terefework et al., 2001; Wolde-meskel et al., 2004).

The AFLP technique was also applied in studying the taxonomic relationship of various groups of bacterial species (e.g. Doignon-Bourcier et al., 2000; Portier et al., 2006).
However, in this case several restriction enzymes might be needed in order to produce composite datasets for comparison and to determine a clear similarity or difference among the taxa (Terefework et al., 2001). It was reported to have similar discriminating power as DNA-DNA hybridization and therefore it was proposed to be used as an alternative to the DNA-DNA hybridization for identifying bacterial species (Portier et al., 2006; Willems et al., 2001). AFLP was used successfully to delineate the genomic species of the genus \textit{Agrobacterium} that were previously classified by DNA-DNA hybridization relatedness (Portier et al., 2006). Similarly, good agreement between the DNA-DNA hybridization and the AFLP method was reported in the taxonomic studies of \textit{Bradyrhizobium} species (Doignon-Bourcier et al., 2000; Gueye et al., 2009; Willems et al., 2001). Recently, the AFLP method was also applied in studying the phylogenetic structure and in identifying of the genospecies of the genus \textit{Frankia} (Bautista et al., 2011).

1.8.5. Whole genome sequencing

Complete genome sequencing is a molecular technique that involves the sequencing of the entire chromosomal DNA of an organism as well as DNA found in the mitochondria (for higher organisms, eukaryotic), chloroplast (for plants) or plasmids (for single cell organisms, prokaryotic) at a single time. Recently, due to several whole genomes sequencing efforts, a large numbers of bacterial genome are sequenced and available in public nucleotide databases such as GenBank (http://www.genomesonline.org/cgi-bin/GOLD/sequencing_status_distribution.cgi). So far (7 August 2013), 8767 prokaryote genomes have been completed and 9142 other sequencing projects are in progress. About 8390 of the completed sequences are obtained from the bacteria domain. This technology is also applied increasingly for nitrogen-fixing and related bacteria. At the time of writing this thesis (August 2013), the complete genome sequences of seven strains belonging to the genus \textit{Rhizobium} including three from \textit{R. leguminosarum} (WSM1325, 3841, WSM2304), two \textit{R. etli} (CFN42, CIAT 652), \textit{R. rhizogenes} K84 and \textit{R. tropici} CIAT899 were deposited in the GenBank database. The genome of nine strains in the genus \textit{Sinorhizobium} comprising seven \textit{S. meliloti} strains (1021, Ak83, BL225c, SM11, 2011, GR4, RM4), \textit{S. medicae} WSM419 and \textit{S. fredii} NGR234 have been sequenced for public use. The complete genome sequences of \textit{Mesorhizobium australicum} WSM2073, \textit{Mesorhizobium ciceri} WSM1271, \textit{Mesorhizobium loti} MAFF303099, \textit{Mesorhizobium opportunistum} WSM2075, \textit{B. japonicum} (USDA6, USDA110), \textit{Bradyrhizobium oligotrophicum} S58, \textit{Bradyrhizobium} sp. (WSM471, BTAi1, ORS 278, S23321), \textit{Azorhizobium cauliformans} ORS 571 and \textit{Methyllobacterium nodulans} ORS 2060 are also among the sequences deposited in the database (http://www.ncbi.nlm.nih.gov/genome/browse/).

Whole genome sequencing provides the complete genetic variation of the organisms and the sequence data can be used in comparative genomic studies for identification and taxonomic purpose. With sharp falling in the price of the technology, in the future many more bacteria including rhizobial complete genomes will be sequenced. Nevertheless, at the moment sequencing the whole genome is far from applicable in many laboratories since its price is still expensive and it needs skilled personnel to analyze the sequence data. Consequently, the current phylogenetic and diversity studies of bacteria including studies in this thesis are still
based on nucleotide sequences of various genes and genetic fingerprints though these markers contain limited molecular information.

1. 8.6. Accessory genetic elements: symbiotic genes

Accessory genomes are groups of genes, which are mainly located in bacterial secondary replicons known as plasmids and also in genomic islands of the chromosome. They are also called the flexible gene pool (Dobrindt et al., 2004). While core genes encode essential cellular functions of the bacteria, accessory genes code for extra (accessory) functions that are needed for bacteria to adapt various environmental conditions and ecological niches (Lan and Reeves, 2000). These are mobile genetic elements that can move between bacterial cells but can have long-term association within a bacterial species or group of species (Campbell, 1981; Lindström et al., 2010; Young et al., 2006). A continuous gain or loss of mobile genetic elements between different bacteria species is believed to be a common phenomenon. These processes are associated to several lateral gene transfer mechanisms, such as conjugation by which fragments of DNA or whole plasmids can transfer from a donor bacterium to the recipient, excision or integration of transposons into the chromosome, and transduction of phages or transformation of free DNA from one cell to another (Dobrindt et al., 2004). Consequently, accessory genomes are involved in spreading traits, such as virulence factors in pathogenic bacteria, antibiotic resistance and xenobiotic-degrading properties and traits enabling bacteria to form symbiotic relationship with higher organisms. Accessory genomes therefore enhance adaptive evolution in bacteria (Dobrindt et al., 2004).

In rhizobia, the symbiotic genes that are required for nodule formation (nod) and for nitrogen-fixation (fix and nif genes) in symbiosis with the host legume are located either on plasmids or symbiosis islands in the chromosomes of the bacteria. These genes, for example in species of Cupriavidus, Mesorhizobium amorphae, Rhizobium and Sinorhizobium are found in the symbiosis plasmids of the organisms (Galibert et al., 2001; Young et al., 2006; Wang et al., 1999). The symbiotic genes in Azorhizobium cauliodans, Bradyrhizobium spp. and Mesorhizobium loti are located on symbiotic islands in the chromosome flanked by insertion sequences (Amadou et al., 2008; Kaneko et al., 2011, 2002, 2000; Lee et al., 2008; Sullivan et al., 1995).

A transmittable 500 kb symbiosis island was originally identified in M. loti ICMP3 (Luck et al., 2001). Later larger symbiotic islands were also found on the chromosomes of B. japonicum strains USDA 110 (860 kb) and USDA 6 (695-kb) (Kaneko et al., 2011; 2002). As comparative genomic analysis results showed, most of the genes required for Nod-factor formation, nodulation and nitrogen fixation were highly conserved between the symbiotic islands of the strains USDA 110 and USDA 6, indicating that the symbiotic islands and the genes required for nodulation and symbiosis interaction might have long term relationships within B. japonicum species. In addition, the symbiotic islands of strains of genus Bradyrhizobium were found to contain large numbers of insertion sequences and genes required for island transfer, reflecting the transmittable nature of the symbiotic islands. Horizontal transfer of symbiotic genes was evident between inoculant B. japonicum and indigenous S. fredii and Bradyrhizobium elkanii (Barcellos et al., 2007). Rhizobium strain
IRBG74 and Ensifer sp. nodulating Sesbania sp. showed similarity in their nodA sequences (Cummings et al., 2009). Cupriavidus nodule bacterial species obtained from native legumes (Mimosa sp., Mimosa asperata, Mimosa strigillosa) in Texas and from legumes in Costa Rica displayed monophyletic nodA and nifH phylogenies that branched among Neotropical Burkholderia strains. As a result, Cupriavidus was suggested to acquire symbiotic genes by horizontal transfer from Burkholderia (Andam et al., 2007). Similarly, indigenous soil mesorhizobia found to have identical symbiosis island insertion sequences with M. ciceri sv. biserrulae strain WSM1497 inoculated on the pasture legume Biserrula pelecinus L (Nandasena et al., 2007). This and similar findings by Sullivan et al. (1995) also indicated that diverse non-symbiotic soil bacteria might be evolved into symbiotic following the lateral transfer of symbiotic genes from symbiotic to non-symbiotic bacteria in the field.

In contrast to the conserved core genes of the chromosome that have similar organization in closely related species, the organization and composition of the accessory genome can even vary among related species. Based on the genomic sequence analysis of R. leguminosarum strain 3841, Young et al., (2006) revealed that the accessory genomes are sporadic in distribution and have also low G+C contents. In addition to the chromosome, the genome of S. meliloti strain Rm1021 has two replicons termed pSymA and pSymB. The symbiotic genes are located on mega plasmid pSymA (Galibert et al., 2001). The genomic contents of the plasmids are varied among S. meliloti strains. Some of the symbiotic genes located on pSymA in strain Rm1021 were found to be absent in the accessory genomes of strains AK83 and BL225C. The strain Rm1021 was also lacking some symbiosis related genes found on the accessory genome of the strains AK83 and BL225C. In return, the genes required for bacterial microaerophilic adaptation inside nodules were absent in the symbiotic plasmid of strain AK83. A number of genes, which are necessary to enhance bacterial competition, plant invasion and growth of differentiated bacteroids inside nodules, were recognized in AK83 and BL225C strains but not in strain Rm1021. The variation in genomic content of the plasmids caused phenotypic divergences between strains. Thus, strain AK83 induced many immature nodules or showed low plant growth promoting efficiency, whereas nodule formation was delayed when alfalfa plant was inoculated with strain BL225C (Amadou et al., 2008; Galardini et al., 2011).

In line with the organizational and gene content differences that are shown between different accessory genomes of closely related bacterial species, the symbiotic genes sequences can also vary between closely related rhizobial species or strains. The phylogeny of nodulation and nitrogen–fixing genes do not usually follow the phylogeny of housekeeping genes that show the taxonomy of the rhizobia (Cummings et al., 2009; Degefu et al., 2011; Menna and Hungria, 2011; Moulin et al., 2004; Muñoz et al., 2011; Stepkowski et al., 2012, 2007, 2005; Suominen et al., 2001). Mostly, the phylogeny of the symbiotic gene sequences (the nod genes) of the rhizobia follows the phylogeny of the host plants or geographic origin of the hosts. Accordingly, different rhizobial species nodulating the same host plant usually show high similarity in their nodulation genes. Taxonomically closely related species can show differences in their symbiotic genes if they nodulate different host plants.
1.8.7. Genetic markers and methods used in this thesis

- Partial 16S rRNA gene sequence for preliminary bacterial identification to the genus level
- MLSA of protein coding housekeeping gene (recA, rpoB, glnII) sequences for detailed bacterial phylogenetic analysis at species level
- AFLP technique to study intra-species diversity of the bacteria
- The nodulation (nodA, nodC, nodY/K) and nitrogen-fixation (nifH) gene sequences to study diversity of the rhizobia respect to their host plant

1.9. Rationale of this study

Ethiopian economy is highly dependent on agriculture, providing 85% the country’s employment and 90% of its export earnings (CSA, 2005; World Bank, 2010). However, the agriculture and food production are predominantly practiced by smallholder farmers. Mostly, this sector has been affected by recurrent loses of soil fertility due to deforestation, land degradation and erosion. Ethiopia is described as one of the sub-Saharan countries with the highest rate of soil nutrient depletion. For example, the national reduction rate in the content of soil nitrogen was reported to be on average 122 kg N ha\(^{-1}\) year\(^{-1}\) (Haileslassie et al., 2005). The farmlands are deficient in soil fertility mainly in nitrogen content of the soils (Beyene and Tsige, 1986; Katungi et al., 2011; World Bank, 2006). This fact compounded with limited access to agricultural inputs, such as fertilizers and improved production technologies, leads to low crop productivity of the farmlands in the country (Elias, 2002; Mekonnen et al., 2012).

The smallholder farmers are usually unable to amend the nutrient contents of their farmlands due to either shortage or unaffordable price of the chemical fertilizers (Katungi et al., 2011; World Bank, 2006). Thus, cheaper and environmentally friendly alternatives, such as BNF in farming systems is essential to mitigate the fertility problems of the farmlands and to tackle ecological problems that are associated with deforestation and land degradation. Nitrogen-fixing leguminous trees and shrubs, such as *E. brucei*, *Crotalaria* spp. and *Indigofera* spp., are found to be versatile, used as forage for cattle or can be applied in different agroforestry systems, such as in intercrop (alley cropping) or fallow crops in low-input agriculture. Thereby they restore degraded farmlands and soil fertility by increasing content of nitrogen and organic carbon in the soil (Diabate et al., 2005). The usefulness of legumes could be boosted by inoculating them with effective nitrogen-fixing rhizobia. The yield of food legumes such as common bean and soybean can also partly be improved through the use of efficient rhizobial inoculants. However, rhizobial inoculants which can be applied in field in order to boost the nitrogen fixation ability of the legumes are also lacking. To our knowledge there was so far not detailed information about the indigenous rhizobia nodulating *E. brucei*, *Crotalaria* spp. and *Indigofera* spp., common bean and soybean, although a few publications indicated the presence of a high diversity of rhizobia that nodulating other woody and herbaceous legumes of Ethiopia (Degefu et al. 2012, 2011; Wolde-meskel et al. 2005, 2004). This thesis therefore specifically deals with the characterization of rhizobia isolated from the woody and shrub legumes *E. brucei*, *Crotalaria* spp. and *Indigofera* spp., and the food legumes common bean and soybean growing in Ethiopia.
2. OBJECTIVES OF THE STUDY

The main objective of the study was to isolate and explore genetic diversity, phylogeny, plant growth promoting activities and nodulation capability of bacterial strains obtained from root nodules of the indigenous woody and shrub legumes *E. brucei*, *Crotalaria spp.*, *Indigofera* spp. and the food legumes common bean and soybean growing in Ethiopia. The information gained from this study benefits our future research for improved production of inoculum (biofertilizer) that can be applied in the field conditions (1), in order to boost the yields of common bean and soybean and (2) to reinforce seedling establishment of agroforestry legumes *E. brucei*, *Crotalaria* spp. and *Indigofera* spp. in order to restore the degraded farmlands and to develop sustainable ecosystem services in the country.

Specific objectives

1. To isolate bacterial strains from root nodules of *E. brucei*, *Crotalaria* spp. and *Indigofera* spp., common bean and soybean (Paper I, II, III).

2. To clarify the genetic diversity and taxonomic position of *Rhizobium* strains nodulating common bean (Paper I) and *Bradyrhizobium* strains isolated from root nodules of *E. brucei*, *Crotalaria* spp. and *Indigofera* spp., and soybean (Paper II) using AFLP fingerprints analysis and MLSA of housekeeping genes.

3. To study the genetic diversity of sporadic symbionts and endophytic bacterial strains isolated from root nodules of *E. brucei*, *Crotalaria* spp. and *Indigofera* spp., soybeans and common beans (Paper III).

4. To study the nodulation capacity and phylogeny of symbiotic genes (*nod* and *nif* genes) of the rhizobial strains (Paper I, II, III).

5. To determine the plant growth promoting properties of endophytic bacteria and sporadic rhizobial strains (Paper III).

6. To evaluate nodule colonization ability of endophytic bacterial strains by co-inoculating them with nodule forming and nitrogen-fixing rhizobia (Paper III).
3. MATERIAL AND METHODS

List of bacterial strains used in this study and their geographic origins are presented in each paper (Paper I-III). Bacterial isolation techniques, nodulation test, plant growth promoting (PGP) activities, molecular genetic techniques and the methods used to analyze AFLP fingerprints and gene sequences are summarized in Table 1 and details of these methods are also presented in each of the individual papers (Paper I-III).

Table 1 Methods used in this study, the specific papers where they appear and the references

<table>
<thead>
<tr>
<th>Method</th>
<th>Important references</th>
<th>paper/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacterial isolation and cultivation</td>
<td>Lindström et al., 1985; Somasegaran and Hoben, 1994; Vincent, 1970</td>
<td>I,II,III</td>
</tr>
<tr>
<td>Nodulation tests</td>
<td>Räisänen et al., 2001; Somasegaran and Hoben, 1994; Tauro et al., 2009; Vincent, 1970</td>
<td>I,II,III</td>
</tr>
<tr>
<td>DNA extraction</td>
<td>Li et al., 2012</td>
<td>I,II,III</td>
</tr>
<tr>
<td>AFLP-fingerprinting</td>
<td>Li et al., 2012; Terefework et al., 2001; Vos et al., 1995</td>
<td>I,II,III</td>
</tr>
<tr>
<td>Genes sequenced and primers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16S rRNA gene</td>
<td>Weisburg et al., 1991</td>
<td>I,II,II</td>
</tr>
<tr>
<td>recA</td>
<td>Stepkowski et al., 2005</td>
<td>II</td>
</tr>
<tr>
<td></td>
<td>Gaunt et al., 2001</td>
<td>I,III</td>
</tr>
<tr>
<td>rpoB</td>
<td>Martens et al., 2008; Vinuesa et al., 2008</td>
<td>I, II, III</td>
</tr>
<tr>
<td>glnII</td>
<td>Stepkowski et al., 2005</td>
<td>II</td>
</tr>
<tr>
<td></td>
<td>Turnier and Young, 2000; Vinuesa et al., 2005</td>
<td>I,II,III</td>
</tr>
<tr>
<td>nodA, nodY/K</td>
<td>Stepkowski et al., 2005</td>
<td>II</td>
</tr>
<tr>
<td>nodC</td>
<td>Sarita et al., 2005</td>
<td>I, III</td>
</tr>
<tr>
<td>nifH</td>
<td>Vinuesa et al., 2005</td>
<td>II</td>
</tr>
<tr>
<td></td>
<td>Rivas et al., 2002</td>
<td>I, III</td>
</tr>
<tr>
<td>AFLP-fingerprint analysis</td>
<td></td>
<td>I, II, III</td>
</tr>
<tr>
<td>Phylogenetic analysis of the gene sequences</td>
<td>Dereeper et al., 2010, 2008; Kimura, 1980; Huelsenbeck and Ronquist, 2001; Posada, 2008; Tamura et al., 2011</td>
<td>I, II, III</td>
</tr>
<tr>
<td>Antibiotic resistance test</td>
<td>Räisänen et al., 2001</td>
<td>III</td>
</tr>
<tr>
<td>Enzyme β-glucuronidase (GUS) activity test</td>
<td>Pitkäjärvi et al., 2003; Räisänen et al., 2001</td>
<td>III</td>
</tr>
<tr>
<td>Preparation of GUS-transposon marker gene and transformation</td>
<td>Wilson et al., 1995</td>
<td>III</td>
</tr>
<tr>
<td>Electro-transformation of plasmid DNA</td>
<td></td>
<td>III</td>
</tr>
<tr>
<td>Carrying GUS marked gene to recipients</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td><strong>Determining PGP activities</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IAA production</td>
<td>Egamberdieva and Kucharova, 2009</td>
<td></td>
</tr>
<tr>
<td>Protease activity</td>
<td>Brown and Foster, 1970; Li et al., 2012</td>
<td></td>
</tr>
<tr>
<td>Lipase activity</td>
<td>Salkinoja-Salonen et al., 1999; Sierra, 1957</td>
<td></td>
</tr>
<tr>
<td>Cellulase activity</td>
<td>Ruijssenaars and Hartmans, 200; Wenzel et al., 2002</td>
<td></td>
</tr>
<tr>
<td>Mineral phosphate solubilization</td>
<td>Pikovskaya, 1948; Srividya et al., 2009</td>
<td></td>
</tr>
<tr>
<td>Production of siderophore</td>
<td>Alexander and Zuberer, 1991</td>
<td></td>
</tr>
</tbody>
</table>
4. MAIN RESULTS AND DISCUSSIONS

4.1. Bacterial identification

In total, 143 bacterial isolates were obtained from surface-sterilized root nodules of woody, shrub and food legumes collected in Ethiopia. Based on the 16S rRNA gene sequence similarity with the nucleotide database references, the majority of isolates obtained from the nodules of *P. vulgaris* (common bean) were identified as *Rhizobium* (27) and *Agrobacterium* (5) (Paper I). Fifty-six isolates obtained from nodules of *E. brucei*, *Crotalaria* spp. and *Indigofera* spp. and *G. max* (soybean) were found to represent the genus *Bradyrhizobium* (Paper II). In addition to slow-growing *Bradyrhizobium* spp., 13 fast-growing strains representing *Rhizobium*, *Mesorhizobium* or *Agrobacterium* were also isolated from the nodules of *E. brucei*, *Crotalaria* spp. and *Indigofera* spp. The rest of the 42 isolates obtained from root nodules of woody, shrub and food legumes were belonged to diverse non-rhizobial Gram-negative and Gram-positive bacteria (Paper III).

The 16S rRNA gene sequence has been used as a standard marker in identification and classification of prokaryotes (Harris and Hartley, 2003; Garrity et al., 2001). The presence of large 16S rRNA gene sequence deposits in the Genbank database facilitates comparison of test strain sequences with different reference species (Woese et al., 1990; Weisburg et al., 1991). Conversely, the use of 16S rRNA gene as the only genetic marker is not recommended for detailed phylogenetic studies due to its presence as multiple copies in a genome of some bacteria (Haukka et al., 1996), vulnerability to genetic recombination and horizontal gene transfer (van Berkum et al., 2003) and low divergence among strains belong to the same species/genus (Vinuesa et al., 2005; Willems et al., 2003). In this study, partial 16S rRNA gene sequencing helped us to identify the bacterial strains at the genus level. Nonetheless, it was highly conserved among the bacterial isolates and it was not possible to identify them to species level although the strains represented various bacterial species according to the later MLSA (Paper I, II and III). This phenomenon was particularly emphasized among species belonging to the genus *Bradyrhizobium* (Paper II).

4.2. Phylogeny and diversity of the test bacterial strains: MLSA and AFLP-fingerprinting analyses

For detailed phylogenetic studies we used MLSA of protein coding genes that has been considered as a preferred method to study closely related species and strains in the same species (Martens et al., 2008). Thus, in addition to the 16S rRNA gene, housekeeping genes (*recA*, *glnII* and *rpoB*) were sequenced for all *Bradyrhizobium*, *Rhizobium* and *Agrobacterium* strains and aligned with corresponding gene sequences of appropriate type and reference strains retrieved from the nucleotide databases. Subsequently, single genes and combined phylogenetic trees were constructed from sequence alignments. Generally, the strains were grouped consistently in each phylogenetic gene tree and only a few exceptional test and reference strains were grouped aberrantly depending on the housekeeping gene considered (Paper I, Figs.S1A-C; Paper II, Figs.S1A-C; Paper III, Figs.S1-S3). The differences in the grouping might be due to the different evolutionary rates of single genes.
between different bacterial taxa (Nichols, 2001; Rivas et al., 2009). With a few exceptions, the combined recA-glnII-rpoB-16S rRNA gene sequences analysis gave consistent grouping with single gene sequences but the combined gene sequences dataset produced more robust phylogenetic trees with the majority of the branches supported by higher bootstrap and/or likelihood ratio than in the single gene trees. Consequently, based on MLSA of housekeeping genes Rhizobium and Agrobacterium obtained from nodules of common bean were differentiated into eight phylogenetic groups (Paper I, Fig.2 and Figs.S1A-C). The majority of E. brucei, I. arrecta, Crotalaria spp. and soybean nodulating strains were distributed into 15 diverse Bradyrhizobium phylogenetic groups (Paper II, Fig.1.). Besides, minor Rhizobium, Mesorhizobium and Agrobacterium strains obtained from E. brucei, I. arrecta and Crotalaria spp. were classified into six phylogenetic groups (Paper III, Fig.2.). Detailed phylogenetic information about the bacterial species is given in sections 4.2.1, 4.2.2 and 4.2.3.

The genetic diversity of all test bacterial strains was studied using the AFLP fingerprinting technique, since it has been proven the best method to show genomic variation among closely related strains (Boudon et al., 2005; Li et al., 2012). More heterogeneity was observed among closely related strains based on AFLP clustering compared with the phylogenetic grouping based on concatenated core genes sequences. The majority of the test strains produced distinct AFLP fingerprints (Paper I, Fig.1; Paper II, Fig.2; Paper III, Fig.1.) indicating high diversity among strains belonging to the same species. The AFLP-fingerprinting was also proposed as an alternative taxonomic tool for bacterial classification at species level (Doignon-Bourcier et al., 2000; Huys et al., 1996; Jensen et al., 1993). Congruence between AFLP clustering and ITS and recA phylogenies of Bradyrhizobium species was reported by Gueye et al. (2009). In this thesis, usually strains found in the same phylogenetic groups obtained through MLSA, formed also the same AFLP clusters. Therefore, in addition to its role in genetic diversity study, AFLP fingerprint analysis was also supporting the phylogenetic positions of the strains that resulted from MLSA. However, the groupings in some AFLP clusters did not reflect fully the phylogenetic relationship of the strains based on concatenated sequences (Paper I, Fig.1 and Fig.2; Paper II, Fig.1 and Fig.2.). This case was in agreement with the study by Terefework, et al. (2001), which showed incongruence between AFLP clustering and phylogenetic relationships of Rhizobium species.

4.2.1. The Rhizobium leguminosarum complex: True symbionts of common bean in Ethiopia

According to Li et al. (2012) frequently isolated rhizobial species, which form effective nitrogen-fixing nodules on a specific legume species, were called true symbionts, whereas rarely occurring effective or ineffective nodule forming rhizobia species obtained among the true nodulating rhizobia species from a specific legume nodules were called sporadic symbionts. Bacterial strains related to non-symbiotic Gram negative bacteria including Agrobacterium species and Gram-positive bacteria were called nodule endophytes. With similar analogy combined with nodulation test we could classify the bean nodulating rhizobia as true and sporadic symbionts (Paper I). The results in Paper I showed that the majority of the Ethiopian bean rhizobia (25 strains) belonged to four phylogenetic species
groups *R. leguminosarum*, *R. phaseoli*, *R. etli* and novel *Rhizobium* genospecies. All these were clustered in one large monophyletic clade in the phylogenetic tree (Paper I, Fig.2). For clarity, this clade was named as “the *R. leguminosarum* species complex”. All the rhizobial strains from this study included in the complex formed effective symbioses with common bean plants, indicating that they are true symbionts of common bean in Ethiopia.

*R. etli* has been reported as one of the main bean-nodulating rhizobia in the center of the host origin of diversity, in Latin America (Segovia et al., 1993). In east Africa, Kenyan soils harbored bean-nodulating *R. leguminosarum*, *R. etli* and *R. tropici* (Amann et al., 1992). West African soils were also reported to contain bean nodulating *R. etli* and *R. tropici* (Diouf et al., 2000). *R. leguminosarum* sv. *phaseoli* was first identified in Europe and it was suggested that the symbiotic plasmid may have been transferred to native *R. leguminosarum* from introduced seeds containing *R. etli* sv. *phaseoli* from America (Pérez-Ramírez et al., 1998; Segovia et al., 1993). Therefore, the presence of *R. etli* in Ethiopian soil may indicate the introduction of *R. etli* together with beans to soils of Africa from America (Diouf et al., 2000). Ethiopia has been proposed to be one of the centers of diversity of field pea (Hailu et al., 1991) and therefore, the soils of Ethiopia believed to harbor symbionts of pea, *R. leguminosarum* sv. *viciae* (Beyene et al., 2004). This fact suggests that these native rhizobia may have also acquired molecular and functional characteristics of *R. phaseoli* from the introduced *R. etli* sv. *phaseoli*.

*R. leucaenae* is known as a common bean-nodulating rhizobial species in South America (Ribeiro et al., 2009; Ribeiro et al., 2012). *Rhizobium gardinii* was originally obtained in France in Europe (Amarger et al., 1997) and also inhabits soils of Latin America (Martínez-Romero, 2003). In this study, *R. leucaenae* and *R. gardinii* were represented only by a single strain each (Paper I, Fig.2 and Table1). Thus, these were considered as sporadic symbionts of beans in Ethiopia.

4.2.2. Ethiopian soils harbor phylogenetically diverse groups of *Bradyrhizobium*

The combined recA-glnII-rpoB-16S rRNA sequences phylogeny presented in paper II, Fig.1 was in agreement with previously published MLSA phylogenies of the genus *Bradyrhizobium* (Rivas et al., 2009; Steenkamp et al., 2008; Vinuesa et al., 2008). Together with our strains, the type and reference strains formed well-supported phylogenetic groups. Consequently, the combined sequences analyses showed that *E. brucei*, *I. arrecta* and *Crotalaria* spp. and soybean nodulating *Bradyrhizobium* strains in Ethiopian are phylogenetically diverse, distributed into fifteen genospecies under *B. japonicum* (I) and *B. elkanii* (II) super lineages. The phylogenetic groups were also with a few exceptions corresponding to the AFLP clusters as presented in Paper II, Fig.2. Thus, the AFLP clustering supports generally the distinctness of the *Bradyrhizobium* lineages in each phylogenetic group.

Seven *Crotalaria incana* and *Indigofera* spp. nodulating strains in groups I.1 were classified as *Bradyrhizobium yuanmingense*. Soybean nodulating strains in group I.12 were *B. japonicum* type I. Ethiopian strains in groups II.1 (obtained from *I. arrecta* and soybean) and
group II.2 (obtained from soybean) were classified as two phylogenetically different \textit{B. elkanii} groups (Paper II, Fig. 1). Based on single gene phylogenies (Paper II, Figs. ISA-C) the majority, nine groups in super clade I, were related to different unnamed genospecies or lineages of the genus \textit{Bradyrhizobium} isolated from tropical soils in Brazil, Mexico or South Africa (Menna et al., 2009; Ormeño-Orrillo et al., 2012; Steenkamp et al., 2008). However, eight strains obtained from nodules of \textit{E. brucei}, \textit{Crotalaria} spp. and \textit{Indigofera} spp. (group I.3) and a single strain isolated from \textit{C. incana} (group I.6) were identified as new, Ethiopian \textit{Bradyrhizobium} genospecies ETH1 and ETH2, respectively. Therefore, further study is needed to clarify the species boundary of Ethiopian genospecies and \textit{Bradyrhizobium} strains related to unnamed genospecies.

In Ethiopia, soybean cultivation is relatively new and therefore rhizobia nodulating soybeans were assumed to be missing from the soils of Ethiopia. The discovery of twelve native soybean symbionts (eight strains representing \textit{B. elkanii}, two strains representing \textit{B. japonicum} type I and \textit{Bradyrhizobium} sp. strains SBR6 and SBR1B), in the soils of Ethiopia was an interesting aspect of Paper II. All of these were as effective symbionts as inoculant strain TAL 379 on local soybean cultivar in the growth chamber. These strains might be ecologically well adapted to local conditions and could be good potential inoculants for soybean in the country. The \textit{Bradyrhizobium} strains indigenous to African soils were reported to induce effective nodules on promiscuous soybean cultivar Tropical \textit{Glycine} cross (TGx) (Abaidoo et al., 2000). Three of the Ethiopian soybean-nodulating strains were taxonomically closely related to other Ethiopian \textit{Bradyrhizobium} strains isolated from \textit{C. incana} or \textit{I. arrecta} (Paper II, Fig.1 group I.4 and II.1), suggesting that soybean might be nodulated promiscuously by indigenous symbionts of other legume species.

4.2.3. Diverse non-symbiotic endophytic bacteria obtained from nodules of woody, shrub and food legumes

In paper III, the majority of non-nodulating bacteria isolated from woody, shrub and food legumes including non-symbiotic \textit{Rhizobium} and \textit{Agrobacterium} species obtained from nodules of \textit{E. brucei} and \textit{Crotalaria} spp. were dealt with as endophytic bacteria. Only few representatives of \textit{Rhizobium} and \textit{Mesorhizobium} obtained from nodules of \textit{E. brucei} and \textit{I. arrecta} were sporadic symbionts. The endophytic bacterial strains were identified as diverse genera of Gram-negative and Gram-positive bacteria (Paper III, Table 1). Generally, the distribution pattern of the nodule endophytic bacteria was related to the host plants, for instance most of the \textit{Enterobacter} strains were obtained from \textit{Crotalaria} spp. and common bean (Paper III, Table 1) indicating that the host plant may dictate the nodule endophytic bacterial community as is the case for rhizobia.

Each endophytic bacterial strain produced a distinct AFLP genetic fingerprint indicating that strains within a species were distinct (Paper III, Fig.1 and Table 1). Based on MLSA results, the sporadic symbionts were classified as \textit{R. phaseoli} strains ERR16 and ERR17, \textit{Rhizobium} sp. IAR30 and \textit{Mesorhizobium} sp. ERR6. Among endophytic bacteria included in the phylogenetic tree (Fig. 2), five strains in group III were identified as putative new \textit{Rhizobium} species and four as \textit{Agrobacterium} species, \textit{A. radiobacter} strains CIR5 and ERR18 and \textit{Agrobacterium} sp. strains ERR15 and ERR8. (Paper III, Fig.2).
Endophytic bacteria such as *Agrobacterium* (Liu et al., 2010), *Bacillus* (Bai et al., 2003), and *Enterobacter* and *Pseudomonas* species (Ibáñez et al., 2009) were reported to coexist with nodulating rhizobial species in the same legume nodules. In this study, nine non-symbiotic strains representing seven genera: *Agrobacterium* sp. CIR5, ERR15 and ERR18, *Burkholderia* sp. CIR2, *Paenibacillus* sp. IAR22, *Pantoea* sp. HBR8, *Pseudomonas* sp. HBR44, *Rhizobium* sp. CSR8B and *Serratia* sp. HBR15 were re-isolated from nodules of plants after co-inoculation with symbiotic bacteria, confirming their ability to enter and coexist with symbiotic bacteria in the same nodule (*Paper III, Table 3*). The non-nodulating *Rhizobium* species can be found in the rhizosphere of legumes and sometimes they were found to nodulate legume plants (Segovia et al., 1991). These types of *Rhizobium* strains were isolated from nodules of soybean (Wu et al., 2011). In this study, strains classified as novel *Rhizobium* species in paper III could not form nodules on their original host. Therefore, these strains were also considered as endophytic bacteria, as their representative strains were re-isolated from the nodules of co-inoculated plants.

### 4.3. Plant growth promoting activities of the sporadic symbionts and endophytic bacteria

Endophytic bacterial strains have the ability to promote plant growth, although the mechanisms relating to this phenomenon are not yet well understood. Several different activities or traits of the bacterial strain have been thought to be involved in the promotion of plant growth, such as production of plant growth hormone IAA, siderophores, plant cell wall degrading cellulase and fungal cell wall degrading lipase and protease as well as ability to solubilize phosphate (Kuklinsky-Sobral et al., 2004; Li et al., 2008; Mishra et al., 2009; Spaepen et al., 2007; Rodriguez et al., 2006). Most of the endophytic and nodulating rhizobial strains presented in *Paper III* showed two or more of the plant growth promoting properties. Most bacterial strains (44 out of 54 tested) showed cellulase enzyme activity and IAA production was the second most common PGP activity among the tested strains, from which 42 were capable of producing IAA (*Paper III, Table 1*). These properties might be used by endophytic bacteria to penetrate the plant defense mechanisms and to create localized lesions in legume cell walls (Sindhu and Dadarwal, 2001; Spaepen et al., 2007). By these mechanisms endophytic bacteria may enhance the infection process of symbiotic bacteria and nodulation on legumes (Sindhu and Dadarwal, 2001). The phosphate solubilizing bacteria might increase the plant phosphorus uptake by mineralizing the insoluble phosphates in the soil (Rodriguez et al., 2006). The siderophore producing bacteria compete for mineral sources by chelating iron and other essential metals and thereby they can suppress growth of plant pathogens (Loper and Henkels, 1997). In this study, many of the endophytic bacterial strains showed phosphate solubilization and siderophore production ability, indicating their potential role in promotion of plant growth while colonizing the nodules or rhizospheres of the plants.
4.4. Phylogeny of symbiotic genes

4.4.1. Genes *nodC* and *nifH* of Ethiopian *Rhizobium* species

The majority of Ethiopian test strains and reference *Rhizobium* species nodulating common bean were closely related with respect to their *nifH* and *nodC* gene sequences. The *nodC* gene sequences among Ethiopian strains were slightly more divergent than their *nifH* gene sequences. Most Ethiopian test strains and reference *Rhizobium* species that nodulate common bean were clustered into two closely linked *nodC* clades, Clade I and Clade II (99% bootstrap value) (Paper I, Fig. 3). Two *R. phaseoli* strains (ERR16, ERR17) that were isolated from the nodules of the leguminous tree *E. brucei* had 100% *nodC* sequences similarity with the type strains *R. phaseoli* ATCC 14482 and *R. etli* CFN 42 (Paper III). In the *nifH* phylogenetic tree, most Ethiopian bean nodulating test strains and type strains of reference species were placed into one well-supported major clade (93% bootstrap value) (Paper I). The close relatedness of *nodC* or *nifH* gene among bean and *E. brucei* nodulating rhizobia may be explained by horizontal transfer and a common evolutionary history of *nifH* or *nodC* gene among these bacterial species (Amarger et al., 1997; Laguerre et al., 2001).

Li et al. (2008) showed the presence of high similarity in *nifH* gene sequence between endophytic *Bacillus* and symbiotic *Bradyrhizobium*. Similarly, in our study the nodulating and non-nodulating, endophytic *Rhizobium* species and two *Agrobacterium* strains (ERR8 and CIR5) obtained from nodules of *Crotalaria* spp. and *E. brucei* were found to have identical *nifH* gene sequences (Paper III, Fig.3) with the Ethiopian true symbionts of bean presented in Paper I. This might also indicated the horizontal transfer of *nifH* gene between symbiotic and non-symbiotic, nodule endophytic bacteria. Conjugal plasmid exchange was reported to happen between *R. meliloti* strains inside alfalfa nodules (Pretorius-Güth et al., 1990), reflecting that the contact between bacteria inside the nodules may facilitate conditions for horizontal gene transfer between them. As a result, the nodule non-symbiotic bacteria may also switch role and evolve into novel microsymbionts (Li et al., 2008; Pretorius-Güth et al., 1990).

The former *R. tropici* type A (now *R. leucaenae*, Ribeiro et al., 2012), which can induce nitrogen-fixing nodules on common bean, were reported to have different *nod* genes than other more common bean nodulating *Rhizobium* species (Poupot, et al., 1995, 1993). Similarly, in this study the sporadic *R. leucaenae* HBR12 and its closely related reference strains were different from most common bean symbionts with respect to their *nifH* and *nodC* gene sequences. The *nifH* gene sequence of the other sporadic *R. giardinii* HBR21 was closely related to the major common bean nodulating rhizobia, indicating that this strain may belong to sv. *phaseoli*. However its *nodC* was clustered tightly with the type strain of *R. giardinii* sv. *giardinii* which is contradictory to the result showing that each of the sv. *phaseoli* and *giardinii* formed distinct lineages on the *nodC* tree (Laguerre et al., 2001) (Paper I, Fig.3 and Fig.4).
4.4.2. *Bradyrhizobium nodA* phylogeny: Implications on geographic origin or host legumes

The phylogeny of *nodA* gene in the genus *Bradyrhizobium* has been found to follow the geographic origin and/or host range of the bradyrhizobial strains (Moulin et al., 2004; Muñoz et al., 2011; Stepkowski et al., 2012, 2007, 2005). Accordingly, ten monophyletic *nodA* clades (I-X) were described for *Bradyrhizobium* spp. The *nodA* gene clades I, IV and X were found to be formed by *Bradyrhizobium* symbionts of legumes native to Australia. The *nodA* clade II belonged to Genisteae and *Serradella* symbionts of European and Mediterranean origin and *nodA* clades V, VII and VIII were designed for symbionts isolated from native legumes of the subtropical and tropical parts of the America. Clade VI included photosynthetic *Bradyrhizobium* species. Clade III was highly diversified and therefore further classified into sub-clades III.1, III.2 and III.3, and the sub-clades III.1 and III.3 were dominant among strains isolated from sub-Saharan African countries (Moulin et al., 2004; Steenkamp et al., 2008; Stepkowski et al., 2012, 2007, 2005). In this study, all Ethiopian *Bradyrhizobium* strains were clustered in *nodA* Clade III.3 along with strains originating from tropical and sub-tropical regions (Paper II, Fig.3). This finding is in agreement with the previous work presented by Steenkamp et al. (2008) who showed that *nodA* sequences of the *Bradyrhizobium* strains originating from the sub-Saharan African countries Botswana, South Africa, Guinea, Zimbabwe and Kenya belonged to clade III.3 as well.

In clade III.3, Ethiopian *Bradyrhizobium* species were further classified into 14 distinct *nodA* phylogenetic sub-groups (Paper II, Fig.3 groups A-N). These groups were mostly formed according to host clades or origin of the strains, for instance, *nodA* groups G, I, L, M and N comprised only symbionts of Milletoids and sub-groups I, G, M and N specifically included symbionts of soybean (*G. max*). However, both sub-clades I and M contained a single *nodA* sequence of Ethiopian origin, suggesting the need for specific *nodA* genes by different *Bradyrhizobium* strains that nodulate soybean. On the contrary, the *nodA* sub-groups A-D were intermixed with strains obtained from *Crotalaria* sp., *Indigofera* sp. and/or *E. brucei*, but included either only Ethiopian *nodA* sequences or grouped with those originating from other African countries. In conclusion, our findings support the previous results which showed the presence of a relationship between the *nodA* phylogeny of the *Bradyrhizobium* strains and their host range and/or the geographic origin of the hosts.

### 4.4.3. Phylogenetic congruence of *Bradyrhizobium nodA, nodY/K* and *nifH* gene sequences: Inference for the monophyletic origin of the symbiotic genes

The nodulation (*nod*) and nitrogen fixation (*nif*) genes of *B. japonicum* USDA 110 and USDA 6 are arranged similarly on symbiotic islands on the chromosome (Kaneko et al., 2011). In previous studies the phylogeny of nodulation genes *nodA, noeI* and *nodZ* were found to be congruent in *Bradyrhizobium* spp. (Moulin et al., 2004; Stepkowski et al., 2007, 2005). According to Menna and Hungria (2011) the tree built with *Bradyrhizobium* nitrogenase gene (*nifH*) was also found to be consistent with trees built with the nodulation genes *nodY/K, nodA* and *nodZ*. In Paper II, Ethiopian *Bradyrhizobium* test strains as well as the references were consistently grouped respect to their *nodA* (Fig. 3), *nodY/K* (Fig.4) and *nifH* (Fig.5) gene phylogenies. The *nodY/K* and *nodA* phylogenies were highly
congruent and all the 14 sub-groups that were formed in the *nodA* tree were also found in the *nodY/K* tree (Paper II, Table 1). Generally, the *nifH* gene sequences also produced a tree consistent with the tree built with *nodA* sequences, except for a few grouping differences (Paper II, Table 1). From 13 *nifH* clusters, 11 were the same as in the *nodA* and *nodY/K* trees, though strains in *nifH* groups A, B, D, E, F, I, and M were clustered without any reference strains due to the lack of *nifH* sequences in nucleotide databases at the time of this study for certain reference strains that were included in the *nodA* tree (Paper II, supplementary Table S3). The phylogenetic congruence of *nodA*, *nodK/Y* and *nifH* in this study together with the previous findings (Moulin et al., 2004; Stepkowski et al., 2007, 2005, Menna and Hungria, 2011) suggest a monophyletic origin of the symbiotic genes in the *Bradyrhizobium*.

4.5. Comparative analysis of the phylogeny based on symbiotic and housekeeping genes

The symbiotic genes of root-nodule bacteria are located on transmissible plasmids or on chromosomal islands. The evolution of symbiotic genes is thought to take place mainly via lateral gene transfer. Vertical gene transfer mechanisms with frequent homologous gene exchange are the main means for exchange of housekeeping genes within lineages. Consequently, chromosomal and symbiotic related genes frequently show different evolutionary history (Menna and Hungria, 2011; Parker, 2012; Steenkamp et al., 2008; Stepkowski et al., 2005; Vinuesa et al., 2005; Young and Wexler, 1988; Young et al., 2006). In this study, the phylogeny of the test strains based on symbiotic genes did not follow their taxonomic classification based on housekeeping genes (Paper I, Fig. 3; Paper II, Figs 3, 4; Paper III, Fig. 3). The symbiotic genes (*nod* and *nif*) sequences were highly conserved among the *Rhizobium* species compared to the protein coding housekeeping genes.

The phylogenetic pattern of symbiotic genes for test strains belonging to the genus *Bradyrhizobium* was relatively consistent with the phylogeny of the housekeeping genes of the respective strains (Paper II, Fig. 1 & Fig. 3 and Table 1). Some of the *nodA*, *nodY/K* and *nifH* groups, however, were not congruent with the phylogeny of the housekeeping genes of the *Bradyrhizobium* strains. For example, sub-groups L and N in the *nodA* clade IIII or L and K in the *nifH* gene tree were congruent with their concatenated core gene phylogeny, suggesting vertical gene transfer of the symbiotic genes. On the other hand, strains in group I.4 according to the concatenated core gene sequences were found in four different *nodA* or *nifH* groups, suggesting horizontal gene transfer. Therefore, the high diversity in *nod* and *nif* genes reported in this and previous studies may have been a result of the spread of the symbiotic islands both by vertical and horizontal gene transfer among *Bradyrhizobium* species (Menna and Hungria, 2011; Moulin et al., 2004).
5. CONCLUSIONS

MLSA was a valuable method for the taxonomic study of *Rhizobium*, *Agrobacterium* and *Bradyrhizobium*. By this method phylogenetically diverse *Rhizobium* and *Bradyrhizobium* species were discovered in the soils of Ethiopia. Their genetic diversity within the species was further explored using the whole genome based AFLP-fingerprinting technique. Although sometimes AFLP clustering of the strains did not follow their phylogenetic relationships based on MLSA, this method generally showed high resolution between strains not distinguished by the housekeeping gene sequencing. Thus, the AFLP fingerprinting technique is a powerful method to uncover the diversity of bacterial strains within species. It is recommended to include the AFLP fingerprinting technique along with MLSA in taxonomic studies of rhizobia.

The comparative analysis between housekeeping and symbiotic gene sequences for the test strains belonging to genus *Rhizobium* and *Bradyrhizobium* showed that the phylogenetic relationships based on the housekeeping genes were incongruent with symbiotic gene phylogeny, suggesting that the core and symbiotic genes had different evolutionary histories. However, for strains belonging to *Bradyrhizobium*, the phylogeny based on housekeeping genes was partially consistent with the phylogeny of symbiotic genes, indicating that *Bradyrhizobium* strains may inherit their symbiotic genes both by vertical transfer from their ancestor and horizontal transfer from more distant *Bradyrhizobium* species.

The *nifH* and *nodY/K* phylogenies of the Ethiopian *Bradyrhizobium* strains were generally congruent with the *nodA* gene phylogeny, suggesting the same origin of the symbiotic genes in the Ethiopian *Bradyrhizobium* strain studied here.

All common bean-nodulating *Rhizobium* species and most *Bradyrhizobium* species (except strains from *E. brucei*, which were not tested since seeds did not germinate) from soybean, *Indigofera* spp. and *Crotalaria* spp., formed effective nodules on their respective host legume species in the greenhouse or in the growth chamber, reflecting their potential role in increasing production of food legumes and promoting plant growth of agroforestry woody and shrubs legumes through nitrogen fixation.

The isolation of diverse non-rhizobial endophytic bacteria from root nodules of *Crotalaria* species, *Indigofera* spp., *E. brucei*, *P. vulgaris* and *G. max* and the presence of plant growth promoting activities in them suggest their potential role in promotion of plant growth while colonizing plant rhizospheres or root nodules of the host legume.
6. FUTURE PROSPECTS

Further taxonomic studies, such as the DNA-DNA hybridization technique and phenotypic methods will be performed to describe and name the prospective novel *Bradyrhizobium* (Paper II) and *Rhizobium* (Paper I, III) species that were identified based on the results of the MLSA and AFLP-fingerprint analyses.

Since the common bean-nodulating *Rhizobium* spp. and *Bradyrhizobium* spp. nodulating soybean and agroforestry shrubs (*Crotalaria* spp., *Indigofera* spp.) showed promising results in nodulation experiments under growth chamber or greenhouse conditions when plants were grown axenically, further greenhouse and field experiments are necessary to assess the nitrogen fixation effectiveness and adaptability of the test strains to local soil and field conditions. Ultimately, selection of effective and competent nitrogen-fixing rhizobia that can be applied in the field as inoculants to increase the productivity of food legumes (beans and soybean) and to boost the growth of agroforestry seedlings will be the next challenge.

As most of the non-rhizobial endophytic bacteria obtained from Tree, Shrub and Food legumes in this study showed plant growth promoting traits *in vitro*, further plant tests are needed to uncover the interactions between endophytic and symbiotic bacteria in the nodules and to test their efficacy on plant growth, nodulation and yields.
7. REFERENCES


49


