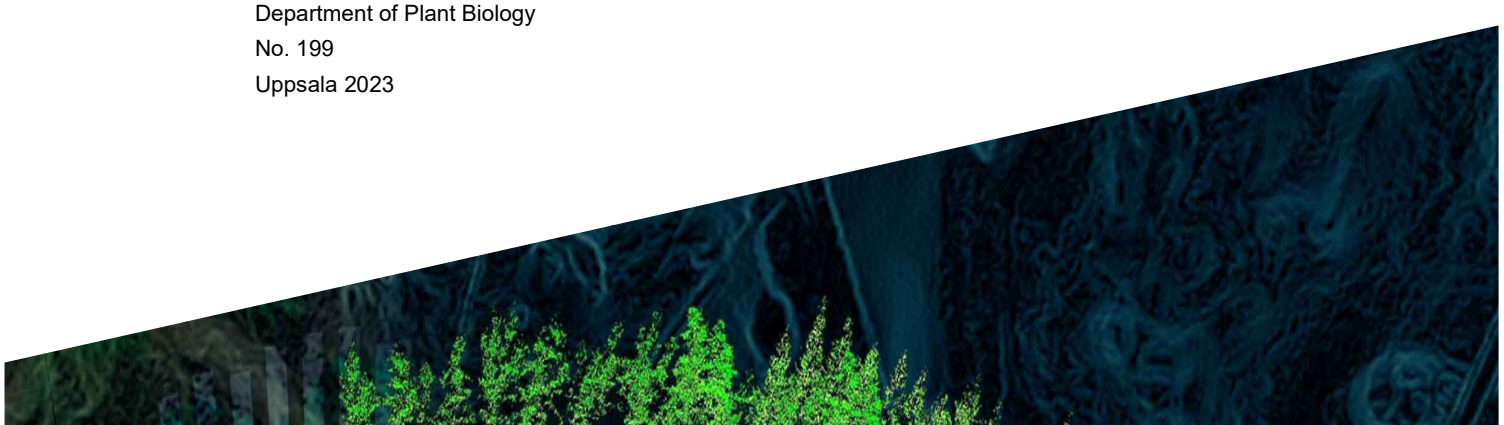
Swedish red clover-associated *Rhizobium leguminosarum* bv. *trifolii* share high genetic similarity in the *16S* gene but divergence in *nod* genes

Tua Lilja

Independent project • 60 credits
Swedish University of Agricultural Sciences, SLU
Department of Plant Biology
No. 199
Uppsala 2023



Swedish red clover-associated *Rhizobium leguminosarum* bv. *trifolii* share high genetic similarity in the *16S* gene but divergence in *nod* genes

Tua Lilja

Supervisor:	**Christina Dixelius, Swedish University of Agricultural Sciences, Department of Plant Biology**
Examiner:	Jens Sundström, Swedish University of Agricultural Sciences, Department of Plant Biology

Credits:	60
Level:	A2E
Course title:	Independent project
Course code:	EX0900
Course coordinating dept:	Department of Plant Biology
Place of publication:	Uppsala, Sweden
Year of publication:	2023
No.	199
Copyright:	All featured images are used with permission from the copyright owner.

Keywords:	Genetics, nitrogen fixation, sustainability, symbiosis, taxonomy

Swedish University of Agricultural Sciences

Faculty of Natural Resources and Agricultural Sciences

Department of Plant Biology

Publishing and archiving

Approved students' theses at SLU are published electronically. As a student, you have the copyright to your own work and need to approve the electronic publishing. If you check the box for **YES**, the full text (pdf file) and metadata will be visible and searchable online. If you check the box for **NO**, only the metadata and the abstract will be visible and searchable online. Nevertheless, when the document is uploaded it will still be archived as a digital file.

If you are more than one author you all need to agree on a decision. Read about SLU's publishing agreement here: https://www.slu.se/en/subweb/library/publishand-analyse/register-and-publish/agreement-for-publishing/.

☒ YES, I/we hereby give permission to publish the present thesis in accordance with the SLU agreement regarding the transfer of the right to publish a work.

☐ NO, I/we do not give permission to publish the present work. The work will still be archived, and its metadata and abstract will be visible and searchable.

Abstract

Rhizobium leguminosarum biovar *trifolii* (*Rlt*) is a root-associated soil bacterium (rhizobia) capable of entering a symbiotic relationship with various species of clover (*Trifolium* sp.) and fix nitrogen within root nodules.

Many rhizobial genes required for host infection and nitrogen fixation are located on symbiotic plasmids, which are highly mobile between strains. In this study the genetic diversity of one chromosomal phylogenetic marker gene (16S rRNA) and four symbiotically associated plasmid-located nodulation genes (*nodA, nodB, nodD1,* and *nodD2*) of thirteen isolates were analysed. The isolated *Rlt* were derived from red clover root nodules collected in various locations in Sweden.

It was determined that Swedish *Rlt* isolates were highly similar regarding the chromosomal marker 16S gene sequences but displayed sequence variations regarding the symbiotic *nod* genes *nodA*, *nodB*, and *nodD1.* The gene *nodD2* was not detected. The variations among these genes could be partly geographically grouped in the north and south of Sweden. Similarities in the marker gene confirms a close genetic relationship, whereas the divergence in *nod* genes suggests variations in competitiveness between the strains.

Considering the differences seen in the *nod* genes, comparison to international sequences showed that the Swedish *Rlt* strains are closely related. The high genetic similarities among the isolates suggest that the *Rlt* strains in Sweden may share a common origin.

Further genetic and phenotypic analyses must be conducted to determine whether the *nod* gene variants result in a difference in competitiveness.

Keywords: Genetics, nitrogen fixation, sustainability, symbiosis, taxonomy

Table of contents

List of tables

List of figures

Abbreviations

BNF	Biological nitrogen fixation
Bv	Biovar
EPS	Exopolysaccharides
GTR	General Time Reversible model
HGT	Horizontal gene transfer
IT	Infection thread
NCBI	National Centre for Biotechnology Information
PCR	Polymerase chain reaction
PGPB	Plant growth promoting bacteria
RAxML	Randomized Axelerated Maximum Likelihood
Rlp	*Rhizobium leguminosarum* bv. *phaseolus*
Rlt	*Rhizobium leguminosarum* bv. *trifolii*
Rlv	*Rhizobium leguminosarum* bv. *viciae*

Introduction

1.1 A brief history of the legume-rhizobia friendship

To put some perspective on the extent of the research on biological nitrogen fixation: the first comprehensive attempts at reviewing the topic were written in the 1880s and 1890s. Remarkably, publications on the subject predate that by over two millennia. The philosopher Theophrastus (370–285 B.C.) wrote of the reinvigorating effects of legume crops on soil in his text "Enquiry into Plants":

> "Of the other leguminous plants, the bean best reinvigorates the ground, even if it is sown thick and produces much fruit,"

and

> "Beans ... are not a burdensome crop to the ground; they even seem to manure it, because the plant is of loose growth and rots easily; wherefore the people of Macedonia and Thessaly turn over the ground when it is in flower."

Cato (234–149 B.C.) continued with

> "Lupines, field beans, and vetches manure corn land,"

> "Sow, for feed for cattle, clover, vetch, fenugreek, field beans, and pulse. Sow these crops a second and a third time."

This early understanding of crop rotation and green manure is further documented by, for example, Jia Sixie, who in 500 AD wrote in "Qimin Yaoshu" about the

benefit of growing *Phaseolus mungo* (now *Vigna mungo*) as green manure (Fred et al. 1932).

However, the method by which legumes enrich the soil with nitrogen was not demonstrated until 1838, when Boussingault found that the nitrogen in the plant must come from the air itself and that this enrichment occurred with peas and clover, but not with wheat (Erisman et al. 2008). A publication by Berthelots in 1885 was the first that suggested microorganisms as the agents of nitrogen fixation, and the same implications were made in other studies of the time.

At the same time, nodules had been remarked upon for centuries. The first published picture of root nodules was by Fuchs in his "De Historia Stirpium Commentarii Insignes" in 1542, however, he did not comment upon them. In "Historia Generalis Plantarum", published 1586 and 1587, Dalechamps and Rouillé specifically mention root nodules as they describe an *Ornithopodium* species. This is the earliest remark of root nodules in literature, but it took another three centuries to elucidate their importance.

Hellriegel and Wilfarth are known as the team that connected nodules with nitrogen fixation in 1886. They also suggested that microorganisms residing in the nodules that were responsible.

The first species, a bean symbiont, of the microorganisms was identified in 1889 by Frank and named *Rhizobium leguminosarum*. More on the history of the subject can be found in 1932's "Root Nodule Bacteria and Leguminous Plants" by Fred and coworkers, where the details of all the publications mentioned above are found (Fred et al. 1932).

1.2 Clover

Clovers, the *Trifolium* genus of the Fabaceae, are a relatively new agricultural addition, first domesticated in Moorish Andalusia in 1000 A.D and slow to spread throughout Europe (Kjærgaard 2003). In the 17th and 18th centuries, the cultivation

of clover was expanded to re-fertilise the nitrogen-depleted soils in agricultural areas, an unfortunate result of the increased needs of an expanding population and the demands of the ever-growing cities. Clover remained an essential part of European agriculture until the Haber–Bosch ammonia synthesis method of 1913 made enormous amounts of industrially fixed nitrogen available, facilitating the establishment of the monoculture crop methods dominating the last century. Currently, inorganic nitrogen feeds about half of the world's population (Erisman et al. 2008). While the sustainability factor in farming is coming strongly into focus, clover may return to being as essential a crop as previously.

More than 250 annual and perennial species are found in the *Trifolium* genus. Most of the diversity of these species is found in the temperate northern hemisphere, but the genus is found worldwide. The agricultural use of *Trifolium pratense, Trifolium resupinatum*, *Trifolium hybridum,* and *Trifolium repens* are mainly as fodder crops but also as green manure plants for improving soil fertility by nitrogen fixation. Red clover additionally contributes to soil health by promoting a high soil microbe activity, improving soil structure, and it contributes to local areas by encouraging a large population of pollinators (Riggi et al. 2021).

Fodder based on red clover, or supplemented with it, increases growth performance of lambs (Weinert-Nelson et al. 2023), average daily gains in steers (Harlow et al. 2020), and milk yield in cows (Johansen et al. 2018). Red clover also increases the equol content, a metabolite of pharmacological interest (Mayo et al. 2019) in eggs (Tosar et al. 2021) and milk (Mustonen et al. 2009). However, these positive results are mainly connected to a mixed diet, where a higher percentage of red clover in the feed did not proportionally increase the product benefits.

Intercropping clover species with other crop varieties can increase biomass and yield with a reduced application of exogenous nitrogen fertilization (Kintl et al. 2018). Additionally, intercropping can reduce nitrogen loss through leaching. Mixed sowing of forage grasses and clover species reduces the need for nitrogen supplementation while producing higher-quality forage (Evers 2011). Intercropping of grasslands with legume- and non-legume crops has shown transgressive overyielding (overyielding the monoculture of the best-performing

species in the mixture) in a variety of environmental conditions (Finn et al. 2013). When breaking up forage land where legumes have grown, much nitrogen is lost from the exposed soil. It is important to sow the right crop afterwards to take up the recently freed nitrogen. There should be as short a time as possible between plowing and sowing the new crop to minimize loss.

1.3 Rhizobium leguminosarum

Two hundred thirty-eight species of legume-nodulating bacteria belonging to 18 genera and two clades, as well as 33 closely related non-symbiotic species have been reported (Shamseldin et al. 2017).

Pseudomonadota (previously Proteobacteria) are a major phylum of Gram-negative bacteria. The classes of legume-nodulating bacteria are the α, β and γ-Proteobacteria. The α-Proteobacteria are the largest class among the three and contain most nodulating species. It is the class containing the family Rhizobiaceae, in which the genus *Rhizobium* can be found.

The *Rhizobium* genus currently contains 98 species, of which 69 have been isolated from legume hosts and 29 have not been found symbiotic (Shamseldin et al. 2017). Rhizobium are motile, Gram-negative, non-sporulating bacteria capable of living both free in the soil and as nitrogen-fixing bacteroids in symbiotic organs on legume roots called root nodules (**Figure 1**).

Figure 1. Hydroponically grown red clover inoculated with and successfully nodulated by a previously isolated Swedish *Rlt* strain. Photograph kindly provided by Shridhar Jambagi.

Biovars, biological variants (previously known as biotypes), can be defined as symbiotically distinct subgroups of a species. They share a common chromosomal background, but their symbiotic plasmids vary. The term symbiovar, for symbiotic variant, is also used as a parallel term to pathovar (Rogel et al. 2011). The term biovar will be used throughout this text. The existence of multiple biovars per species is a trait common among *Rhizobium*. Three biovars (bv) can be clearly distinguished in the species complex *Rhizobium leguminosarum*: *R. leguminosarum* bv. *trifolii, R. leguminosarum* bv. *viciae* (*Rlv*), and *R. leguminosarum* bv. *phaseoli* (*Rlp*).

These biovars have defined host ranges of various narrowness, with *Rlt* only establishing symbiosis with *Trifolium* species, *Rlv* with *Pisum*, *Vicia*, *Lens*, and *Lathyrus*, and *Rlp* with *Phaseolus* species (Acosta-Jurado et al. 2021). *Rlp* forms round determinate nodules with a transient meristem, whereas *Rlt* and *Rlv* form rod-like indeterminate nodules with an apical meristem. The *Rlp* nodules are among the three that alone encourage the bacteria that differentiate into bacteroids to retain their morphology and easier return to a free-living state, in contrast to the bacteroids of *Rlt* and *Rlv* that mainly undergo terminal differentiation (Haag et al. 2013).

1.4 The rhizosphere

The areas in the soil most influenced by the plant are the rhizosphere, rhizoplane, and endosphere. These areas are the soil around the roots and on the root surfaces, and the area between the plant cells. The soil surrounding the plant is richer in microbial life than bulk soil and much more species diverse (Reinhold-Hurek et al. 2015). The plant and its microbiome are defined as the holobiont, a complex and dynamic ecological unit connected through symbiosis, and its collective genome is the hologenome (Baedke et al. 2020). The holobiont contains all organisms that contribute to the function of the units, not limited to microbes. Thus, the symbiotic unit can be looked at as an organism in itself.

The diverse rhizo-environment is shaped by the plants through, for example, root exudates rich in organic acids, vitamins, sugar, and signal molecules. In turn, the microbe community contributes to plant nutrition by fixing nitrogen and solubilizing phosphates, to stress response by providing or degrading phytohormones, and to biocontrol through competing with pathogens or releasing antibiotics (Reinhold-Hurek et al. 2015). These bacteria contributing to plant growth are known as plant growth-promoting bacteria (PGPB) or plant growth-promoting rhizobacteria (PGPR). They significantly contribute to more sustainable agriculture with a reduced input of fertilizers and other agrochemicals (Shah et al. 2021). The use of PGPB has a history spanning more than a century, with the first commercial product developed in 1896. The first PGPB carrier material used was gelatine, later changed to peat-based material, and to today's liquid inoculants with synthetic polymers protecting the cells. Biopolymers in the form of bacteria-derived exopolysaccharides (EPS) are an alternative for inoculant development and production as EPS protect against desiccation, salinity, pH, and other stress components (Palhares Farias et al. 2022). Survival of the bacteria has been a constant issue and the storage and handling of these living materials have been a limiting factor in the success of their use (Santos et al. 2019).

Microbial inoculants and pre-inoculated seeds are available from many distributors. The products promise, for example, plant growth promotion (*Cladosporium tenuissimum*), disease management (*Streptomyces lydicus* and *Pseudomonas chlororaphis*), and pest management (*Paenibacillus popilliae* and *Serratia entomophila)* (O'Callaghan et al. 2022). The main inoculant products today and historically are rhizobial species that nodulate legumes and provide nitrogen through biological nitrogen fixation (BNF).

1.5 Nitrification

Nitrogen deficiency is an often-limiting issue for plants, even though the atmosphere comprises approximately 80% N_2. This atmospheric elemental nitrogen is unavailable to plants and can only be utilized after nitrification: conversion of dinitrogen to nitrogen-containing organically incorporable compounds. Dinitrogen

has a low level of reactivity due to its triplet covalent bond, which requires 16 moles of ATP per mole of dinitrogen to be broken for nitrogen fixation. This energy-demanding process is carried out by the enzyme nitrogenase in rhizobia (Massa et al. 2022). Most legume species, approximately 90% (Rascio & La Rocca 2013), can fix atmospheric nitrogen via their symbiosis with rhizobia.

The natural nitrogen cycle involves several steps. Nitrogen fixation, ammonification, nitrification, and denitrification are all stages in this process. Nitrogen fixation converts atmospheric elemental nitrogen (N_2) to organically incorporated nitrogen into compounds, such as amines ($-NH_2$). Ammonification is the process of converting this organically incorporated nitrogen into ammonia (NH_3) and ammonium ions (NH_4+). During nitrification, the ammonia and ammonium are oxidized into nitrite (NO_2-) and nitrate (NO_3-). Finally, denitrification reduces nitrogen oxides back into elemental nitrogen (Hillel 2008).

On average, legumes can annually fix 20–200 kg per hectare. The total nitrogen fixation in the world is estimated to be ~175 Tg; out of this, an approximately ~80 Tg is fixed by legumes (Shah et al. 2021 and references therein). Much of the remaining nitrogen is fixed industrially during the production of synthetic nitrogen fertilizers and a smaller amount by lightning strikes.

Only about half of the applied nitrogen is used by the crops (Lassaletta et al. 2014). The other half escapes the fields by run-off, nitrate leaching, or as nitrous oxide or ammonia, causing coastal eutrophication, air pollution, and increased greenhouse gas emission.

1.6 The legume-rhizobia symbiosis

1.6.1 Flavonoids and NodD

The symbiosis of legume-rhizobia requires a complex and highly specific exchange of signal molecules, beginning with the plant flavonoids being released into the rhizosphere. The flavonoids are secondary metabolites of variable phenolic structures and act both as chemoattractants and inducers of rhizobial *nod* gene

transcription. The flavonoids are recognized and bind to NodD, a LysR-type transcriptional regulator. The LysR family of transcriptional activators are approximately 35 kDa, possess an N-terminal helix-turn-helix DNA binding domain, most often require an inducer for activation, and frequently auto-repress its own transcription (Schell 1993). The *nodD* gene encodes the rhizobial transcriptional activator that recognizes the plant flavonoids and induces *nod* gene expression. The flavonoid-bound form of NodD activates transcription of the *nodABCIJ* operon, which lies adjacent to the *nodD* gene but is transcribed divergently through interaction with the upstream cis-regulatory element called the *nod* box. The *nod* box is a highly conserved DNA sequence located in the promoter region of the NodD-activated genes (**Figure 2**).

NodD simultaneously represses its own transcription when activating transcription of the *nodABC* operon. The promoter regions of *nodD* and *nodA* overlap and when binding to the *nodA nod*-box, NodD also inhibits its own expression. This autoregulation is produced by NodD competing with RNA polymerase for binding sites (Hu et al. 2000) and physically altering the spatial structure of the DNA (Chen et al. 2005).

The *nodD* gene is an important determinator of the specificity of the legume-rhizobia symbiosis; some species contain up to five isoforms of the gene, which broadens their host range through the possibility of detection of additional flavonoid variants. A second *nodD*, *nodD2*, in *Rlt* has been shown to enhance nodule colonization competitiveness, but not all strains possess this second form (Ferguson et al. 2020).

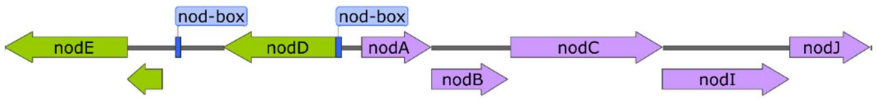


Figure 2. The *nodABC* region is transcribed divergently from *nodD*. Two *nod* boxes are found flanking *nodD*: one located in the overlapping promoter area upstream of *nodD* and *nodA*, and one upstream of *nodFE*. Figure created using Snapgene 6.2 (GSL Biotech LLC).

1.6.2 Function, synthesis, and evolution of the Nod factors

The products of the *nod* genes are the Nod factors (NFs); they recognize flavonoids secreted from the plant. This recognition leads to the expression of *nod* genes and subsequent production of NFs. The process is induced and controlled by the self-regulating NodD transcriptional regulators.

The NFs of all *Rhizobium* species belong to the same chemical family, the lipochitooligosaccharides (LCOs). The enzyme products of the canonical *nodABC* operon synthesize the basic NF structure. NodC is an N-acetyl-d-glucosamine transferase that synthetizes the chitin oligomer backbone by combining N-acetylglucosamine monomers into a single tetra- or pentameric molecule with β-1,4 glycosidic bonds. NodB is a chitooligosaccharide deacetylase that removes the acetyl residues from the backbone nonreducing end, and NodA is an acyl transferase that N-acylates the nonreducing end.

The additional *nod* genes responsible for the host specificity are called host-specific nodulation genes (*hsn*), and include *nodFE*, *nodH*, *nodG*, *nodPQ,* and others, depending on the species and strain. Their enzyme products contribute with adding additional chemical groups to the NF backbone such as fucose, sulphate, and acetate. They also modify the fatty acid chain length and saturation in a strain-specific manner. Depending on the species and strain, rhizobia may produce two to sixty variants of NFs (D'Haeze & Holsters 2002).

The *nodABC* genes are conserved among most rhizobia; when missing, the result is a total lack of nodulation. An exception is found in some *Bradyrhizobia* strains, who without *nodABC* genes successfully nodulate their hosts (Giraud et al. 2007) with a method that is suggested to predate NF-dependent symbiosis (Okazaki et al. 2016). NF-dependent symbiosis may have evolved from this NF-independent strategy, which in turn may have evolved from a system originally developed to overcome host resistance. It is regarded as likely that the NF-dependent strategy was evolved from the mycorrhizal symbiosis between plants and fungi, as fungi produce their own LCOs.

Over 90% of land plants can form a mycorrhizal symbiosis with Glomeromycota (Bonfante & Genre 2010). Some fungal species produce Myc factors that are

lipochitooligosaccharides structurally almost identical to NFs. As with species-specific NFs, they differ mostly in their additional substituents. Arbuscular mycorrhizas predate legume-rhizobia symbiosis by 340 million years, and the legume-rhizobia symbiosis is thought to have evolved from it (Maillet et al. 2011).

1.7 Rhizobial nitrogenase and its variants

The rhizobial nitrogenase protein complex encoded by the *nif* genes comprises six subunits: two copies each of NifH, NifD, and NifK. It contains four iron-sulphur clusters, two [4Fe–4S] and two [Fe8S7], and two iron-molybdenum cofactors called FeMoco (FeMo cofactor). The N_2 reduction takes place at the FeMoco metallocentre. The metallocentres are easily oxidized and must therefore operate in a low-oxygen environment. This conflicts with the high level of energy that needs to be supplied by aerobic respiration for the reduction process, as 16 ATP molecules are required for one reduced N_2 molecule.

The contradiction is solved by the host root production of leghaemoglobins (legume haemoglobins): oxygen-carrying enzymes that buffer oxygen concentration in the nodule by binding and releasing O_2. Leghaemoglobins allow the aerobic respiration to continue while the nitrogenase still has the low-oxygen environment it requires (Larrainzar et al. 2020). Further, expression of the *nif* genes is negatively regulated by free oxygen in the nodule, thus the leghaemoglobins are essential throughout the BNF process (Ott et al. 2005).

The root nodules actively fixing nitrogen are distinguished by their pink colour from the leghaemoglobins (**Figure 3**).

Figure 3. Pink active nodules on clover roots. Photograph kindly provided by Shridhar Jambagi.

The three nitrogenase subunits share a significant sequence similarity to the three subunits of dark-operative protochlorophyllide reductase, which light-independently converts protochlorophyllide to chlorophyll (Li et al. 1993). Although no longer present in angiosperms, it is extant in gymnosperms, photosynthetic bacteria, and algae.

Nitrogen fixation is not limited to the plant-associated microbes in the soil, but rather widespread throughout various environments, including, *e.g.*, marine sediments, the termite gut, and human pathogens (Jasniewski et al. 2020). An increasing number of genomes have become available, resulting in more organisms carrying putative alternative nitrogenases have been identified (McRose et al. 2017). Instead of molybdenum, the alternative nitrogenases use iron or vanadium. These alternative nitrogenases are suggested to contribute more than 23% to the world's total nitrogen fixation.

1.8 Host infection process

After the rhizobia recognise the plant exudates, NFs are produced. The detection of the NF by the LysM motif-containing plant receptor-like kinases (chitooligosaccharide-detecting cell surface receptors involved in plant immunity) of the root hair initiates the organogenesis of the lateral root organ called the root nodule. Inhibition of growth at the root hair tip leads to growth deformation causing curling of the hair which traps rhizobial cells in the resulting infection pocket. Through localized degradation of the cell wall and rearrangement of the cytoskeleton, an invagination called the infection thread (IT) is formed, leading into the epidermis and further into the cortex. During the extension of the IT, rhizobia near the growing tip replicate and populate the IT as it progresses. Meanwhile, a nodule meristem is formed through de-differentiation and division of the cortical cells. After the initial rhizobial invasion of the intercellular space, this primordial nodule is colonized with rhizobia cells. A host-derived membrane encloses the rhizobia inside the developing primordia as they exit the IT. This forms the symbiosome, a de novo organelle in which the rhizobia differentiate into round, large bacteroids: capable of nitrogen fixation but no longer replicating. More details of the infection process can be found in Walker et al. (2020), and references therein.

Compatibility between rhizobia and legume hosts goes much further than the initial flavonoids and NFs. Effector-triggered immunity (Cui et al. 2015), exopolysaccharides (EPS), nodule-specific cysteine-rich antimicrobial peptides (Roy et al. 2020), and transfer RNA -derived small RNA fragments (Ren et al. 2019) are a few additional systems utilized to ensure that only the properly functioning symbiosis is initiated and supported.

The host ranges of rhizobia differ between species. The *Rlt* biovar is only capable of symbiosis with *Trifolium* species, whereas, for example, the *Rhizobium* strain NGR234 can nodulate 112 genera, an amazing range believed to result from the strain possessing a large number of secretion systems (Schmeisser et al. 2009). Strain NGR234 also encode 19 nod-boxes, twice as many as typical, and a second NodD transcriptional activator, which offers an additional suggestion of the broad

host range caused by a more complex signal network with a temporal shift of expression (Kobayashi et al. 2004).

Additionally, the symbiosis range of legumes varies. *Lotus burttii* is capable of symbiosis with rhizobia from five distinct genera, and is thus more promiscuous than its relative, the model plant *Lotus japonicus* (Zarrabian et al. 2022). This promiscuous trait can be traced to a single QTL. However, the extended symbiosis range can be detrimental to the host due to nodule formation with insufficient nitrogen fixers can lead to reduced plant benefit or even loss of growth.

The symbiosis range also depends on what infection process is utilized. In *Lotus japonicus*, the root hair infection process with ITs is more highly controlled by the host when it comes to partner compatibility than that of crack entry (with or without ITs), which is more controlled than that of single-cell infection (Madsen et al. 2010). The single cell infection is thought to be the most primitive of the infection processes.

1.9 Rhizobial exopolysaccharides

An important ingredient in the formation of the symbiotic relationship between legumes and rhizobia is the bacteria's production of extracellular polysaccharides. These include the surface polysaccharides, such as EPS, lipopolysaccharides, capsule polysaccharides, glucomannans, cyclic β-glucans, and cellulose fibrils that play several important roles in species-specific interactions, affecting host-rhizobium signalling, root tissue invasion, nodule formation, and the suppression of host immunity, as well as extracellular survival (Downie 2010). The EPS are acidic polysaccharides secreted outside the cell with little or no cell association (López-Baena et al. 2016). Production of EPS requires the expression of several *exo* and *pss* genes. The function of EPS include protection against biotic and abiotic stresses, attachment to surfaces, and nutrient gathering. Additional diverse functions of EPS during the interaction with the host plant have been suggested, such as calcium ion sequestration, reactive oxygen species scavenging, prevention of cell clumping, pH tolerance, and host surface attachment (Acosta-Jurado et al.

2021). EPS production is essential for biofilm formation, and reduced EPS production decreases the rhizobia's ability to form nodules and fix nitrogen. A gene duplication of the EPS production and regulatory genes *rosR* and *pssA* in *Rlt* increased the competitiveness of the strain, increased nodule formation, and caused increased biomass of the infected plants (Janczarek et al. 2009).

The EPS of *Rlt* is a heteropolymer consisting of octasaccharide subunits composed of five glucose residues, one galactose, and two glucuronic acid residues, and further modified with acetyl, pyruvyl, and hydroxy-butanoyl groups (Hollingsworth et al. 1988). The EPS structure is species- and sometimes strain-dependent (Wanke et al. 2021). In most *Rlt* strains, the EPS structure is repeating units of octosaccarides that contain D-glucose, D-glucuronic acid, and D-galactose, joined by β-1,3 and β-1,4 glycosidic bonds in a molar ratio of 5:2:1 (Acosta-Jurado et al. 2021). The structure is further modified with non-sugar (acetyl and pyruvyl) groups. This galactose unit is missing in some *Rlt* strains, and in *Rlv* there is an extra D-glucuronic residue. In *Rlp*, EPS appears not to be essential for symbiosis as it is in *Rlt* and *Rlv*.

1.10 Horizontal gene transfer

Metabolic flexibility is vital in the structurally diverse environment of the soil. The soil is a setting where slow growth due to a large genome may not be as disadvantageous as in other environments (Konstantinidis & Tiedje 2004). Rhizobial genomes are large and complex, as required to adapt to and survive the everchanging environment of the rhizosphere. The large genome size is a common trait among soil bacteria (Mazur et al. 2011). These large genomes are enriched in genes for transport, secondary metabolism, and gene regulation.

The symbiotic bacteria have an even larger demand of a dynamic genome, as they cycle between intracellular and extracellular lifestyles. In *Rlv*, a total of 603 genetic regions including 593 genes, 5 transfer RNAs, and 5 RNA features, are required to competitively nodulate their host (Wheatley et al. 2020). This corresponds to roughly a tenth of the coding genome.

In *Rlt*, the 6.9 Mbp multipartite genome is divided into five separate replicons consisting of one chromosome of approximately 4.5 Mbp and four plasmids of 1.3 Mbp, 500 kbp, 309 kbp, and 2.6 kbp, respectively (Reeve et al. 2010). The complete genome contains 6643 protein-coding genes and 62 RNA-encoding genes. The *R. leguminosarum* genome has two basic components: the core genome, which is mainly chromosomal, and the accessory genome, which is mainly plasmid-located (Mazur et al. 2011). The bacterial genes responsible for pathogenic or symbiotic relationships with eukaryotic hosts are often a part of the accessory gene pool and often acquired by horizontal gene transfer (HGT). They can be located in the genome in so-called symbiosis islands or clustered on plasmids which may vary significantly even between close strains. The HGT of bacterial DNA can be triggered by plant flavonoids released into the rhizosphere (Ling et al. 2016).

One could classify the core genome as the "non-mobile" genome and the accessory as the "mobile" genome. The core genome has a higher G+C content and has a consistent phylogeny compared to the accessory genome, which may differ widely even between closely related rhizobia (Young et al. 2006). The accessory genome is mostly located on symbiotic plasmids, but there are also chromosomal regions that belong to this "mobile" genome. Most core genes are located on the chromosome, but some are found on plasmids, further complicating the distinction between the two. The *nod* genes and the *nif* (nitrogen fixation) genes, among others, are part of the accessory genome and located on one of the symbiotic plasmids (pSym) in *Rlt* and is therefore easily laterally spread throughout a population. In contrast, in *Mesorhizobium loti* strain MAFF303099 (Kaneko et al. 2000) and in *Bradyrhizobium japonicum* USDA110 (Kaneko et al. 2002) the symbiosis genes are located on the chromosome in symbiosis islands.

The exchange of accessory genes between strains confers adaptations to the environment, such as host adaption and adaption to the local soil conditions. Complex traits can in that way spread throughout a population, such as changing a non-symbiotic tribe of bacteria into capable symbionts. One example of horizontal transfer of symbiosis genes between strains comes from New Zeeland (Sullivan et al. 1995). *Lotus corniculatus* seeds were inoculated with a single strain of *Mesorhizobium loti* and sown in an area where no indigenous nodulating symbionts

had been previously detected. Uninoculated lotus plants died of nitrogen deficiency, whereas inoculated plants were healthy. After seven years without interference, nodules from lotus plants at the site were analysed and found to contain genetically diverse *Mesorhizobium*, all containing a 500 kb chromosomally located symbiosis island. This island had been transferred from the original inoculation strain into the native non-symbiotic *Mesorhizobium*. A P4-type integrase, encoded in the symbiosis island, mediated the island's integration into a phenylalanine-tRNA gene in the recipient native strain.

As the mobile accessory genome may differ even among closely related strains, there is a need to analyse multiple DNA sequences from both the chromosomal core and the plasmid accessory genome of a strain if a confident taxonomy is to be determined. The 16S rRNA gene is commonly used as a chromosomal core gene, and the *nodABC* genes as plasmid accessory genes.

Aim

The aim of this study was to determine whether there are genetic variations among geographically diverse *Rlt* isolates collected throughout Sweden. The hypothesis was that there would be no genetic variations between the isolates. One phylogenetic marker gene (16S rRNA) and four symbiotically relevant accessory genes (*nodA*, *nodB*, *nodD1* and *nodD2*) of sixteen isolates were analysed to determine this.

Materials and methods

3.1 Microbial culturing

3.1.1 Media composition

The complex yeast extract mannitol media was used both as solid and liquid growth media. Its composition is described in **Table 1**. Jensen's nitrogen free media (Jensen 1942) was used as a defined solid media. Both are mainly selective for rhizobia. The addition of Congo Red dye facilitated identification of the correct, non-dye absorbent *Rhizobium* colonies.

Table 1. Composition of selective growth media.

Growth media	Composition g/L
Yeast Extract Mannitol media (YEM)	Mannitol 10 g, dipotassium phosphate 0.5 g, magnesium sulphate 0.2 g, yeast extract 1 g, sodium chloride 0.1 g
YEM Agar media (YEMA)	YEM with addition of 20 g Bactoagar
YEMA CR	YEMA with 0.025 g/L Congo red dye
Jensen's nitrogen-free liquid media	Sucrose 20 g, dipotassium phosphate 1 g, magnesium sulphate 0.5 g, sodium chloride 0.5 g, ferrous sulphate 0.1 g, sodium molybdate 0.005 g, calcium carbonate 2 g
Jensen's solid media	Jensen's with addition of 15 g Bactoagar
Jensen's CR	Jensen's solid media with 0.025 g/L Congo red dye

3.1.2 Isolation of *Rhizobium*

Samples of *Trifolium pratense* root systems were previously collected from various locations in Sweden (Jambagi et al. 2023). The roots were gently washed with room-temperature tap water. Pink and firm nodules were excised using a sterile

scalpel. The nodules were sterilized with 90% EtOH for 30 s, 10% sodium hypochlorite for 90 s, and then washed with sterile deionized water five times. Three to five nodules per sample (approximating the same tissue volume in each sample) were crushed in 100 µl of sterile tap (non-deionized) water with a sterile plastic rod until a milky suspension was formed. The suspension was diluted in steps from 10^{-1} to 10^{-10} with sterile tap water. The dilutions were spread in 100 µl portions onto YEMA and Jensen's solid media, and the plates were incubated at 28˚C in the dark. When colonies could be visually detected, at two to seven days, a selection of large, medium, and small colonies were re-streaked twice to obtain pure cultures. Culture plates were stored at 4˚C for a maximum of four weeks, then sub-cultured onto fresh media.

3.1.3 Culturing conditions

Liquid cultures were grown in 5 mL of YEM in 10 mL Falcon tubes at 28˚C with agitation of 180 rpm, for 48 hours. The cells were pelleted by centrifugation, resuspended in YEM with 25% sterile glycerol, and frozen in aliquots. Cultures were stored short-term at -20˚C and long-term at -70˚C and maintained at 4°C as active source cultures.

Liquid cultures for genomic DNA extraction were grown in 8 mL of YEM with the same culture conditions as described above.

3.2 Molecular analysis

3.2.1 Gene amplification

Thermocycler settings for PCR DNA amplification with Phusion High-Fidelity PCR Master Mix (Thermo Scientific) were as follows: initial denaturation 98˚C for 30s, denaturation at 98˚C for 10s, annealing at 65˚C for 30s, extension at 72˚C for 30s per kb, repeated 35x from second denaturation step, ending with final amplification at 72˚C for 5 minutes. Amplified products were visualized by 1.5% agarose gel electrophoresis to confirm product size.

A gradient PCR was performed to analyse the annealing temperature of the primers with annealing temperatures from 58˚C to 70˚C.

3.2.2 DNA extraction methods and sequencing

Genomic bacterial DNA for molecular analysis was extracted from liquid cultures using GeneJet Genomic DNA Purification Kit (Thermo Scientific) according to the manufacturer's instructions. Purified DNA was stored at -20˚C.

PCR amplification products were extracted after separation on 1.5% agarose gel with Midori green dye by E.Z.N.A. Gel Extraction Kit (Omega Bio-tek) or GeneJet Gel Extraction kit (Thermo Scientific) according to the manufacturer's instructions. Extracted DNA was stored at -20˚C.

16S rRNA PCR products were purified from PCR reactions using GeneJet PCR Purification Kit (Thermo Scientific) according to the manufacturer's instructions.

3.3 In silico analysis

3.3.1 Primer design

Genes chosen for molecular analysis and the primers for amplification of bacterial DNA are listed in

Table 2. Primers were designed using MEGA v17.1.1, Snapgene 6.2 (GSL Biotech LLC), and the NCBI primer design tool (Ye et al. 2012) (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). The primers were designed to flank the genes of interest with high specificity, as well as to have a uniform Tm (melting temperature) of 65˚C when PCR amplifying DNA using Phusion High-Fidelity enzyme to allow for a standardised PCR protocol.

Table 2. Genes and the primers used for PCR amplification and gene sequencing.

Gene	Function	Primer	Primer sequence (5'->3')	Amplicon
16S rRNA	Taxonomy	338-F	ACTCCTACGGGAGGCAGCA	855 nt
		1193-R	ACGTCATCCCCACCTTCC	
nodD1	Transcriptional regulator of *nod* genes	nodD1_F	GCGTTTTAAGGGCCTGGAC	976 nt
		nodD1_R	AGCCGCACACCATTCGG	
nodD2	Potential transcriptional regulator of *nod* genes	nodD2- F	GGGCCTTGATCTAAACCTGCT	946 nt
		nodD2- R	TCAGGAACCGGATTGTTGTAGC	
nodA	*nod* gene involved in NF production	nodA2-F	ATGTCTGCTGGAGTGCGGT	591 nt
		nodA2-R	TCATAGCTCGCACCCGTTT	
nodB	*nod* gene involved in NF production	nodB2-F	ATATAAGCGAAGTGCCGTTGGA	635 nt
		nodB2-R	TTACTGGGGAAGTGAACGGAG	

3.3.2 DNA sequencing and analysis

The amplified products were Sanger sequenced by Lightrun Genomics (Eurofins Genomics, Germany). The sequencing chromatograms were visually examined using Snapgene V6.2.1 (GSL Biotech LLC) to manually detect artefacts and the sequences were trimmed accordingly. An initial alignment of sequences was performed using MegAlign Pro v17.1.1 (DNASTAR, Inc.) with the Clustal Omega method (Sievers & Higgins 2018) to identify and correct remaining sequencing artefacts. Full-length DNA sequences were assembled from forward and reverse fragments.

The taxonomy of the isolates was determined using BLASTN (https://blast.ncbi.nlm.nih.gov/Blast.cgi) searches with the partial 16S rRNA sequences as query. The identity of the amplified *nodA, nodB* and *nodD* genes was likewise confirmed by BLASTN.

Trimmed, combined, and confirmed DNA sequences were compared by multiple alignment using MegAlign Pro. The open reading frames of the DNA sequences were visualised using Snapgene. The amino acid sequences were analysed by multiple alignment using Clustal Omega to determine whether the nucleotide variants resulted in amino acid variations. Phylogenetic trees were constructed by the Randomized Axelerated Maximum Likelihood method (RAxML) (Stamatakis 2014). Settings in MegAlign Pro were random starting seed

number, bootstrap analysis with 1000 replicates, and using the General Time Reversible (GTR) substitution model. The RAxML method generated trees with the highest bootstrap support compared to other methods. The trees were edited with FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

Pairwise comparisons between amino acid sequences from two isolates were performed using MegAlign pro with the Smith-Waterman Algorithm (Smith & Waterman 1981) to determine the percentage of variation expected within nodulation proteins.

Results

4.1 Isolation and culturing of *Rhizobium leguminosarum bv. trifolii*

Sixteen isolates from nine distinct clover root samples were selected for genetic analysis (**Table 3**). The geographic sites are shown in **Figure 4**.

The CFU (colony forming units) resulting per sample ranged from zero to 10^{10}. Several samples were tested repeatedly but yielded no viable isolates.

Table 3. The *Rlt* isolates and the red clover root sample origin.

Sample	Isolates	Sequences analysed	Province
141	141:6	*16S, nodA, nodD*	Medelpad
	141:7	*nodA, nodD*	
146	146:1	*16S*	Medelpad
	146:2	*16S, nodA, nodD*	
170	170:1	*16S, nodB, nodA, nodD*	Norrbotten
174	174:7	*16S, nodA, nodD*	Norrbotten
239	239:1	*16S*	Västergötland
	239:4	*16S, nodA, nodD*	
	239:13	*nodA, nodD*	
259	259:8	*nodA, nodB, nodD*	Gästrikland
	259:9	*nodA, nodD*	
270	270:7	*nodA, nodD*	Södermanland
291	291:2	*nodA, nodD*	Örebro
	291:10	*16S, nodA, nodD*	
293	293:8	*nodA, nodD*	Örebro
	293:13	*16S*	

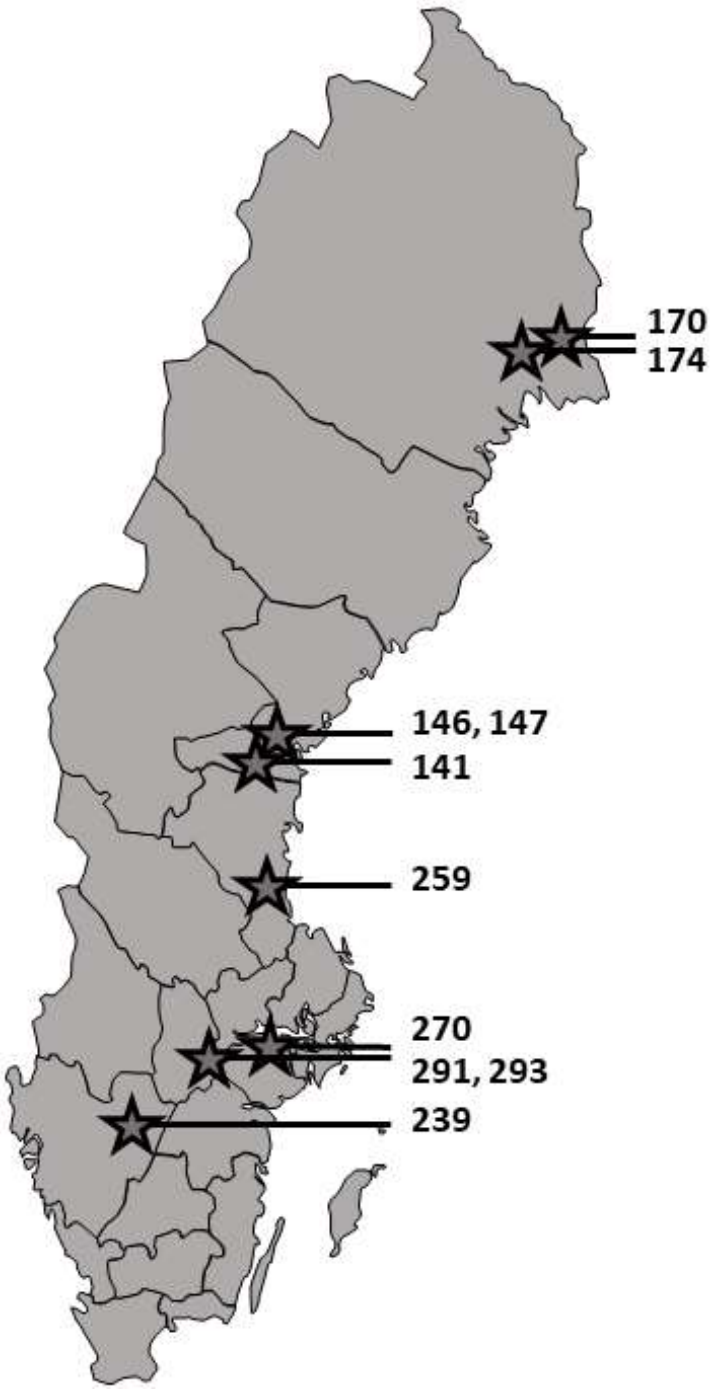


Figure 4. Geographic distribution of red clover root sample sites.

4.2 Data processing for phylogeny

Amino acid sequences were collected from NCBI GenBank with search parameters targeting *Rlt*, *Rlv*, and *Rlp*. The collected sequences were analysed to identify and remove identical entries. Identical entries were determined as 1. Having 100% sequence identity, 2. Being annotated as the same biovar, 3. Originate from samples collected from the same host species, 4. Samples collected from the same region. The remaining sequences were used for phylogenetic analysis.

The collected sequences were aligned using MegAlign Pro (v17.1.1) alongside translated nucleotide sequences from the isolates of this study. Phylogenetic trees were constructed with MegAlign Pro using the Maximum likelihood method RAxML with bootstrap analysis of 1000 replicates and edited for clarity using FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

4.3 In silico analysis

4.3.1 16S rRNA analysis

The taxonomy of nine of the isolates was confirmed through partial sequencing of the 16S rRNA gene. Multiple alignments constructed using MegAlign Pro with the Clustal Omega method showed that the sequences from all sampling points were identical, except for two sequences showing a one nucleotide deletion in the 3' end of the sequence. When analysing the chromatograms, the sequencing results of this specific region were often found ambiguous. Therefore, that deletion was not regarded as significant. When the sequences of four earlier Swedish isolates (Jambagi et al. 2023) were added to the alignment, three of them were found to be 100% identical, with one exception being 99.6% identical.

Sequences of the 16S rRNA gene were collected from NCBI GenBank with a query targeting the three biovars of *R. leguminosarum*.

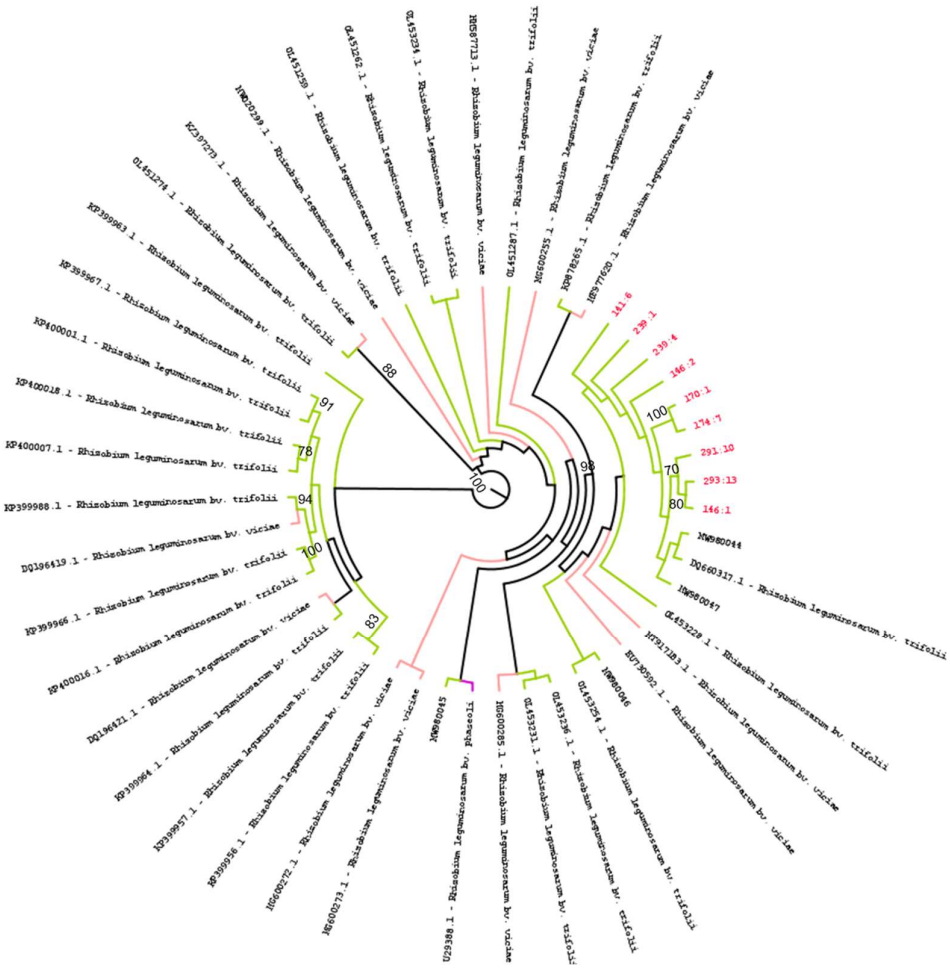


Figure 5. Phylogeny relationship of *Rlt* (green), *Rlv* (pink), and *Rlp* (purple) strains according to 16S rRNA sequences with isolates of this study marked in red. Sequences were aligned using MegAlign Pro by the Clustal Omega method. Tree constructed using the RAxML method with the GTR model with MegAlign Pro with bootstrap analysis based on 1000 replicates and edited using FigTree. Bootstrap values >70 are indicated.

When comparing the 16S rRNA sequences belonging to the three biovars collected from GenBank, the isolates from this study grouped together with very low branch support (**Figure 5**). This tree visualised the high similarity between the core genes of the biovars, where strains of different biovars can have a more similar core genome than two strains of the same biovar. The isolates are closely related to *Rlt* strains as well as *Rlv* strains, suggesting that the 16S rRNA gene is inefficient in determining the actual symbiotic behaviour of *R. leguminosarum.*

4.3.2 NodB

The PCR amplification of *nodB* from two isolates resulted in two fragments visible on agarose gel after electrophoresis: one of the expected approximate size of 650 nt, and one of an unexpectedly larger size at approximately 1000 nt (**Figure 6Error! Reference source not found.**A) and 900 nt (**Figure 6**B).

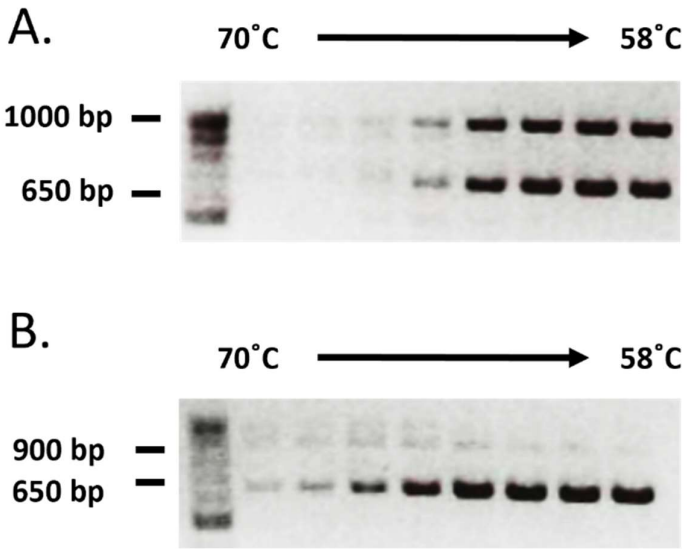


Figure 6. Temperature gradient PCR amplification of *nodB*. A) Isolate 170:1. B) Isolate 259:8

To determine whether the second larger band was an effect of too low temperature during the annealing step of the PCR program, causing primers to anneal non-specifically, a gradient PCR was performed with annealing temperatures from 70˚C down to 58˚C, and the products were visualised using gel electrophoresis. The results of the gradients showed that the amplified bands in isolate 170:1 remained largely the same at all annealing temperatures, suggesting that the primer binding is as efficient at the intended target sites as at the second unintended site. The visible bands of isolate 259:8 have another pattern, where the amplification of the intended target was successful at slightly higher temperatures than in 170:1. In contrast, the amplification of the unintended target was weak but approximately equal at all temperature steps. These variations indicate that there might be high sequence variability in the targeted primer regions of *nodB*, which

demands a redesign of the *nodB* primers if further analysis of the gene is to be conducted.

The *nodB* gene of two isolates was successfully sequenced. Unfortunately, attempts to sequence the large, unexpected, amplified fragment (**Figure 6A**) to elucidate its origin have been unsuccessful.

A NodB amino acid sequence pairwise comparison between the isolates revealed thirteen amino acid sequence variants with an identity of 93.6%. To confirm the percentage of identity between the isolates, pairwise alignments with the two isolates' NodA sequences, as well as NodD sequences, were performed, and the identity was found to be 95.5% and 93.5%, respectively. The level of variation in the pairwise comparisons remained consistent throughout the three Nod genes. We can conclude that these genes have a similar pattern of diversity. Sequences from additional isolates could confirm the pattern of sequence evolution.

4.3.3 NodA

The phylogenetic relationship of the NodA amino acid sequences of the isolates was determined (**Figure 7Error! Reference source not found.**). The NodA sequences among the Swedish isolates were highly similar, but a division could be seen geographically, as southern sequences were most similar. A grouping of the middle Sweden isolates is also seen. However, the bootstrap analysis provided no branch support value above 55, which demonstrates the low confidence of the groupings as the sequences are highly similar.

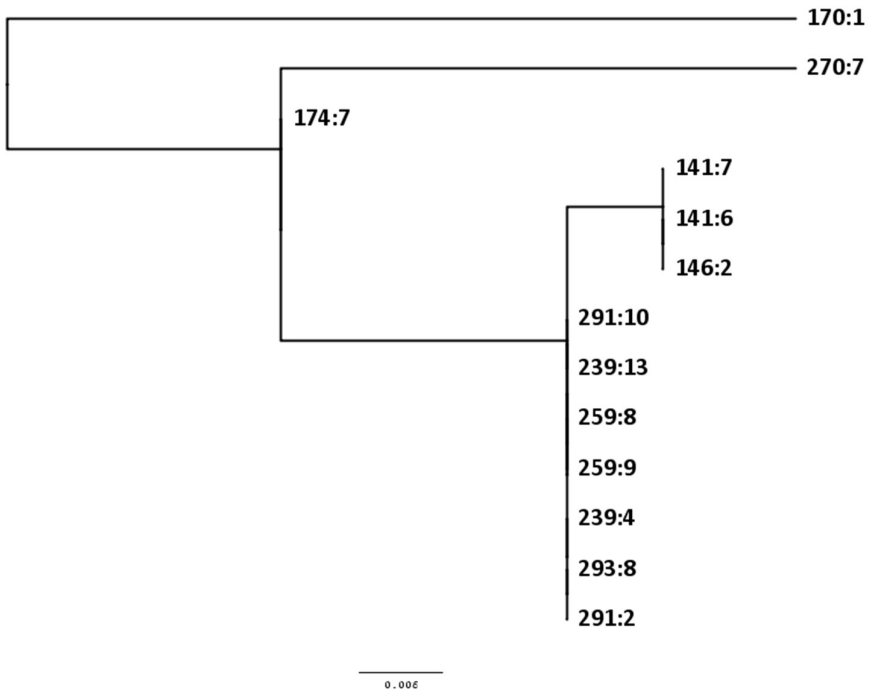


Figure 7. Phylogenetic tree of thirteen NodA sequences. Sequences were aligned using MegAlign Pro by the Clustal Omega method. Tree constructed using the RAxML method with the GTR model with MegAlign Pro with bootstrap analysis based on 1000 replicates and edited using FigTree. Bootstrap values >70 are indicated.

Amino acid sequences for NodA were collected from GenBank. The search returned 689 FASTA sequences longer than 150 aa, with the search parameters targeting NodA sequences belonging to *Rhizobium leguminosarum* and its biovars. The collected sequences were analysed to identify and remove identical entries. Identical entries were determined as: 1. Having 100% sequence identity, 2. Being annotated as the same biovar, 3. Originate from samples collected from the same host species, 4. Samples collected from the same region. The remaining sequences were used for phylogenetic analysis. When comparing international NodA sequences, there was no clear geographical grouping. The genotypes appeared to have a more random distribution (**Figure 8**).

There is overlap in sequences between the biovars when it comes to NodA (**Figure 8**) but not in NodD (**Figure 10**). There are three *Rlt* sequences that group with *Rlv*, as well as three *Rlv* that groups with *Rlt*.

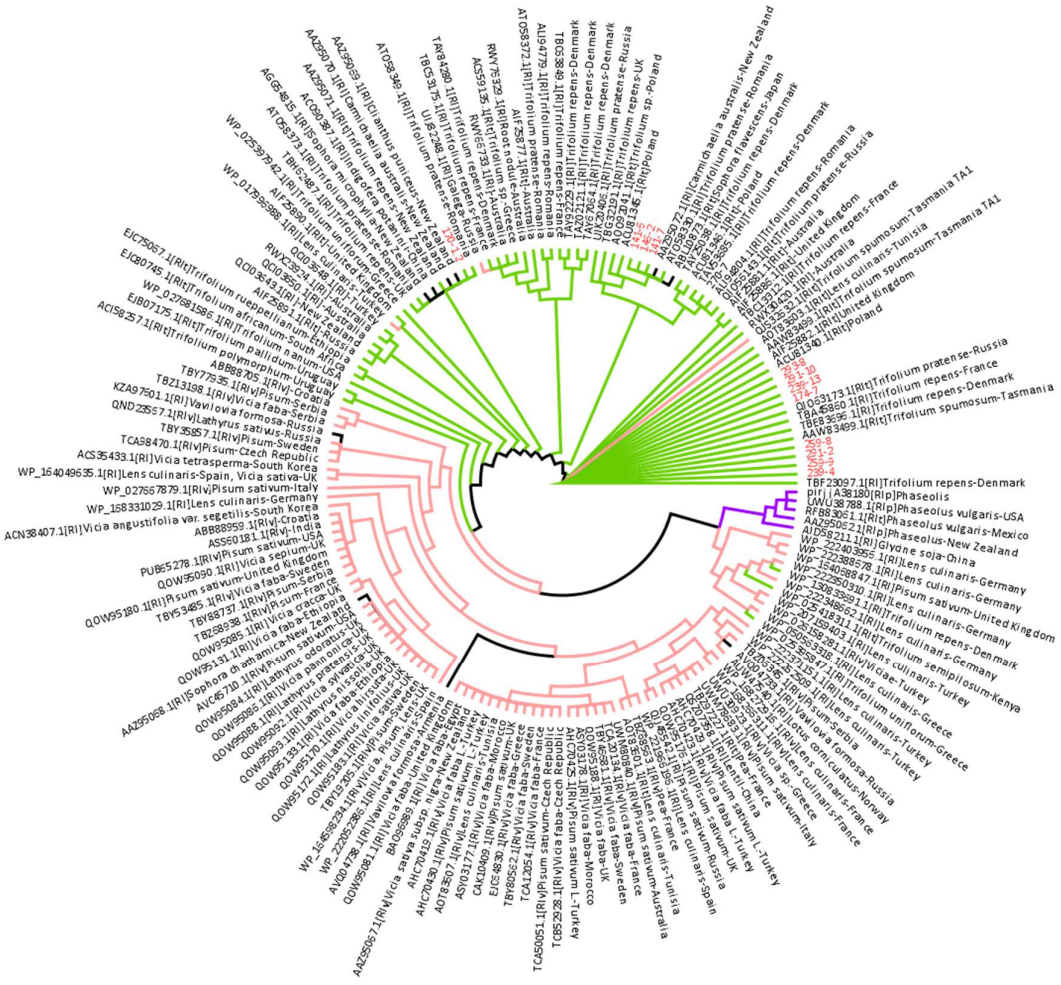


Figure 8. Phylogenetic tree of NodA sequences. Biovars *Rlt* (green), *Rlv* (pink), *Rlp* (purple), unclear biovar (black). Isolates of this study marked in red. Sequences were aligned using MegAlign Pro by the Clustal Omega method. Tree constructed using the RAxML method with the GTR model with MegAlign Pro and edited using FigTree.

4.3.4 NodD

The phylogenetic relationship of the NodD amino acid sequences of the isolates was determined. The variation between the isolates were mainly geographically divided, with northern and southern Swedish isolates sharing sequence variants (**Figure 9**). Isolate 270:7 was outside of this grouping, continuing the pattern of the NodA sequences.

Amino acid sequences for NodD were collected as described above, and 913 sequences longer than 200 amino acids were obtained. The collected sequences were analysed to identify and remove identical entries as described above.

The phylogenetic analysis of NodD protein sequences revealed a clear distinction between the biovars (**Figure 10**). The thirteen isolates place within the *Rlt* group, and follow the same pattern as seen in **Figure 9**, with two main groups, north and south, with isolate 270:7 as an exception.

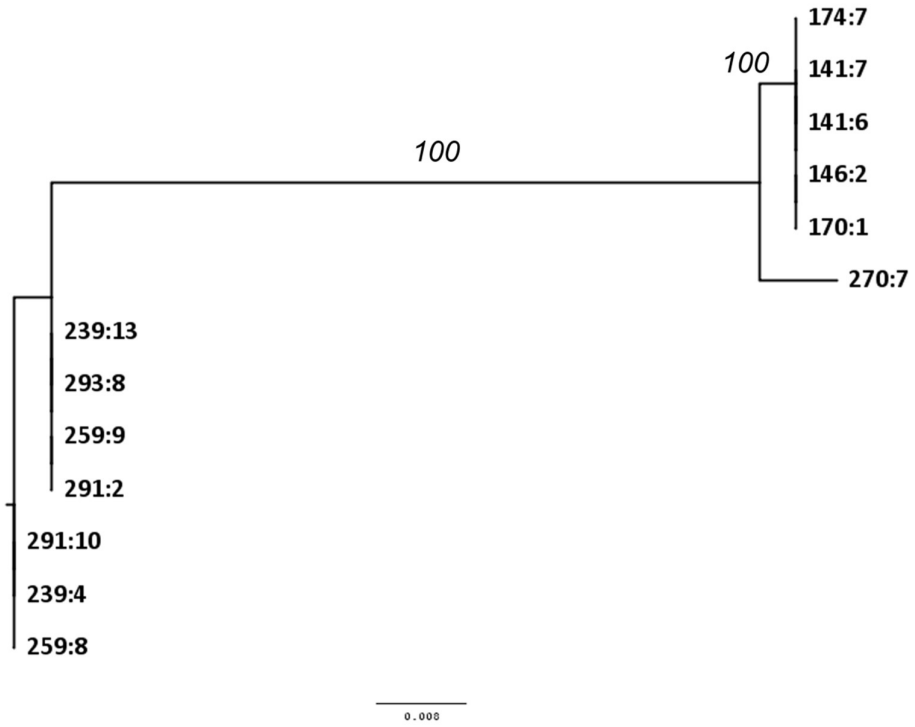


Figure 9. Phylogenetic tree of thirteen NodD sequences. Sequences were aligned using MegAlign Pro by the Clustal Omega method. Tree constructed using the RAxML method with the GTR model with MegAlign Pro with bootstrap analysis based on 1000 replicates and edited using FigTree. Bootstrap values >70 are indicated.

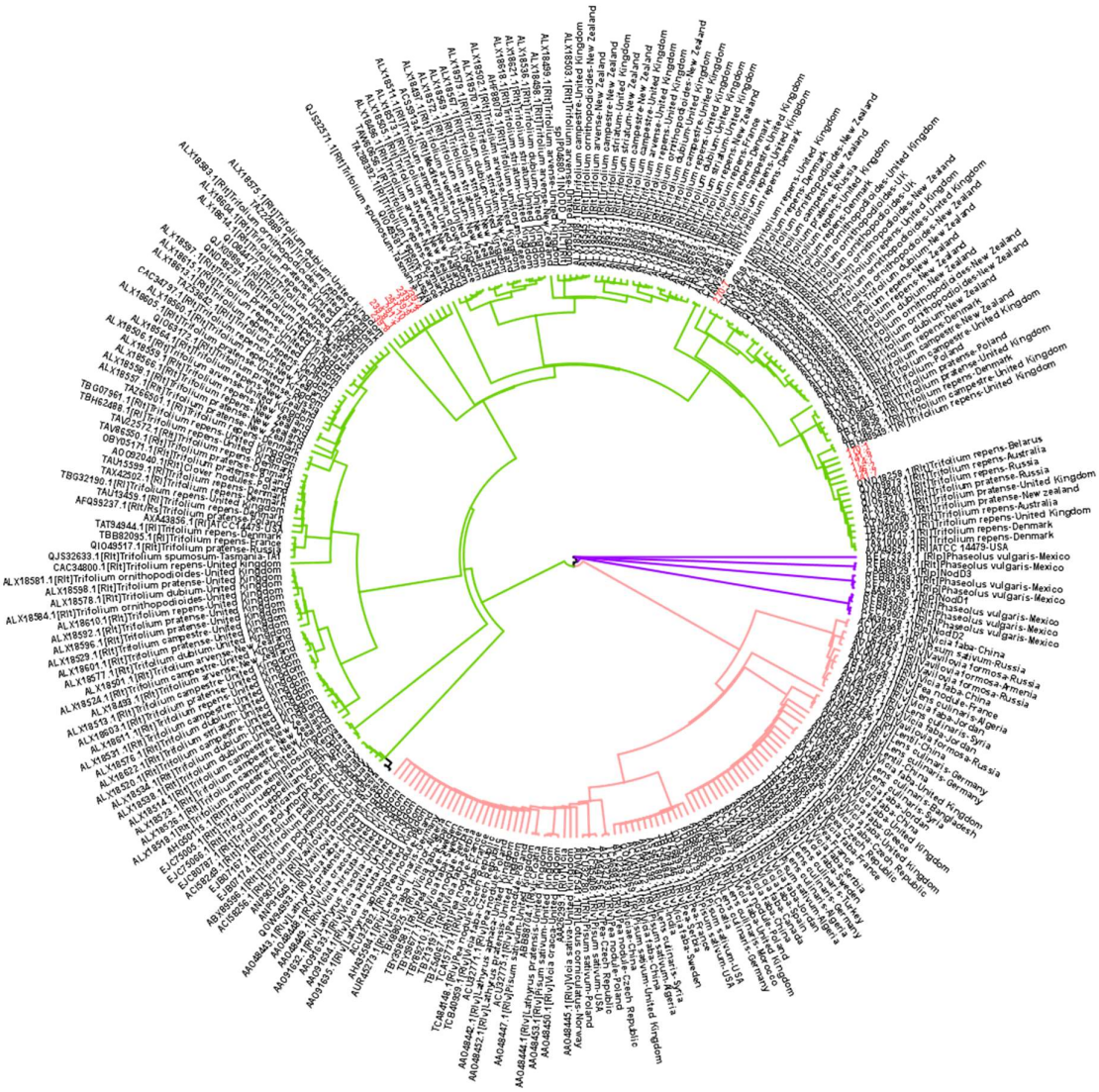


Figure 10. Phylogenetic tree of NodD sequences of biovars *Rlt* (green), *Rlv* (pink), *Rlp* (purple), unclear biovar (black), with the isolates of this study marked in red. Sequences were aligned using MegAlign Pro by the Clustal Omega method. Tree constructed using the RAxML method with the GTR model with MegAlign Pro and edited using FigTree.

Additional to the analysis of the canonical *nodD1*, attempts were made to detect a potential second *nodD2* sequence in the isolates' genomes with PCR amplification. Unfortunately, no detectable amplification of the potential gene was achieved.

Discussion

In this study, the genetic diversity of one chromosomal phylogenetic marker gene (16S rRNA) and four symbiotically relevant plasmid-located nodulation genes (*nodA, nodB, nodD1, and nodD2*) of sixteen isolates were analysed. The hypothesis of this study was that there would be no sequence divergence within the Swedish isolates. They are, after all, collected from the same host plants in the same country with a shared history of agricultural practices.

Based on the 16S rRNA sequences, it could be determined that the genetic diversity of red clover isolated *Rlt* strains is indeed low in Sweden (**Figure 5**). This needs to be further supported by additional core gene analysis, with genes such as *recA*, *atpD*, and *glnII* commonly used for the purpose (Cordeiro et al. 2017). When distinguishing between the biovars within the *R. leguminosarum* species complex, a single gene analysis is insufficient (Young et al. 2021). Thus, additional sequencing of other core genes, alternatively complete genome sequencing, is required for a proper taxonomy of strains. It is generally not possible to determine the biovar through 16S sequences alone, as illustrated in **Figure 5**, which demonstrates the species' close genetic relationship regardless of host preferences.

Despite the close-to-identical 16S sequences, some clear genetic variation could be seen among the *nod* genes. In published studies, the core genome generally shows a higher conservation level than the accessory genes (Young et al. 2006). The exchange of plasmids and symbiosis islands between strains and species creates a genetically dynamic rhizobia more adaptable to local hosts and environments. The genetic variants seen in the *nod* genes show that the sequences of these symbiotic genes are not conserved, even though the genes exist. Interestingly, the phylogenetic analyses of *nodA* and *nodD* do not follow the same pattern; the complete separation between biovars seen in *nodD* is less stringent in

nodA. This could be the result of the function of the proteins, as NodD is responsible for the recognition of the host flavonoids and thus requires more specificity for host infection than the NodA acyl transferase that performs a less species-specific role when it acetylates the non-reducing end of the NF backbone. The activity of NodA does not contribute to the species-specificity of the NFs, so the diversity of the *nodA* gene may be less regulated within the biovars.

One factor that cannot be ignored when analysing the geographic distribution of *Rlt* is the use of commercial inoculant strains during the last twelve decades. Based on a handful of strains, these inoculants have been used worldwide, and the expected frequent horizontal gene transfer between imported and native strains may have impacted the genetic relationship of the international *Rlt* population.

Both YEMA and Jensen's have been used as selective media extensively throughout the history of *Rhizobium* research, often with an addition of antibiotics targeting Gram-positive bacteria and fungi. This study omitted antibiotics to reduce the risk of selecting against potentially sensitive *Rlt* strains. The solid media used most successfully during the isolation process was Jensen's nitrogen-free media, which displayed a higher selectivity towards *Rlt* than YEMA. Using Jensen's also resulted in a reduced growth rate and slowed the production of EPS, which otherwise interferes with the handling of the isolates. The high growth rate on YEMA often caused cultures and their EPS to overflow standard Petri dishes. Although soil bacteria often display low generation times, with growth rates limited by a harsh environment, the nutrient-rich environment of the complex YEMA media permitted short doubling times. It supplied abundantly of source material for EPS production. Although previous studies have reported that in late-stage fermentation, the EPS is reduced by catabolism by the cells until the content is depleted (Sayyed et al. 2011), this was not the observation in this study. The rate of production of EPS slowed down with age (>30 days) but did not cease. Whether the nutrient content was altered during this time is not analysed here, but the volume continued increasing throughout culturing in temperatures of 28˚C, 20˚C, 4˚C, and 2˚C.

To detect whether this excessive EPS production is due to genetic variants common among Swedish isolates, sequencing of EPS-relevant genes such as the *rosR* regulator and the *pssA* glucose-IP-transferase could be informative (Janczarek et al. 2009). Previous work determined that gene duplication increased EPS production and competitiveness, thus, searching for multiplication events among these genes could be revealing. If this is a genetic variant of EPS genes established throughout Sweden, it would suggest that the climate, for example, may have selected for such a trait. EPS has been described previously as an important factor in cold resistance (Tribelli & López 2018), which would be an important trait for survival in Swedish soil and would explain continued production through lowered temperatures. In cold soil, the lack of moisture can be as harmful as the low temperatures, and hosts having a natural moisture-retaining root biofilm may have increased survival.

Although root nodules from several additional root samples were collected, attempts to isolate strains from other sample sites were unsuccessful. Possible reasons for this include poor survivability of the bacteria in the original root samples, bacteroids incapable of de-differentiating back into extracellular life, fatal nutrition limitation by the growth media, or improper preparation of media causing pH or osmotic shock. The stress of the isolation process may be simply overwhelming for some strains.

The problem of sequencing the large fragment produced when amplifying *nodB* could be resolved through cloning of the PCR fragments into recipient plasmids. The sequencing primers for the plasmid bind outside of the amplified fragment and could resolve a double binding site issue with the gene-specific primers. This cloning step could be performed with standard cloning kits such as CloneJET PCR Cloning Kit from Thermo Scientific.

The majority of *Rlt* strains with sequence data available have been isolated from the root nodules of various clover species, mainly *T. repens and T. pratense*, but also *T. polymorphum, T. spumosum, T. nanum, T. semipilosum, T. uniflorum, T. africanum*, and others. Root nodule extraction is the method recommended if the strains to be analysed should be effective in root colonization and survival within

the nodule, but it is not gentle on the bacteria and many strains may be lost in the process. This approach may limit the insight into the nodule microbial communities if the isolated bacteria always belong to the sturdier strains, whereas sensitive strains are discriminated against throughout the process. Unculturable rhizobia are included with population genetic analyses of the rhizosphere, however, analysis beyond DNA sequencing becomes impossible without the isolation of the strain.

Inside the nodule there is a mixed population of bacteria and differentiated bacteroids. When isolating the rhizobia, it is the non-differentiated cells that are retrieved. Among the *Rlp* strains, the limitations of the bacteroids incapable of returning to the extracellular lifestyle is less of an issue. The bacteroids are not terminally differentiated in the *Rlp* determinate nodules. However, *Rlv* and *Rlt* have indeterminate nodules populated with mainly terminally differentiated bacteroids. Such strains unable to return to a free-living state may have been studied during the long history of *Rlt* research, but if they have, they make up a minority of analysed strains. We get only a partial image of the populations in the nodule, and so the available data is skewed when it comes to analysis of the actual population.

Limitations to the phylogenetic analysis regarding the *nod* genes included the low number of unique protein sequences available in the NCBI database. Although the initial numbers of sequences for all genes appeared to approach several hundreds, and some above a thousand, the sequences fit for analysis were far fewer. The limited geographical distribution of the isolates reduced the confidence in the conclusions that could be drawn as the data was derived mainly from strains from European countries. Further, many available sequences resulted from only a few studies, many from the same samples and the same strains. For each collection of sequences, several separate searches had to be conducted to capture the diversity among *R. leguminosarum* while excluding unwanted data. If the data analysis conducted in this study was to be repeated, the biovars of additional *Rhizobium* species should be included to clarify the genetic variation.

A difficulty arising during sequence analysis was the inconsistency in naming isolates found in genetic databases. Naming conventions have changed over time and some classifications remain debated. *Rlt* and *Rlp.* were previously considered

separate species from *R. leguminosarum* and the redefinition of them as *R. leguminosarum* biovars was generally accepted in 1987 (Ramírez-Bahena et al. 2008). *Rlp* is often discussed as belonging to a separate species due to low genetic similarity to the other biovars, with some strains reclassified as *R. etli* or simply *R. phaseoli* (Tong et al. 2018). However, the genetic similarity or dissimilarity between species and biovars is a complicated subject.

The symbiosis genes may be highly similar between biovars, but not within the species (Rogel et al. 2011). Some of the different *Rhizobium* species share biovars with others. All three biovars bv. *phaseoli*, bv. *trifolii* and bv. *viciae* are found in *R. anhuiense*. Bv. *trifolii* is found in *R. laguerreae* which also contains bv. *viciae*, and in *R. hidalgonense* and *R. acidisoli* which also contain bv. *phaseoli*. Bv. *phaseoli* is present in *R. etli* and *R. sophoriradicis* (Tong et al. 2018). These examples, identified by multi locus sequence analysis, illustrate the difficulties of defining many of the known *Rhizobium* species and their strains without a wide range of genetic data. Perhaps it is time to update the naming conventions to reflect the partitioned state of the rhizobial genetic material.

The nitrogen-fixing ability of the rhizobia does not always correlate with the nodulation efficiency (Mendoza-Suárez et al. 2020). Inefficient nitrogen fixers might be the most competitive at colonizing root nodules and thus decrease the growth efficiency of the crop. To improve the chances of an inoculant strain being the main nodulating one, it must be applied generously to outcompete any native strains. Native strains can dominate the population of the nodules despite soil rhizobia content being less than 10^2 rhizobia per gram of soil. One source reporting inoculation failures with native rhizobia existing in concentrations as low as 50 rhizobia per gram of soil (Thies et al. 1991). The strain most successfully nodulating the host is commonly the one already successfully adapted to the local rhizosphere (Mwenda et al. 2023). The drastic lifestyle change for an inoculant strain from life in the moist and rich peat-based inoculant carrier to the soil and its diverse harsh conditions is a demanding process, and survivability of the non-native strains is often low, further reducing inoculant efficiency (Bashan et al. 2014; Santos et al. 2019). The native strain already occupies the available niches; a new strain may simply have no space to occupy. With the high genetic similarity of the Swedish

isolates, one can assume that they are highly established in the Swedish soil, and new strains may be difficult to establish unless they display significantly higher fitness and host compatibility.

Most legumes can enter symbiosis with one or several rhizobia. Fully understanding how to optimize both nodule competitiveness and nitrogen fixing and how to utilize these traits could have a large impact on agricultural sustainability. Further studies of the relationship between efficient nitrogen fixers and high nodule competitiveness among isolated strains could steer us toward developing a more efficient inoculant strain for Swedish legumes. Such a study focused on *Rlt* could be translated to other species of nitrogen fixers to optimise BNF. Reducing the need for application of exogenous nitrogen would lower costs both for the farmers and for the climate.

Conclusion

This study revealed that the genetic similarity within Swedish *Rlt* isolates is high compared to the diversity of international strains. It was determined that the isolates were highly similar regarding the 16S rRNA gene sequences but displayed sequence variations regarding the symbiotic *nod* genes *nodA*, *nodB*, and *nodD1*. In conclusion, these genetic variants suggest that there may be differences in competitiveness among these otherwise closely related isolates. Further genetic and phenotypic studies are necessary to elucidate the full effect of these genetic variants.

References

Acosta-Jurado, S., Fuentes-Romero, F., Ruiz-Sainz, J.-E., Janczarek, M. & Vinardell, J.-M. (2021). Rhizobial exopolysaccharides: genetic regulation of their synthesis and relevance in symbiosis with legumes. *International Journal of Molecular Sciences*, 22 (12), 6233. https://doi.org/10.3390/ijms22126233

Baedke, J., Fábregas-Tejeda, A. & Nieves Delgado, A. (2020). The holobiont concept before Margulis. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 334 (3), 149–155. https://doi.org/10.1002/jez.b.22931

Bashan, Y., de-Bashan, L.E., Prabhu, S.R. & Hernandez, J.-P. (2014). Advances in plant growth-promoting bacterial inoculant technology: formulations and practical perspectives (1998–2013). *Plant and Soil*, 378 (1), 1–33. https://doi.org/10.1007/s11104-013-1956-x

Bonfante, P. & Genre, A. (2010). Mechanisms underlying beneficial plant-fungus interactions in mycorrhizal symbiosis. *Nature Communications*, 1, 48. https://doi.org/10.1038/ncomms1046

Chen, X.-C., Feng, J., Hou, B.-H., Li, F.-Q., Li, Q. & Hong, G.-F. (2005). Modulating DNA bending affects NodD-mediated transcriptional control in *Rhizobium leguminosarum*. *Nucleic Acids Research*, 33 (8), 2540–2548. https://doi.org/10.1093/nar/gki537

Cordeiro, A.B., Ribeiro, R.A., Helene, L.C.F. & Hungria, M. (2017). *Rhizobium esperanzae* sp. *nov*., a N2-fixing root symbiont of *Phaseolus vulgaris* from Mexican soils. *International Journal of Systematic and Evolutionary Microbiology*, 67 (10), 3937–3945. https://doi.org/10.1099/ijsem.0.002225

Cui, H., Tsuda, K. & Parker, J.E. (2015). Effector-triggered immunity: from pathogen perception to robust defense. *Annual Review of Plant Biology*, 66, 487–511. https://doi.org/10.1146/annurev-arplant-050213-040012

D'Haeze, W. & Holsters, M. (2002). Nod factor structures, responses, and perception during initiation of nodule development. *Glycobiology*, 12 (6), 79R-105R. https://doi.org/10.1093/glycob/12.6.79R

Downie, J.A. (2010). The roles of extracellular proteins, polysaccharides and signals in the interactions of rhizobia with legume roots. *FEMS Microbiology Reviews*, 34 (2), 150–170. https://doi.org/10.1111/j.1574-6976.2009.00205.x

Erisman, J.W., Sutton, M.A., Galloway, J., Klimont, Z. & Winiwarter, W. (2008). How a century of ammonia synthesis changed the world. *Nature Geoscience*, 1 (10), 636–639. https://doi.org/10.1038/ngeo325

Evers, G.W. (2011). Forage Legumes: forage quality, fixed nitrogen, or both. *Crop Science*, 51 (2), 403–409. https://doi.org/10.2135/cropsci2010.06.0380

Ferguson, S., Major, A.S., Sullivan, J.T., Bourke, S.D., Kelly, S.J., Perry, B.J. & Ronson, C.W. (2020). *Rhizobium leguminosarum* bv. *trifolii* NodD2 enhances competitive nodule colonization in the clover-*Rhizobium* symbiosis. *Applied and Environmental Microbiology*, 86 (18), e01268-20. https://doi.org/10.1128/AEM.01268-20

Finn, J.A., Kirwan, L., Connolly, J., Sebastià, M.T., Helgadottir, A., Baadshaug, O.H., Bélanger, G., Black, A., Brophy, C., Collins, R.P., Čop, J., Dalmannsdóttir, S., Delgado, I., Elgersma, A., Fothergill, M., Frankow-Lindberg, B.E., Ghesquiere, A., Golinska, B., Golinski, P., Grieu, P., Gustavsson, A.-M., Höglind, M., Huguenin-Elie, O., Jørgensen, M., Kadziuliene, Z., Kurki, P., Llurba, R., Lunnan, T., Porqueddu, C., Suter, M., Thumm, U. & Lüscher, A. (2013). Ecosystem function enhanced by combining four functional types of plant species in intensively managed grassland mixtures: a 3-year continental-scale field experiment. *Journal of Applied Ecology*, 50 (2), 365–375. https://doi.org/10.1111/1365-2664.12041

Fred, E.B., Baldwin, I.L. & Mccoy, E. (1932). Root nodule bacteria and leguminous plants. *The University of Wisconsin Press,* https://digital.library.wisc.edu/1711.dl/VDA6UDQAKQJ748T

Giraud, E., Moulin, L., Vallenet, D., Barbe, V., Cytryn, E., Avarre, J.-C., Jaubert, M., Simon, D., Cartieaux, F., Prin, Y., Bena, G., Hannibal, L., Fardoux, J., Kojadinovic, M., Vuillet, L., Lajus, A., Cruveiller, S., Rouy, Z., Mangenot, S., Segurens, B., Dossat, C., Franck, W.L., Chang, W.-S., Saunders, E., Bruce, D., Richardson, P., Normand, P., Dreyfus, B., Pignol, D., Stacey, G., Emerich, D., Verméglio, A., Médigue, C. & Sadowsky, M. (2007). Legumes symbioses: absence of *Nod* genes in photosynthetic Bradyrhizobia. *Science (New York, N.Y.)*, 316 (5829), 1307–1312. https://doi.org/10.1126/science.1139548

Haag, A.F., Arnold, M.F.F., Myka, K.K., Kerscher, B., Dall'Angelo, S., Zanda, M., Mergaert, P. & Ferguson, G.P. (2013). Molecular insights into bacteroid development during *Rhizobium*-legume symbiosis. *FEMS Microbiology Reviews*, 37 (3), 364–383. https://doi.org/10.1111/1574-6976.12003

Harlow, B.E., Flythe, M.D., Kagan, I.A., Goodman, J.P., Klotz, J.L. & Aiken, G.E. (2020). Isoflavone supplementation, via red clover hay, alters the rumen microbial community and promotes weight gain of steers grazing mixed grass pastures. *PloS One*, 15 (3), e0229200. https://doi.org/10.1371/journal.pone.0229200

Hillel, D. (2008). 12. – Soil Biodiversity. In: Hillel, D. (ed.) *Soil in the Environment*. San Diego: Academic Press. 163–174. https://doi.org/10.1016/B978-0-12-348536-6.50017-4

Hollingsworth, R.I., Dazzo, F.B., Hallenga, K. & Musselman, B. (1988). The complete structure of the trifoliin A lectin-binding capsular polysaccharide of *Rhizobium trifolii* 843. *Carbohydrate Research*, 172 (1), 97–112. https://doi.org/10.1016/s0008-6215(00)90845-0

Hu, H., Liu, S., Yang, Y., Chang, W. & Hong, G. (2000). In *Rhizobium leguminosarum*, NodD represses its own transcription by competing with RNA polymerase for binding sites. *Nucleic Acids Research*, 28 (14), 2784–2793. https://doi.org/10.1093/nar/28.14.2784

Jambagi, S., Hodén, K.P., Öhlund, L. & Dixelius, C. (2023). Red clover root-associated microbiota is shaped by geographic location and choice of farming system. *Journal of Applied Microbiology*, 134, 1-17. https://doi.org/10.1093/jambio/lxad067

Janczarek, M., Jaroszuk-Sciseł, J. & Skorupska, A. (2009). Multiple copies of *rosR* and *pssA* genes enhance exopolysaccharide production, symbiotic competitiveness and clover nodulation in *Rhizobium leguminosarum* bv. *trifolii*. *Antonie Van Leeuwenhoek*, 96 (4), 471–486. https://doi.org/10.1007/s10482-009-9362-3

Jasniewski, A.J., Lee, C.C., Ribbe, M.W. & Hu, Y. (2020). Reactivity, mechanism, and assembly of the alternative nitrogenases. *Chemical Reviews*, 120 (12), 5107–5157. https://doi.org/10.1021/acs.chemrev.9b00704

Jensen, H.L. (1942). Nitrogen fixation in leguminous plants. II. Is symbiotic nitrogen fixation influenced by *Azotobacter*? *Proceedings of the Linnean Society of New South Wales*, 67, 205–212

Johansen, M., Lund, P. & Weisbjerg, M.R. (2018). Feed intake and milk production in dairy cows fed different grass and legume species: a meta-analysis. *Animal*, 12 (1), 66–75. https://doi.org/10.1017/S1751731117001215

Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S., Watanabe, A., Idesawa, K., Ishikawa, A., Kawashima, K., Kimura, T., Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A., Mochizuki, Y., Nakayama, S., Nakazaki, N., Shimpo, S., Sugimoto, M., Takeuchi, C., Yamada, M. & Tabata, S. (2000). Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. *DNA Research*, 7 (6), 331–338. https://doi.org/10.1093/dnares/7.6.331

Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T., Sasamoto, S., Watanabe, A., Idesawa, K., Iriguchi, M., Kawashima, K., Kohara, M., Matsumoto, M., Shimpo, S., Tsuruoka, H., Wada, T., Yamada, M. & Tabata, S. (2002). Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA Research*, 9 (6), 189–197. https://doi.org/10.1093/dnares/9.6.189

Kintl, A., Elbl, J., Lošák, T., Vaverková, M.D. & Nedělník, J. (2018). Mixed intercropping of wheat and white clover to enhance the sustainability of the conventional cropping system: effects on biomass production and leaching of mineral nitrogen. *Sustainability*, 10 (10), 3367. https://doi.org/10.3390/su10103367

Kjærgaard, T. (2003). A plant that changed the world: The rise and fall of clover 1000-2000. *Landscape Research*, 28, 41–49. https://doi.org/10.1080/01426390306531

Kobayashi, H., Naciri-Graven, Y., Broughton, W.J. & Perret, X. (2004). Flavonoids induce temporal shifts in gene-expression of *nod*-box controlled loci in *Rhizobium* sp. NGR234. *Molecular Microbiology*, 51 (2), 335–347. https://doi.org/10.1046/j.1365-2958.2003.03841.x

Konstantinidis, K.T. & Tiedje, J.M. (2004). Trends between gene content and genome size in prokaryotic species with larger genomes. *Proceedings of the National Academy of Sciences of the United States of America*, 101 (9), 3160–3165. https://doi.org/10.1073/pnas.0308653100

Larrainzar, E., Villar, I., Rubio, M.C., Pérez-Rontomé, C., Huertas, R., Sato, S., Mun, J.-H. & Becana, M. (2020). Hemoglobins in the legume-*Rhizobium* symbiosis. *The New Phytologist*, 228 (2), 472–484. https://doi.org/10.1111/nph.16673

Lassaletta, L., Billen, G., Grizzetti, B., Anglade, J. & Garnier, J. (2014). 50 year trends in nitrogen use efficiency of world cropping systems: the relationship between yield and nitrogen input to cropland. *JRC Publications Repository*. https://doi.org/10.1088/1748-9326/9/10/105011

Li, J., Goldschmidt-Clermont, M. & Timko, M.P. (1993). Chloroplast-encoded *chlB* is required for light-independent protochlorophyllide reductase activity in *Chlamydomonas reinhardtii*. *The Plant Cell*, 5 (12), 1817–1829. https://doi.org/10.1105/tpc.5.12.1817

Ling, J., Wang, H., Wu, P., Li, T., Tang, Y., Naseer, N., Zheng, H., Masson-Boivin, C., Zhong, Z. & Zhu, J. (2016). Plant nodulation inducers enhance horizontal gene transfer of *Azorhizobium caulinodans* symbiosis island. *Proceedings of the National Academy of Sciences*, 113 (48), 13875–13880. https://doi.org/10.1073/pnas.1615121113

López-Baena, F.J., Ruiz-Sainz, J.E., Rodríguez-Carvajal, M.A. & Vinardell, J.M. (2016). Bacterial molecular signals in the *Sinorhizobium fredii*-soybean

symbiosis. *International Journal of Molecular Sciences*, 17 (5), 755. https://doi.org/10.3390/ijms17050755

Madsen, L.H., Tirichine, L., Jurkiewicz, A., Sullivan, J.T., Heckmann, A.B., Bek, A.S., Ronson, C.W., James, E.K. & Stougaard, J. (2010). The molecular network governing nodule organogenesis and infection in the model legume *Lotus japonicus*. *Nature Communications*, 1 (1), 10. https://doi.org/10.1038/ncomms1009

Maillet, F., Poinsot, V., André, O., Puech-Pagès, V., Haouy, A., Gueunier, M., Cromer, L., Giraudet, D., Formey, D., Niebel, A., Martinez, E.A., Driguez, H., Bécard, G. & Dénarié, J. (2011). Fungal lipochitooligosaccharide symbiotic signals in arbuscular mycorrhiza. *Nature*, 469 (7328), 58–63. https://doi.org/10.1038/nature09622

Massa, F., Defez, R. & Bianco, C. (2022). Exploitation of plant growth promoting bacteria for sustainable agriculture: Hierarchical approach to link laboratory and field experiments. *Microorganisms*, 10 (5), 865. https://doi.org/10.3390/microorganisms10050865

Mayo, B., Vázquez, L. & Flórez, A.B. (2019). Equol: A bacterial metabolite from the daidzein isoflavone and its presumed beneficial health effects. *Nutrients*, 11 (9), 2231. https://doi.org/10.3390/nu11092231

Mazur, A., Stasiak, G., Wielbo, J., Kubik-Komar, A., Marek-Kozaczuk, M. & Skorupska, A. (2011). Intragenomic diversity of *Rhizobium leguminosarum* bv. *trifolii* clover nodule isolates. *BMC Microbiology*, 11, 123. https://doi.org/10.1186/1471-2180-11-123

McRose, D.L., Zhang, X., Kraepiel, A.M.L. & Morel, F.M.M. (2017). Diversity and activity of alternative nitrogenases in sequenced genomes and coastal environments. *Frontiers in Microbiology*, 8, 267. https://doi.org/10.3389/fmicb.2017.00267

Mendoza-Suárez, M.A., Geddes, B.A., Sánchez-Cañizares, C., Ramírez-González, R.H., Kirchhelle, C., Jorrin, B. & Poole, P.S. (2020). Optimizing *Rhizobium*-legume symbioses by simultaneous measurement of rhizobial competitiveness and N_2 fixation in nodules. *Proceedings of the National Academy of Sciences*, 117 (18), 9822–9831. https://doi.org/10.1073/pnas.1921225117

Mustonen, E.A., Tuori, M., Saastamoinen, I., Taponen, J., Wähälä, K., Saloniemi, H. & Vanhatalo, A. (2009). Equol in milk of dairy cows is derived from forage legumes such as red clover. *British Journal of Nutrition*, 102 (11), 1552–1556. https://doi.org/10.1017/S0007114509990857

Mwenda, G.M., Hill, Y.J., O'Hara, G.W., Reeve, W.G., Howieson, J.G. & Terpolilli, J.J. (2023). Competition in the *Phaseolus vulgaris-Rhizobium* symbiosis and the role of resident soil rhizobia in determining the outcomes of inoculation. *Plant and Soil*. https://doi.org/10.1007/s11104-023-05903-0

O'Callaghan, M., Ballard, R.A. & Wright, D. (2022). Soil microbial inoculants for sustainable agriculture: Limitations and opportunities. *Soil Use and Management*, 38 (3), 1340–1369. https://doi.org/10.1111/sum.12811

Okazaki, S., Tittabutr, P., Teulet, A., Thouin, J., Fardoux, J., Chaintreuil, C., Gully, D., Arrighi, J.-F., Furuta, N., Miwa, H., Yasuda, M., Nouwen, N., Teaumroong, N. & Giraud, E. (2016). *Rhizobium*-legume symbiosis in the absence of Nod factors: two possible scenarios with or without the T3SS. *The ISME Journal*, 10 (1), 64–74. https://doi.org/10.1038/ismej.2015.103

Ott, T., van Dongen, J. T., Günther, C., Krusell, L., Desbrosses, G., Vigeolas, H., Bock, V., Czechowski, T., Geigenberger, P., & Udvardi, M. K. (2005). Symbiotic leghemoglobins are crucial for nitrogen fixation in legume root nodules but not for general plant growth and development. *Current Biology*, 15(6), 531–535. https://doi.org/10.1016/j.cub.2005.01.042

Palhares Farias, T., de Melo Castro, E., Marucci Pereira Tangerina, M., Quintino da Rocha, C., Brito Bezerra, C.W. & de Souza Moreira, F.M. (2022). Rhizobia exopolysaccharides: promising biopolymers for use in the formulation of plant inoculants. *Brazilian Journal of Microbiology*, 53 (4), 1843–1856. https://doi.org/10.1007/s42770-022-00824-z

Ramírez-Bahena, M.H., García-Fraile, P., Peix, A., Valverde, A., Rivas, R., Igual, J.M., Mateos, P.F., Martínez-Molina, E. & Velázquez, E. (2008). Revision of the taxonomic status of the species *Rhizobium leguminosarum* (Frank 1879) Frank 1889AL, *Rhizobium phaseoli* Dangeard 1926AL and *Rhizobium trifolii* Dangeard 1926AL. *R. trifolii* is a later synonym of *R. leguminosarum*. Reclassification of the strain *R. leguminosarum* DSM 30132 (=NCIMB 11478) as *Rhizobium pisi* sp. nov. *International Journal of Systematic and Evolutionary Microbiology*, 58 (11), 2484–2490. https://doi.org/10.1099/ijs.0.65621-0

Rascio, N. & La Rocca, N. (2013). Biological nitrogen fixation. In: *Reference Module in Earth Systems and Environmental Sciences*. *Elsevier*. https://doi.org/10.1016/B978-0-12-409548-9.00685-0

Reeve, W., O'Hara, G., Chain, P., Ardley, J., Bräu, L., Nandesena, K., Tiwari, R., Malfatti, S., Kiss, H., Lapidus, A., Copeland, A., Nolan, M., Land, M., Ivanova, N., Mavromatis, K., Markowitz, V., Kyrpides, N., Melino, V., Denton, M., Yates, R. & Howieson, J. (2010). Complete genome sequence of *Rhizobium leguminosarum* bv *trifolii* strain WSM2304, an effective microsymbiont of the South American clover *Trifolium polymorphum*. *Standards in Genomic Sciences*, 2 (1), 66–76. https://doi.org/10.4056/sigs.44642

Reinhold-Hurek, B., Bünger, W., Burbano, C.S., Sabale, M. & Hurek, T. (2015). Roots shaping their microbiome: global hotspots for microbial activity. *Annual Review of Phytopathology*, 53, 403–424. https://doi.org/10.1146/annurev-phyto-082712-102342

Ren, B., Wang, X., Duan, J. & Ma, J. (2019). Rhizobial tRNA-derived small RNAs are signal molecules regulating plant nodulation. *Science*, 365 (6456), 919–922. https://doi.org/10.1126/science.aav8907

Riggi, L.G.A., Lundin, O. & Berggren, Å. (2021). Mass-flowering red clover crops have positive effects on bumblebee richness and diversity after bloom. *Basic and Applied Ecology*, 56, 22–31. https://doi.org/10.1016/j.baae.2021.06.001

Rogel, M.A., Ormeño-Orrillo, E. & Martinez Romero, E. (2011). Symbiovars in rhizobia reflect bacterial adaptation to legumes. *Systematic and Applied Microbiology*, 34 (2), 96–104. https://doi.org/10.1016/j.syapm.2010.11.015

Roy, P., Achom, M., Wilkinson, H., Lagunas, B. & Gifford, M.L. (2020). Symbiotic outcome modified by the diversification from 7 to over 700 nodule-specific cysteine-rich peptides. *Genes*, 11 (4), 348. https://doi.org/10.3390/genes11040348

Santos, M.S., Nogueira, M.A. & Hungria, M. (2019). Microbial inoculants: reviewing the past, discussing the present and previewing an outstanding future for the use of beneficial bacteria in agriculture. *AMB Express*, 9 (1), 205. https://doi.org/10.1186/s13568-019-0932-0

Sayyed, R.Z., Jamadar, D.D. & Patel, P.R. (2011). Production of exo-polysaccharide by *Rhizobium* sp. *Indian Journal of Microbiology*, 51 (3), 294–300. https://doi.org/10.1007/s12088-011-0115-4

Schell, M.A. (1993). Molecular biology of the LysR family of transcriptional regulators. *Annual Review of Microbiology*, 47, 597–626. https://doi.org/10.1146/annurev.mi.47.100193.003121

Schmeisser, C., Liesegang, H., Krysciak, D., Bakkou, N., Le Quéré, A., Wollherr, A., Heinemeyer, I., Morgenstern, B., Pommerening-Röser, A., Flores, M.,

Palacios, R., Brenner, S., Gottschalk, G., Schmitz, R.A., Broughton, W.J., Perret, X., Strittmatter, A.W. & Streit, W.R. (2009). *Rhizobium* sp. strain NGR234 possesses a remarkable number of secretion systems. *Applied and Environmental Microbiology*, 75 (12), 4035–4045. https://doi.org/10.1128/AEM.00515-09

Shah, A., Nazari, M., Antar, M., Msimbira, L.A., Naamala, J., Lyu, D., Rabileh, M., Zajonc, J. & Smith, D.L. (2021). PGPR in agriculture: A sustainable approach to increasing climate change resilience. *Frontiers in Sustainable Food Systems*, 5. https://www.frontiersin.org/articles/10.3389/fsufs.2021.667546 [2023-01-05]

Shamseldin, A., Abdelkhalek, A. & Sadowsky, M.J. (2017). Recent changes to the classification of symbiotic, nitrogen-fixing, legume-associating bacteria: a review. *Symbiosis*, 71 (2), 91–109. https://doi.org/10.1007/s13199-016-0462-3

Sievers, F. & Higgins, D.G. (2018). Clustal Omega for making accurate alignments of many protein sequences. *Protein Science*, 27 (1), 135–145. https://doi.org/10.1002/pro.3290

Smith, T.F. & Waterman, M.S. (1981). Identification of common molecular subsequences. *Journal of Molecular Biology*, 147 (1), 195–197. https://doi.org/10.1016/0022-2836(81)90087-5

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30 (9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Sullivan, J.T., Patrick, H.N., Lowther, W.L., Scott, D.B. & Ronson, C.W. (1995). Nodulating strains of *Rhizobium loti* arise through chromosomal symbiotic gene transfer in the environment. *Proceedings of the National Academy of Sciences of the United States of America*, 92 (19), 8985–8989

Thies, J.E., Singleton, P.W. & Bohlool, B.B. (1991). Influence of the size of indigenous rhizobial populations on establishment and symbiotic performance of introduced rhizobia on field-grown legumes. *Applied and Environmental Microbiology*, 57 (1), 19–28.

Tong, W., Li, X., Huo, Y., Zhang, L., Cao, Y., Wang, E., Chen, W., Tao, S. & Wei, G. (2018). Genomic insight into the taxonomy of *Rhizobium* genospecies that nodulate *Phaseolus vulgaris*. *Systematic and Applied Microbiology*, 41 (4), 300–310. https://doi.org/10.1016/j.syapm.2018.03.001

Tosar, V., Rousseau, G., Decruyenaere, V., Beckers, Y., Larondelle, Y. & Froidmont, E. (2021). Red clover-rich grassland increases equol concentration in eggs from free-range laying hens. *British Poultry Science*, 62 (6), 840–845. https://doi.org/10.1080/00071668.2021.1929069

Tribelli, P.M. & López, N.I. (2018). Reporting key features in cold-adapted bacteria. *Life*, 8 (1), 8. https://doi.org/10.3390/life8010008

Wanke, A., Malisic, M., Wawra, S. & Zuccaro, A. (2021). Unraveling the sugar code: the role of microbial extracellular glycans in plant-microbe interactions. *Journal of Experimental Botany*, 72 (1), 15–35. https://doi.org/10.1093/jxb/eraa414

Weinert-Nelson, J.R., Ely, D.G., Flythe, M.D., Hamilton, T.A., May, J.B., Ferrell, J.L., Hamilton, M.C., Jacks, W.L. & Davis, B.E. (2023). Red clover supplementation modifies rumen fermentation and promotes feed efficiency in ram lambs. *Journal of Animal Science*, skad036. https://doi.org/10.1093/jas/skad036

Wheatley, R.M., Ford, B.L., Li, L., Aroney, S.T.N., Knights, H.E., Ledermann, R., East, A.K., Ramachandran, V.K. & Poole, P.S. (2020). Lifestyle adaptations of *Rhizobium* from rhizosphere to symbiosis. *Proceedings of the National*

Academy of Sciences, 117 (38), 23823–23834. https://doi.org/10.1073/pnas.2009094117

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S. & Madden, T.L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, 13, 134. https://doi.org/10.1186/1471-2105-13-134

Young, J.P.W., Crossman, L.C., Johnston, A.W., Thomson, N.R., Ghazoui, Z.F., Hull, K.H., Wexler, M., Curson, A.R., Todd, J.D., Poole, P.S., Mauchline, T.H., East, A.K., Quail, M.A., Churcher, C., Arrowsmith, C., Cherevach, I., Chillingworth, T., Clarke, K., Cronin, A., Davis, P., Fraser, A., Hance, Z., Hauser, H., Jagels, K., Moule, S., Mungall, K., Norbertczak, H., Rabbinowitsch, E., Sanders, M., Simmonds, M., Whitehead, S. & Parkhill, J. (2006). The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome Biology*, 7 (4), R34. https://doi.org/10.1186/gb-2006-7-4-r34

Young, J.P.W., Moeskjær, S., Afonin, A., Rahi, P., Maluk, M., James, E.K., Cavassim, M.I.A., Rashid, M.H.-O., Aserse, A.A., Perry, B.J., Wang, E.T., Velázquez, E., Andronov, E.E., Tampakaki, A., Flores Félix, J.D., Rivas González, R., Youseif, S.H., Lepetit, M., Boivin, S., Jorrin, B., Kenicer, G.J., Peix, Á., Hynes, M.F., Ramírez-Bahena, M.H., Gulati, A. & Tian, C.-F. (2021). Defining the *Rhizobium leguminosarum* species complex. *Genes*, 12 (1), 111. https://doi.org/10.3390/genes12010111

Zarrabian, M., Montiel, J., Sandal, N., Ferguson, S., Jin, H., Lin, Y.-Y., Klingl, V., Marín, M., James, E.K., Parniske, M., Stougaard, J. & Andersen, S.U. (2022). A Promiscuity Locus Confers *Lotus burttii* nodulation with rhizobia from five different genera. *Molecular Plant-Microbe Interactions,* 35 (11), 1006–1017. https://doi.org/10.1094/MPMI-06-22-0124-R0

Popular science summary

Every year the crops of the world are supplied with a hundred million tons of nitrogen in the form of inorganic fertilizer which is mainly produced using fossil fuels. Regrettably, half of it is leached out of the fields and is causing pollution, eutrophication, and increased greenhouse gas emissions.

Thankfully, there is another way to provide the crops with essential nitrogen. Certain soil bacteria in symbiosis with legumes can transform ("fix") the nitrogen in the air into organic forms that plants can utilize. One group of legumes excellent for this purpose is the clovers, who enter a symbiotic relationship with the species *Rhizobium leguminosarum* variant *trifolii*. The symbiosis requires complex molecular communication and several hundred symbiosis genes to function. Here we attempt to sample the Swedish diversity of red clover-associated *Rhizobium* and determine their level of genetic variation. For this we collected 16 *Rhizobium* isolates from clover roots and performed gene sequencing of one species-determining gene and four symbiosis genes.

The analysed Swedish *Rhizobium* isolates were highly genetically similar; the species-determining gene sequences were near identical, showing that all isolates were closely related. However, the symbiosis-specific genes showed some variability, and that variability could be roughly traced to the sampling location, to north and south of Sweden. By comparing these *Rhizobium* sequences to published sequences worldwide we could determine that the Swedish isolates are indeed very closely related and, considering the similarities, can possibly share a single geographical origin. To understand the variability and its effects in the symbiosis-specific genes, we need to extend our work to include further genetical and functional analysis of our *Rhizobium*.

Acknowledgements

I would like to thank my supervisor Christina Dixelius for guidance throughout this project. I would also like to thank my examiner, Jens Sundström, who gave up skiing for my sake.

Further, I would like to thank my opponent Lina Karm Togo for support throughout this process, as well as the entire Dixelius group of

Shailja Singh

Xinyi Hu

Fredrik Dölfors

and special thanks to Shridhar Jambagi for kindly providing me with the beautiful clover root photographs.