Resource allocation in the legume-rhizobia symbiosis: an integration of modelling and experimental approaches



Annet Westhoek
Trinity College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Trinity 2017

Resource allocation in the legume-rhizobia symbiosis: an integration of modelling and experimental approaches

Annet Westhoek, Trinity College, University of Oxford, Trinity 2017

A thesis submitted for the degree of Doctor of Philosophy

Abstract

The symbiosis between plants of the legume family and nitrogen-fixing rhizobia underpins global food security. Legume crops are a major source of protein in human diets, either directly or indirectly as feed for livestock. Application of inoculant rhizobial strains is common practice in many areas, as plant growth is often nitrogen limited and the symbiosis can significantly enhance yields. However, rhizobial strains and outcomes of the symbiosis vary widely. This variation has also been studied by evolutionary biologists interested in the stability of mutualisms. They proposed that plants may prevent establishing symbioses with ineffective strains (partner choice), or provide them with fewer resources (sanctioning). I studied both mechanisms, combining modelling and experimental approaches.

Mathematical modelling was used to predict how plants should allocate resources to maximise growth rates, depending on rhizobial nitrogen provision and carbon requirements and on soil nitrogen conditions. The use of marked mutant strains – easily distinguishable and differing in a single rhizobial characteristic – overcame previous experimental difficulties. It was found that pea (*Pisum sativum* L.) plants are not able to exert partner choice, but do sanction in a more complex way than was previously established. In line with model predictions, resources were preferentially allocated to the single – best available – strain, so that resources allocated to an intermediate-fixing strain depended on whether or not a strain providing more nitrogen was available. Contrary to model predictions, there was no indication of discrimination based on rhizobial carbon requirements.

The results cannot be explained by resource allocation in proportion to nitrogen received, and indicate systemic integration of information from different nodules. I formulate a hypothesis about the underlying plant regulatory mechanisms, and discuss implications of the results for selecting inoculant strains and enhancing yields in the field. Future work will rely on further integration of theoretical and applied methods and perspectives.

Acknowledgements

According to an African proverb, it takes a village to raise a child. The same holds true for educating a researcher. Forgive me if I fail to name everyone who stayed in that village while I was there, but let me thank everyone who made that village what it has been.

First and foremost, I would like to thank my supervisors Dr Lindsay A. Turnbull, Prof. Phil S. Poole and Dr Gail M. Preston. I am deeply grateful for the trust you have given me in willing to supervise this project which sat between fields, for the freedom you've given me in shaping it, for the unique ways you contributed to it, and for the valuable lessons I learned from each of you. Gail, you have an aptitude for sharing little gems of advice at exactly the right time, which has not only helped to produce this thesis, but also enriched the time I spent working on it. Phil, listening to you thinking out loud, producing 5,000 new ideas while displaying an encyclopaedia of knowledge, has been tremendously inspiring. Lindsay, somehow you manage to smuggle bits of wisdom about life into conversations about the appropriateness of a statistical method. You've been critical in an uplifting way, and I've always left your office just that little wiser and lighter.

I am grateful for funding from the Engineering and Physical Sciences Research Council (EP/G03706X/1), and for the support of the Systems Biology Doctoral Training Centre. The first year of courses gave me the skills and knowledge, and the confidence to acquire more of each of them, to pursue this endeavour. The contact with a wide range of scientists and the opportunity to try things formed the perfect environment to define and design my own project – a luxury not given to many starting researchers. I am also indebted to Trinity College, for generous support and for being my home in Oxford.

The groups of all my supervisors provided a rich research environment, but a very special thank you goes to the Poole group – you've taught me everything I know about lab work. Together you exude a mix of passion, diligence, warmth and humour which makes tedious work so much lighter. Marcela Mendoza-Suárez, Carmen Sánchez-Cañizares, Barney Geddes, Isabel Webb, Vinoy Ramachandran, Andrzej Tkacz, Joshua Roworth, Rachel Wheatley, Khusboo Borah, Alison East, and everyone who joined later: thank you! Elsa Field, Finn Rehling and Megan Griffiths; your injections of curiosity and enthusiasm during your projects have been uplifting. John Baker, Helen Prescott, Gem Toes and other staff at the Department of Plant Sciences; none of the research would be possible without you, and you have made me smile in all our interactions. I owe Prof. Ken E. Giller and Dr Peter A. Leffelaar from Wageningen University for my interest in both legumes and modelling. Contact with you, Prof Ton Bisseling, Dr René Geurts, and others from your groups has been truly helpful, and reaffirmed that studying legumes means studying food security. A special thank you also to Dr Neil Dalchau and Dr Drew Purves for inspiring discussions and the visits to Microsoft Research, which have given me another perspective on the research community. I would also like to thank Prof. E. Toby Kiers for helpful comments and interesting discussions.

Lastly, I would like to thank everyone who was not directly involved in this project, but without whom pursuing it would not have been possible. Eleonara (Ellie) Aldegheri and Susie Crow; dancing in your classes has meant more to me than you can imagine. Friends, both in Oxford and abroad: thank you for sharing with me. My family has raised me to be where I am now. Thank you for your love, relativism, and unwavering support.

I trust that in going forward, everything I have learned in that proverbial village will stay with me.

Table of contents

ABSTRACT **3**

ACKNOWLEDGEMENTS **5**

TABLE OF CONTENTS **8**

1 INTRODUCTION **15**

1.1 Overview 15

1.2 The legume-rhizobia symbiosis 17

1.2.1 Organisms involved *17*

1.2.2 Establishment of the symbiosis *18*

1.2.3 Nutrient exchange in the symbiosis *20*

1.2.4 Outcome of the symbiosis *22*

1.3 Different perspectives on the legume-rhizobia symbiosis 24

1.3.1 The legume-rhizobia symbiosis from an agricultural perspective *24*

1.3.2 The legume-rhizobia symbiosis from an evolutionary biology perspective *25*

1.3.3 Combining the two perspectives *26*

1.4 Carbon and nitrogen metabolism 30

1.4.1 Carbon and nitrogen: the backbone of life *30*

1.4.2 Plant nitrogen uptake and assimilation *30*

1.5 Mathematical modelling of the legume-rhizobia symbiosis 33

1.5.1 Why mathematical modelling can be useful *33*

1.5.2 How the mathematical modelling is used *33*

1.6 Experimental model system 37

1.7 Aims and objectives 38

1.8 References 40

2 GENERAL MATERIALS AND METHODS **57**

2.1 Overview 57

2.2 PLANT GROWTH 58
2.2.1 Seed sterilisation and germination 58
2.2.2 Nitrogen free nutrient solution 58
2.2.3 Planting and growth conditions 58
2.2.4 Harvest 60
2.3 BACTERIAL STRAINS AND CULTURE CONDITIONS 61
2.3.1 Bacterial strains 61
2.3.2 Culture conditions 62
2.3.3 Preparation of rhizobial inocula 63
2.3.4 Storage 63
2.4 POST-HARVEST PROCEDURES 64
2.4.1 Shoot dry weight 64
2.4.2 Measuring nitrogen fixation by acetylene reduction 64
2.4.3 Distinguishing strains based on GusA and CelB staining 65
2.4.4 Photographing root systems 67
2.4.5 Microscopy 67
2.4.6 Counting nodules and measuring nodule area 67
2.5 REFERENCES 69
3 PARTNER CHOICE IN THE LEGUME-RHIZOBIA SYMBIOSIS 71
3.1 OVERVIEW 71
3.2 ABSTRACT 71
3.3 INTRODUCTION 73
3.4 RESULTS 76
3.4.1 Partner choice 77
3.4.2 Nodule size 79
3.4.3 Total number of nodules 79
3.4.4 Shoot mass 80
3.4.5 Frequency of mixed nodules 80

3.4.6 Fitness of rhizobial strains within mixed nodules *83*

3.5 DISCUSSION 84

3.6 METHODS 89

3.6.1 Bacterial strains and culture conditions *89*

3.6.2 Plant growth *90*

3.6.3 Experimental designs *90*

3.6.4 Harvest *91*

3.6.5 Data collection *92*

3.6.6 Statistical analyses *92*

3.7 REFERENCES 96

4 MATHEMATICAL MODELLING OF THE LEGUME-RHIZOBIA SYMBIOSIS 103

4.1 OVERVIEW 103

4.2 ABSTRACT 103

4.3 INTRODUCTION 105

4.4 METHODS 108

4.4.1 Model description *108*

4.4.1.1 Basic model: constant soil nitrogen and a single rhizobial strain 108

4.4.1.2 Extension with variable soil nitrogen concentration 110

4.4.1.3 Extension with multiple rhizobial strains 111

4.4.2 Model implementation *112*

4.5 RESULTS 117

4.5.1 Allocation to a rhizobial strain depends on its effectiveness and efficiency and on plant nitrogen status *117*

4.5.2 Allocation to a rhizobial strain depends on other strains present on the same plant *123*

4.6 DISCUSSION 129

4.7 REFERENCES 133

5 RESOURCE ALLOCATION IN THE LEGUME-RHIZOBIA SYMBIOSIS 139

5.1 OVERVIEW 139

5.2 ABSTRACT 139
5.3 INTRODUCTION 141
5.4 RESULTS 144
5.4.1 Characterisation of intermediate-fixing strain LMB827 144
5.4.2 Resource allocation to an intermediate-fixing strain depends on whether or not a more effective strain is available to the plant. 145
5.4.3 Reduced resource allocation results in reduced nitrogen fixation rates of the intermediate-fixing strain 149
5.4.4 Soil nitrate availability affects resource allocation to nodules in a strain-dependent manner 151
5.4.5 Resource allocation to a strain with increased carbon costs does not depend on whether or not a more efficient strain is available to the plant. 152
5.5 DISCUSSION 156
5.6 METHODS 160
5.6.1 Bacterial strains and culture conditions 160
5.6.2 Plant growth 160
5.6.3 Experimental designs 161
5.6.4 Harvest 163
5.6.5 Data collection 164
5.6.6 Statistical analyses 164
5.7 REFERENCES 166
6 GENERAL DISCUSSION 171
6.1 OVERVIEW 171
6.2 SUMMARY OF RESULTS 172
6.3 IMPLICATIONS AND FUTURE WORK 174
6.3.1 Plant regulation of sanctioning 174
6.3.2 The legume-rhizobia symbiosis from an evolutionary biology perspective 179
6.3.3 The legume-rhizobia symbiosis from an agricultural perspective 183
6.4 GENERAL CONCLUSIONS 187

6.5 References 188

APPENDIX A. PUBLISHED MANUSCRIPT CHAPTER 3 199

Overview 199

APPENDIX B. SUPPLEMENTARY MATERIAL CHAPTER 3: PARTNER CHOICE IN THE LEGUME-RHIZOBIA SYMBIOSIS 221

Overview 221

1. Supplementary material 222
2. ReadMe document for data 228
3. R code used to analyse the data 231

APPENDIX C. SUPPLEMENTARY MATERIAL CHAPTER 4: MATHEMATICAL MODELLING OF THE LEGUME-RHIZOBIA SYMBIOSIS 247

Overview 247

R code 247

APPENDIX D. SUPPLEMENTARY MATERIAL CHAPTER 5: RESOURCE ALLOCATION IN THE LEGUME-RHIZOBIA SYMBIOSIS 255

Overview 255

1. Results based on randomly selected nodules 256
2. ReadMe document for data 259
3. R code used to analyse the data 264

1 Introduction

'[…] there is never a single story […]'

Chimamanda Ngozi Adichie, 2009

1.1 Overview

In this chapter the context for the research conducted in this thesis is provided, and the aims and objectives are defined. The legume-rhizobia symbioses are of tremendous importance in agriculture, but have also been studied by evolutionary biologists interested in the stability of mutualisms. These two perspectives have in common that questions are asked about how the natural variability in rhizobial strains, especially in how much nitrogen they provide, affects plant growth. The approach in this thesis is to combine methods and ways of thinking from those two perspectives, and to ask and address questions relevant for both.

The aim is to understand how plants respond to variability in rhizobial strains, from the perspective of optimising plant growth. The symbiosis and plant growth are inextricably linked, since carbon and nitrogen are the main constituents of plant growth, and in the symbiosis carbon from the plant is exchanged for nitrogen from rhizobia. A large part of understanding the plant response to rhizobia is therefore about understanding resource allocation. Hypotheses on possible ways of resource allocation tend to be simple to formulate, but their implications for plant growth much harder to predict. Mathematical modelling is therefore used to convert simple ideas about resource allocation into measurable outcomes for plant growth. Thus, modelling is used to formulate testable hypotheses which are addressed in growth room experiments.

Background on the symbiosis and organisms involved is provided (1.2) before the literature from the two perspectives on the symbiosis is reviewed and it is explained how they are combined in this thesis (1.3). Carbon and nitrogen metabolism in plants is briefly discussed given their importance for plant growth and the symbiosis (1.4). Then, it is explained why mathematical modelling can be helpful in addressing the questions that arise (1.5), and the choice of model system for the experiments is introduced (1.6). The chapter concludes with setting the aims and objectives, and an overview of the thesis (1.7).

1.2 The legume-rhizobia symbiosis

1.2.1 Organisms involved

Bacteria living near, on and in plant roots can tremendously affect plant growth. This is illustrated at one extreme by pathogenic bacteria[1], and at the other extreme by the positive effect of plant growth promoting (rhizo)bacteria (PGPR), which enhance plant growth via diverse mechanisms[2]. The term *symbiosis*, 'the living together of unlike organisms', covers both extremes while *mutualism* refers only to the situation in which both organisms benefit from the interaction[3–5]. Both terms are applied to situations in which atmospheric nitrogen-fixing bacteria nodulate plant roots, likely because it is not always easy to define benefits received by both plant and bacteria[4]. In contrast, the organisms involved in the interaction are clearly defined and limited to particular groups of both plants and bacteria.

Nitrogen-fixing bacteria were first isolated from a root nodule in 1888 and called *Rhizobium leguminosarum*[6]. Since then, many more bacteria have been isolated and continue to be discovered[7]. Classification was initially based mostly on phenotype on the host plant, but is now based on genetics using 16S and core housekeeping genes[6]. Defining species is complicated by lateral transfer of genes involved in the symbiosis[6], and rather than using *species*, the term *strain* is often used, also to denote mutants of naturally occurring strains. The term rhizobia is applied to any species nodulating plant roots (and in some species shoots[8]) in the Gram-negative phylum Proteobacteria, which includes genera from the Alpha subdivision such as *Rhizobium*, *Azorhizobium*, *Bradyrhizobium*, *Mesorhizobium*, and *Ensifer* (formerly *Sinorhizobium*) and from the Beta subdivision such as *Burkholderia*[6,9]. There are bacteria capable of fixing nitrogen which are not considered rhizobia, such as species in the genus *Frankia* which also nodulate plants but are distinct from rhizobia as they are Gram-positive and form a

different type of nodule on actinorhizal plants such as *Alnus* and *Casuarina*[10,11]. In addition, there are numerous nitrogen-fixing bacterial genera which are free living, or live associatively or endosymbiotically with plants, as in the case of sugar cane and the fern *Azolla*.[12]

For plants, the ability to engage in a symbiosis with nodulating nitrogen-fixing bacteria first arose around 70 million years ago[13] and is limited to the Rosid I clade[14]. Within this clade, the symbiosis with rhizobia is almost exclusively limited to the legume family (*Leguminosae* or *Fabaceae*)[14,15]. Species in the genus *Parasponia* (*Cannabaceae*, but formerly classified as *Ulmaceae*[16]) are the only non-leguminous plants forming symbioses with rhizobia[17]. Eight other families in the Rosid I clade are actinorhizal and form symbioses with *Frankia*[14]. The majority of the 19,500 species in the legume family form nitrogen-fixing symbioses, but not all species do[15]. It is generally agreed that the ability to form the symbiosis has been gained and lost several times[7,9,18], which raises questions about the evolutionary stability of the legume-rhizobia symbiosis (see 1.3.1). Legume phylogenies continue to be updated, but a clade of special importance is the Inverted Repeat Lacking Clade (IRLC). This clade arose around 40 million years ago and contains many temperate and agronomically important legumes, such as pea (*Pisum sativum*), vetch (*Vicia* spp.), lentil (*Lens culinaris*) and clover (*Trifolium* spp.), but not common bean (*Phaseolus vulgaris*) and soybean (*Glycine max*)[9,19,20].

1.2.2 Establishment of the symbiosis

The symbiosis between a legume plant and rhizobia needs to be established anew every plant generation, as rhizobia are not transmitted via the seed. The process is well studied and reviewed[21–26]. A germinating legume plant in need of nitrogen will release chemical signals such as flavonoids into the soil[26,27]. Rhizobia in the rhizosphere, usually

saprophytic when free-living[28,29], may use the root exudates as a carbon source or as a guide to the plant root[30,31] and respond by synthesizing Nod factors (lipo-chito-oligosaccharides)[21–26]. Nod factors are perceived by the plant via receptor-like-kinases with an extracellular LysM domain[32–34]. Variable functional groups of the Nod factors are key to host specificity: a given legume will not form a symbiosis with any rhizobial strain[9,35,36]. Some legume species are very selective (for example those in the IRLC, such as pea) and only form symbioses with a few strains, while others are more promiscuous (for example common bean (*Phaseolus vulgaris*)[9,37]), and there is even variation within species[38]. Nod factor binding to the receptor starts off two parallel processes: the infection of the plant root by rhizobia, and the formation of the nodule[22] (**Figure 1-1**).

The infection most commonly occurs via curling of the root hair around bacteria and the formation of an infection thread, providing rhizobia access to cortical cells in the nodule primordium[9,22]. In some species, rhizobia enter via the epidermis without root hair curling and infection thread formation[9]. The formation of the nodule meristem from which the nodule is formed is guided by plant hormones, especially auxin and cytokinin[22]. The nodule meristem is persistent in some plant species, leading to indeterminate nodules with a cylindrical shape, while other species form spherical determinate nodules which cannot continue to grow[9,39,40]. Inside nodules, bacterial cells are contained in a symbiosome membrane within the infected plant cells[41]. Although mixed infection nodules do occur, bacterial cells within a nodule tend to be clonal descendants of a single bacterium[30,42]. In some species, the bacterial cells responsible for nitrogen fixation are significantly enlarged and become unable to reproduce. Whether or not such bacteroids are formed depends on the host plant, as revealed by experiments with the same strains on different hosts[43,44]. In most agriculturally relevant crops, indeterminate nodules contain non-reproductive nitrogen-fixing bacteroids along with

viable non-fixing bacteria, while determinate nodules contain nitrogen-fixing bacteria which are still able to reproduce, but this is not the case for all species[45]. Eventually, nodules senesce and rhizobia are released back into the soil[46–48].

1.2.3 Nutrient exchange in the symbiosis

The core of the symbiosis is the exchange of nitrogen fixed by rhizobia for carbon fixed by the plant. The details of how this exchange is achieved are, however, far from fully understood, and there is considerable variation among species. Exchange of resources takes place in the infected plant cell and involves two membranes: the symbiosome membrane, and the bacterial membrane within[49] (**Figure 1-2**). Depending on plant species, a single symbiosome contains one or multiple bacteria or bacteroids[23]. Infected plant cells are interspersed with uninfected plant cells in the fixation zone, which is surrounded by vascular bundles connecting the nodule to the rest of the plant via phloem and xylem (**Figure 1-1**)[49].

Carbon enters the infected cell as sucrose from the phloem[49], but is provided to the rhizobia as dicarboxylic acids[50,51], especially malate[49]. Ammonia (NH_3) is thought to diffuse across the bacterial membrane, and to be transported across the symbiosome membrane via either an ammonia (NH_3)[52,53] channel, or a non-specific cation channel transporting ammonium (NH_4^+)[54,55]. Outside the symbiosome, ammonium is incorporated into amino acids or ureides by the infected plant cell and exported to the xylem[49,51]. Plants are thus divided into amide or ureide exporters[56].

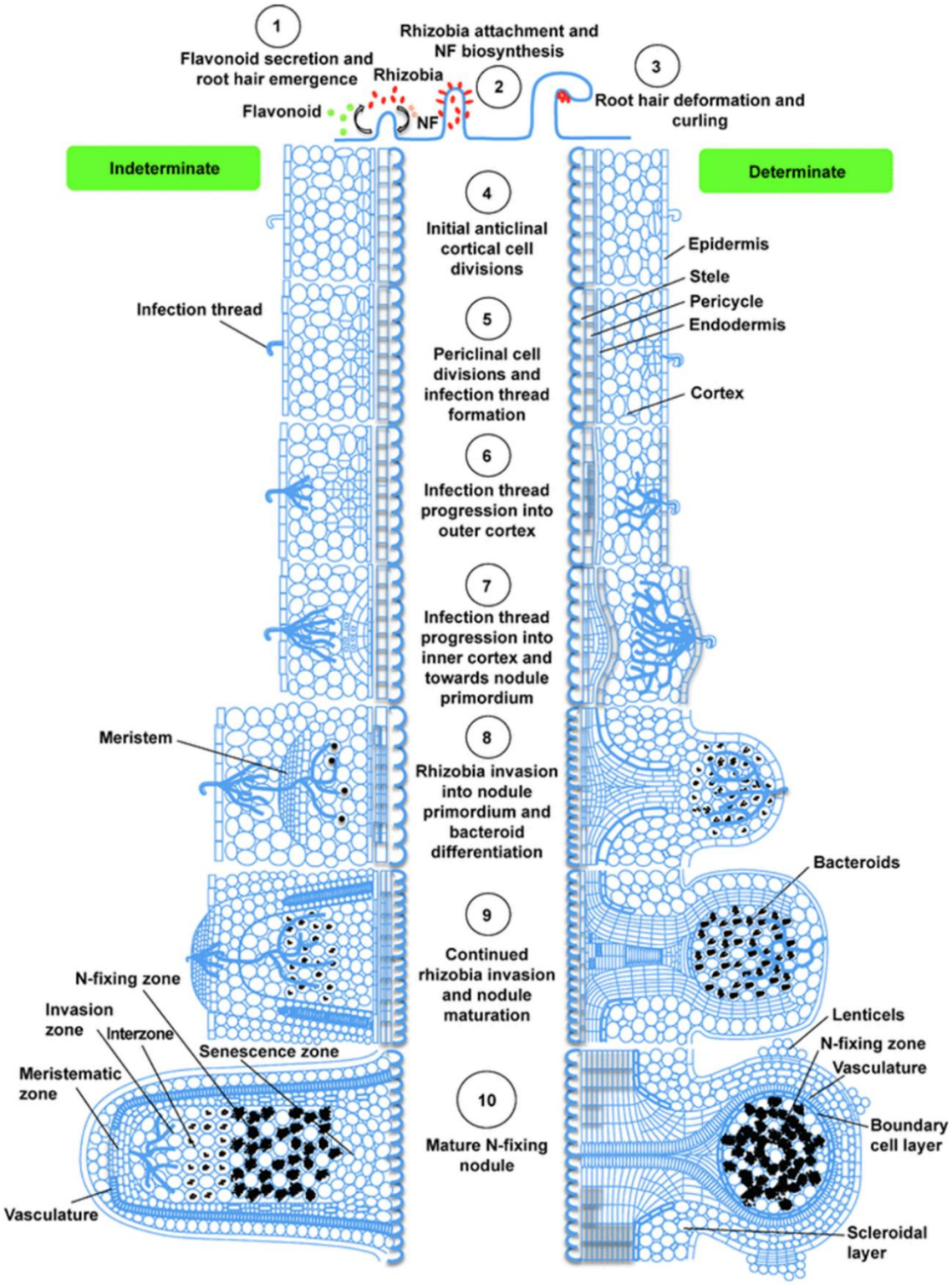


Figure 1-1. Schematic overview of the nodulation process for indeterminate (left) and determinate (right) nodules. Figure reproduced from Ferguson et al., 2010[57].

The bacterial metabolism in the symbiosis is very different from the metabolism in the free-living state and seems to a large extent under plant control[49,58]. Because fixed nitrogen is exported as ammonia and not further incorporated into amino acids, the bacterial cells become dependent on the plant not only for carbon, but also for amino acids[58–61]. Again, the details seem to be dependent on species[23], but transporters for amino acids[62] and various other nutrients such as iron[63], magnesium[64], sulfate[65] and zinc[66] have been identified.

1.2.4 Outcome of the symbiosis

The outcome of the symbiosis has been studied mostly from a plant perspective[4], and more specifically in terms of how much nitrogen the plant has received or how much the growth has been enhanced. Different measures have been adopted. First, plants grown with rhizobia were compared to each other or to non-fixing reference plants in terms of dry matter produced or nitrogen content[67,68]. Later, estimates of nitrogenous compounds in the xylem[69], and direct measurements of nitrogen fixation rates such as acetylene reductions[70] and H_2 evolution[71] also became available. Stable isotope analysis has been used to quantify the proportion of nitrogen provided by rhizobia even under field conditions[69]. Every method has its limitations[68], but it has become clear that there is considerable variation in the outcome of the symbiosis. Some strains provide much more nitrogen than others[72–76], and for a given strain the nitrogen provided may also depend on host plant species[36,77]. The amount of nitrogen provided is denoted the *effectiveness* of the rhizobial strain and variation among strains has been studied from two perspectives: by evolutionary biologists interested in the stability of mutualisms and by agronomists interested in enhancing yields.

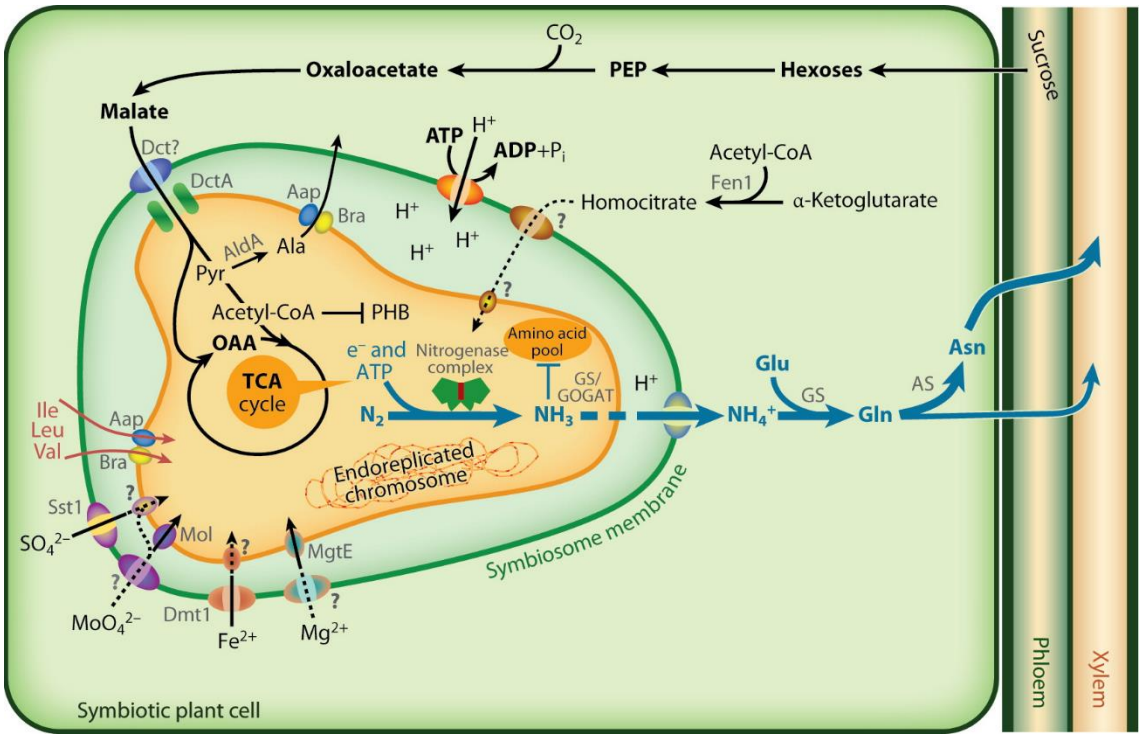


Figure 1-2. Schematic overview of an infected plant cell containing a symbiosome (dark green) which in turn contains a bacteroid (orange). The plant provides the bacteroid with carbon (as malate), amino acids and various other nutrients. In return, the bacteroid provides the plant with nitrogen in the form of ammonia, which is incorporated into amino acids in the infected plant cell, in this case asparagine (Asn), before being exported to the xylem. All compounds exchanged have to pass both the symbiosome and the bacteroid membrane and not all transporters have been identified. Figure reproduced from Udvardi and Poole, 2013[49].

1.3 Different perspectives on the legume-rhizobia symbiosis

1.3.1 The legume-rhizobia symbiosis from an agricultural perspective

Due to the symbiosis with rhizobia, legumes are rich in nitrogen and therefore protein[78–80]. Their nutritional value has been valued since the beginnings of agriculture. The first plants that were domesticated to become crops around 10,000 years ago in the Fertile Crescent not only included grains like wheat and barley, but also legumes like chickpea, lentil, pea and bitter vetch[81]. Later (6,000-9,000 years ago), soybean (*Glycine max*) was domesticated in China[82,83]. In the New World, species of the genus *Phaseolus* were domesticated and especially common beans (*P. vulgaris*) have been cultivated across the Americas[84]. Wherever agriculture developed, legumes were domesticated alongside cereals or tubers[85]. Today, legumes continue to be a crucial source of protein in human diets. Pulses – legumes of which the grains are eaten – contribute 33% to dietary protein nitrogen[79], reason enough for the United Nations to have designated the year 2016 as the International Year of Pulses. But legumes are also used as green manure, and important pasture and fodder crops[79,86,87]. Around 80% of soybean grown globally is processed for animal feed, accounting for 4% of global livestock feed intake[88]. In total, biological nitrogen fixation in agricultural systems amounts to 50-70 Tg nitrogen per year[12]. For comparison, synthetic nitrogen fertiliser inputs into agricultural systems are about 120 Tg nitrogen per year[87]. The energy needed to produce those fertilisers via the Haber-Bosch process accounts for 1.4% of the global energy use[89].

Agricultural research on the legume-rhizobia symbioses is primarily concerned with enhancing crop yields. The variation in effectiveness of rhizobial strains in providing nitrogen and enhancing yields was acknowledged early on, and ineffective strains are common[72,90,91]. A significant effort has therefore been invested in finding and isolating highly effective rhizobial strains to use them as inoculants, first done

commercially in the UK and USA in 1895[67]. The use of inoculants, most commonly in a peat base applied to seed, is standard practice in many areas of the world[67,90]. However, inoculants are not always effective, or only temporarily effective, at enhancing yields. Elite inoculant strains often fail to form sufficient nodules on the plant root, because resident bacterial populations are more competitive in colonizing the plant root, and in surviving in the soil environment over longer periods of time[67,72,90,92–95]. This has been called 'the competition problem'[95]. Research into selecting effective and competitive strains to use as inoculants is ongoing[96–100].

1.3.2 The legume-rhizobia symbiosis from an evolutionary biology perspective

The legume-rhizobia symbiosis has been widely used as a model system to study the evolutionary stability of mutualisms[17]. Mutualisms puzzle evolutionary biologists, as it often seems that one or both partners could gain by providing fewer resources to the symbiotic partner, and instead direct these towards their own fitness. Ultimately, this may lead to the symbiosis breaking down[3,17,42,101–103]. In the legume-rhizobia symbiosis, rhizobia could (partially) forgo energy-costly nitrogen fixation, and instead use resources received from the plant for reproduction or to make storage compounds[42]. Theory predicts that such strains, sometimes called *cheats* in this literature, are likely to arise in the legume-rhizobia symbiosis for two reasons[42,104,105]. First, a single legume plant usually hosts multiple rhizobial strains[106–109]. Therefore, for each of those strains it would be beneficial to obtain plant resources while leaving the nitrogen fixation to the other strains. Second, rhizobia are horizontally transmitted, which means that plant and rhizobial fitness are not closely interdependent[42,104,105].

Two mechanisms have been proposed that would prevent ineffective strains from spreading through the rhizobial population, causing the mutualism to break

down[42,104,105,110,111]. By exerting partner choice, plants might prevent the establishment of ineffective strains. This would mean that plants discriminate against ineffective strains and preferentially form nodules with more effective strains[104,105,110]. Evidence for partner choice has been claimed[105,110,111], but interpretation of results is problematic because natural strains were used, which differ not only in how much nitrogen they provide, but also in how competitive they are at colonising the plant root. Alternatively, plants might sanction ineffective rhizobial strains by allocating fewer resources to those strains once the symbiosis is established and nodules have been formed[42,104,105]. Modelling studies have shown that sanctioning can be selected for in plants[112], and that this could maintain high levels of nitrogen fixation in rhizobial populations[113]. Experimental evidence for sanctioning has been provided by using argon gas to force rhizobia to fix no or less nitrogen, which resulted in smaller nodules containing fewer viable bacteria[114–116]. Also in naturally occurring strains, ineffective nodules have been reported to be smaller than effective nodules[117]. But it remains unclear how sanctions work and how resources would be allocated among multiple rhizobial strains.

1.3.3 Combining the two perspectives

Although the two perspectives on the legume-rhizobia symbiosis seem very different at first glance, they have in common that both are interested in how the variation in effectiveness of rhizobial strains affects plants. The approach in this thesis is to combine the methods and ways of thinking from both the agricultural and the evolutionary biology perspectives to ask and address questions relevant for both fields. For example, the question of whether plants exert partner choice based on how much nitrogen strains provide is one aspect of rhizobial competitiveness with which agricultural research is concerned. It can be addressed using techniques which were developed to study

competition between rhizobia[118] (**chapter 3**). The idea of sanctioning raises questions about plant resource allocation to different rhizobial strains. Differential resource allocation to nodules has received little attention so far from an agricultural perspective, but may be relevant to understand under what conditions rhizobial strains effectively enhance yields.

Understanding sanctioning has the potential to inform agricultural research. From an evolutionary biology perspective, sanctioning can only be a mechanism to stabilise the symbiosis when plants reduce rhizobial fitness of less beneficial strains. At the same time, sanctioning can only be selected for in plants when it directly increases plant fitness. Plant fitness is increased if, by sanctioning, the plant avoids loss of resources to strains that provide little benefit. Since fitness and yields are often both measured in seed output, understanding how and when plants sanction rhizobial strains to increase their fitness may help in understanding how to increase yields. For example, soil nitrogen availability may affect sanctioning, and understanding how may inform whether elite rhizobial strains for inocula are best selected under low or high soil nitrogen conditions. In order to understand sanctioning, sanctioning is mainly considered from a plant perspective in this thesis.

From a plant perspective, sanctioning is a question of optimising resource allocation among rhizobial strains so that no resources are lost on strains which provide little benefit. Currently, it is unclear what optimal resource allocation would look like. The theory of sanctioning is based on the idea that rhizobial strains may cheat on plants by diverting carbon obtained from the plant away from nitrogen fixation and into rhizobial fitness. The main implication of such a diversion is that carbon costs per gram nitrogen fixed are increased, and thus that efficiency of nitrogen fixation is decreased (in addition to any other natural variation in efficiency that may already exist). In the

literature about sanctioning, it is generally assumed that as a result less nitrogen is fixed, and thus that cheating strains are less effective. But alternatively, plants might simply provide more carbon to less efficient strains to (partially) keep up the rate of nitrogen fixation, as it has been argued that plants are not generally carbon-limited[119]. Natural strains are known to vary both in effectiveness (how much nitrogen they provide)[72–76], and in efficiency (carbon costs per gram nitrogen fixed)[120–122], but it is unclear on what basis plants would sanction strains. This is not helped by the fact that the terms effectiveness and efficiency are sometimes used interchangeably and that knowledge is sometimes confounded, for example when effectiveness is expressed as yield differences, which may be affected by both.

By taking a plant perspective and approaching sanctioning as a question of resource allocation, other aspects of sanctioning receive less attention in this thesis. For example, the effects of differential resource allocation on rhizobial fitness are more briefly considered. It is acknowledged that the effects on rhizobial fitness need to be understood to know how the mutualism has persisted over millions of years, and also to understand how to maintain rhizobial populations effective at enhancing yields in agricultural soils. But plant effects on rhizobial fitness cannot be understood without first understanding how plants allocate resources to rhizobia. The focus of the thesis is therefore on understanding how plants allocate resources to rhizobial strains – which may differ both in how much nitrogen they provide and how efficiently they do so. Whether or not ineffective or inefficient strains fit the definition of a 'cheat'[123], or are for other reasons ineffective or inefficient, is also given less attention.

The question of resource allocation to rhizobia by the plant comes down to balancing obtaining nitrogen and spending carbon, because carbon and nitrogen are the main constituents of plant growth and plants provide rhizobia with carbon and receive

nitrogen in return. Although resource allocation to nodules has received little attention, balancing carbon and nitrogen uptake and use through allocation of resources to shoot and root has been a major field of research for many years given the agricultural relevance of understanding what limits plant growth. So while sanctioning may help in understanding how to enhance legume yields, the agriculturally driven research on carbon and nitrogen metabolism may help in understanding sanctioning.

1.4 Carbon and nitrogen metabolism

1.4.1 Carbon and nitrogen: the backbone of life

Life requires many nutrients, but carbon and nitrogen form the backbone of life and are needed in the greatest quantities[124–126]. All tissue is made with both carbon and nitrogen atoms, meaning that carbon and nitrogen metabolism are intimately linked[126–134]. Plants have access to a potentially infinite amount of carbon: as long as they have access to light and water, they can fix atmospheric carbon via photosynthesis. Nitrogen, in contrast, is usually limiting for plant growth[124,125,135–137]. Plants have therefore evolved intricate ways of monitoring and coordinating carbon and nitrogen uptake[127–131,133,134]. A lot of work has been done on the model plant *Arabidopsis thaliana*, although work has been extended to agriculturally relevant crops such as wheat, maize and rice. Some work has been done on legumes, especially the model legume *Medicago truncatula*[125,126,138,139]. Progress has also been made by comparing plants with bacteria and fungi[128,140], and even mammalian systems[141], since carbon and nitrogen metabolism is so central to life that some aspects are conserved across kingdoms.

1.4.2 Plant nitrogen uptake and assimilation

Plants usually take up nitrogen from the soil via their roots. Although nitrogen-containing organic compounds such as urea and amino acids can be taken up directly, the predominant sources are nitrate (NO_3^-) and ammonium (NH_4^+) ions[131,140,142–144]. Legumes have nitrogen fixation as an additional source, but also take up nitrogen via their roots and the mechanisms seem to be the same as for non-legumes[143,145]. How much of the nitrogen is provided via roots in legumes depends on conditions, and can range from 0 to 100%.

Whatever the original source of nitrogen, ultimately the plant forms organic compounds from the ammonium ion (NH_4^+). The ammonium may originate from direct uptake, from amino acid degradation, from nitrogen fixation, or from the reduction of nitrate via nitrite (by nitrate and nitrite reductase respectively). Ammonium is incorporated into the amino acid glutamate, via the combined action of glutamine synthetase (GS) and glutamate oxoglutarate amidotransferase (GOGAT). Glutamate is consequently used to form other amino acids[143,146–149]. The assimilation of nitrogen into amino acids, combining carbon and nitrogen atoms into a single molecule, is the most direct link between carbon and nitrogen metabolism[146].

On a larger scale, this link translates to the need for coordinating carbon and nitrogen uptake, and therefore shoot and root architecture[132,143,144,150–153]. For example, the shoot to root ratio tends to be smaller in nutrient deficient conditions as more resources are allocated to the root to take up the required nutrients[143,150]. How the root grows also depends on local nutrient availability. The control of lateral root growth in response to local nutrient concentrations is well established[140,143,144,150–152]. Regulation is complex, involving both local and systemic signalling pathways[132,152,154,155]. Recent advances in this field continue to elucidate roles of various signalling molecules[131], plant hormones[132,140,151], transcription factors[140,151,155–157] and regulatory microRNAs[140,158].

Adding nitrogen-fixing nodules to the equation makes the puzzle even more intricate, but there may also be synergies. Regulatory systems are at least partially integrated[159–162], as exemplified by formation of fewer and smaller nodules in nitrate rich environments[163–165]. And the idea of allocating resources to nodules based on how much nitrogen they provide (what evolutionary biologists have called sanctioning) is essentially the same as allocating resources (by growing roots) where nutrients are available. The question is how plants coordinate carbon and nitrogen metabolism and

distribute resources over shoot, root and different nodules. Mathematical modelling can be useful in addressing this question.

1.5 Mathematical modelling of the legume-rhizobia symbiosis

1.5.1 Why mathematical modelling can be useful

The idea that plants orchestrate growth of different tissues to enhance their fitness is simple, but the implications are not. Roots and nodules provide different amounts of nitrogen, while requiring different amounts of carbon, and the shoot provides carbon and requires nitrogen. It becomes impossible to mentally and verbally consider different ways of plant resource allocation, and the effects on plant growth of different rhizobial strains and soil nitrogen concentrations. Mathematical models can be useful in situations like this, where the outcomes are often counterintuitive. Simple ideas can be written as mathematical equations relatively straightforwardly, after which the complex outcomes can be studied. Sometimes it is possible to find analytical solutions, but today computers make it feasible to study model outcomes even if the model is too complex to study analytically[166–168]. The computer-generated numerical solutions are also useful in checking analytical solutions. Mathematical modelling is a tool[167,169], and as with any tool, how useful it is depends on how it is used. It is therefore important to be clear about the aims of the modelling, and how it is to be used[170].

1.5.2 How the mathematical modelling is used

Colloquially, a model is often interpreted as ‘how things should be’, but the scientific definition is usually ‘a simplified version of reality’[167,170,171]. This is a rather broad definition, and encompasses anything from maps, to ball-and-stick models of molecules, verbal arguments, and sets of mathematical equations. Mathematical models are a set of hypotheses, in the form of mathematical equations, which together describe a process and explain an outcome of interest[167,170,172,173]. In some cases, there can be confidence in the hypotheses and assumptions that describe the process and models can be used to

predict future events. Models for space travel that make use of well-established physical laws are an often-used example[167,170,172]. In other cases, there is more uncertainty about the set of hypotheses, but the model is used for predictions anyway, as is the case with weather and climate models[170].

Mathematical models can also be used to compare reality with the hypotheses in the model. Statistical models are used to make sense of observations (for example how the outcome of interest is affected by a treatment) given what is assumed to be known about the data (for example that they are normally distributed). Deterministic (non-stochastic) mathematical models are more often used to probe the hypotheses. The idea is that it is possible to get closer to an understanding of reality by repeatedly going through the process of developing a model, comparing the model outcome with reality, discovering discrepancies, re-developing the model etc. Even though the model will never become a perfect representation of reality, the process can be useful[167–169,171,173]. Developing the model forces one to think through a problem very carefully. The model itself may allow for *in silico* experiments so that actual experiments that need to be done can be fewer and better informed[169,171,174], and modelling can help in generating new hypotheses[168,172].

Modelling is in many ways based on judgement and subjective, and therefore sometimes called an art[167,170,172]. The two major judgements are about what is going to be modelled, and in how much detail. A farm can be modelled by describing corn yields based on temperature and fertiliser regimes. But if market prices are not considered the model won't help to understand why the corn yield dropped when the farmer switched to wheat. So the processes described in the model determine what can be learned. Models can be made to match reality better by adding ever more details, but this is not always desirable. Complex models tend to be applicable only to very specific situations, and

although they may describe an outcome rather well they may not advance understanding, if it is unclear which processes are described by the added complexity[167,170].

To simplify models, complex processes are often condensed into a single parameter value[167,169,171]. For example, the complex process of photosynthesis may be summarised into a single value in a model of forest canopy growth[175], while in another model the rate of photosynthesis may be the outcome of interest and be modelled as a function of atmospheric CO_2 concentrations, which in turn can be the subject of a model concerning climate change with parameters about the car industry. So what is a parameter and what is a model outcome depends on the questions asked in the model, and models can be made more complex by replacing parameters with more equations[167]. Parameters themselves can also be of interest, and models are used to establish probable values of parameters (parameter space)[176], or to determine how much the outcome of a model depends on the parameter value (parameter sensitivity analysis)[167,168]. Determining which processes are to be modelled and in how much detail is the art of modelling. Because there are so many different possibilities, none of which will match reality exactly, it is said (first by G.E.P. Box) that 'all models are wrong, but some models are useful'[170].

In this thesis, a mathematical model of plant growth is used to study hypotheses about plant resource allocation to shoot, root and nodules, depending on soil nitrogen availability and different rhizobial strains. The aim of the modelling is to predict how a plant should allocate resources in order to optimise plant growth. These predictions form hypotheses which are tested in experiments. Actual resource allocation may of course deviate from the predicted optimum for various reasons, but these discrepencies inform further model development and experiments and thus advance our understanding. The model presented in this thesis is intended to capture the essentials of plant growth, but is

not intended to be an accurate or realistic representation of plant growth. The model is not specific to a certain species of legume or rhizobia. Therefore, little emphasis is placed on obtaining precise parameter values of key parameters (for example, the rate of photosynthesis), although values were remained within a range of biologically realistic values. The strength of the modelling lies not in its realism, but in using mathematical logic to structure thoughts and generate testable hypotheses.

1.6 Experimental model system

In order to test hypotheses generated by the model, it is necessary to manipulate how much nitrogen rhizobial strains provide and how much carbon they use. I chose the model rhizobial strain *Rhizobium leguminosarum* bv. viciae (*Rlv*) 3841 for use in experiments. The strain is a spontaneous streptomycin-resistant derivative of *Rhizobium leguminosarum* bv. *viciae* 300[177], of which the genome has been published[178], and its widespread use in labs means that many mutant derivatives are available, or can be generated relatively easily. Rlv3841 readily forms indeterminate nodules with pea (*Pisum sativum*), a legume in the agronomically important IRLC group of legumes[9] which is easy to grow in growth rooms. We used cultivar Avola, which has also been widely used (see for example[58,179–181]). A closely related *Rhizobium leguminosarum* 3841 strain (biovar *phaseoli*) forms determinate nodules with bean (*Phaseolus vulgaris*)[9,178,179], which may be useful in future studies considering possible differences in sanctioning mechanisms between determinate and indeterminate nodule types.

1.7 Aims and objectives

In this thesis, the aim is to understand how plants respond to strains which differ in how much nitrogen they provide, and how efficiently they do so. More specifically, the objectives are to:

i) Establish whether legumes preferentially establish a symbiosis by forming more nodules with strains that provide more nitrogen. Such preferential nodule formation has been called partner choice by evolutionary biologists, and is one aspect of the competitiveness of rhizobial strains with which agricultural researchers are concerned. This objective is addressed in **chapter 3**.

ii) Establish how plant resources should be allocated to different rhizobial strains for optimal plant growth: are resources best divided among multiple strains, or should they only be allocated to the best available strain? Should resources be allocated based on effectiveness, or (also) on efficiency of nitrogen fixation? And how does soil nitrogen availability affect resource allocation to and among rhizobial strains? This objective is addressed through a mathematical model of plant resource allocation in **chapter 4**.

iii) Establish how pea plants actually allocate resources to different rhizobial strains. The model predictions are tested by measuring resource allocation to rhizobial strains differing in how much nitrogen they provide and how much carbon they take, under varying soil nitrogen concentrations. This objective is addressed in **chapter 5** and addresses the question of how sanctioning works, which may also have agriculturally relevant implications.

Each chapter is presented in the form of a research paper, and is preceded by a brief overview of the aims of that chapter. **Chapter 3** has been published in *Scientific Reports*[182]. **Chapter 4** and **chapter 5** are being combined to form a single substantial publication (in preparation). For each of the objectives, findings are discussed in the appropriate chapter, and summarised and synthesised in a general discussion (**chapter 6**). **Chapter 2** describes general methods and provides background on the methods used. Specific methods and experimental designs are presented at the end of each chapter, except for the chapter presenting the mathematical model (**chapter 4**) in which methods are provided first so that results can be better understood. References are provided at the end of each chapter. Supplemental information is gathered at the end of the thesis, but split out per chapter consistent with how it was published if applicable. Formatting of chapters and reference styles have been adapted to create a uniform style throughout the thesis.

1.8 References

1. Oerke, E.-C. Crop losses to pests. *J. Agric. Sci.* **144,** 31 (2006).
2. Timmusk, S. Perspectives and challenges of microbial application for crop improvement. *Front. Plant Sci.* **8,** 49 (2017).
3. Herre, E. A., Knowlton, N., Mueller, U. G. & Rehner, S. A. The evolution of mutualisms: exploring the paths between conflict and cooperation. *Trends Ecol. Evol.* **14,** 49–53 (1999).
4. Garcia, J. R. & Gerardo, N. M. The symbiont side of symbiosis: do microbes really benefit? *Front. Microbiol.* **5,** 1–6 (2014).
5. Martin, F. M., Uroz, S. & Barker, D. G. Ancestral alliances: plant mutualistic symbioses with fungi and bacteria. *Science* 6340 (2017).
6. Peix, A., Ramírez-Bahena, M. H., Velázquez, E. & Bedmar, E. J. Bacterial associations with legumes. *CRC. Crit. Rev. Plant Sci.* **34,** 17–42 (2014).
7. Doyle, J. J. Chasing unicorns: nodulation origins and the paradox of novelty. *Am. J. Bot.* **103,** 1865–1868 (2016).
8. Dreyfus, B., Garcia, J. L. & Gills, M. Characterization of *Azorhizobium caulinodans* gen. nov., sp. nov., a stem-nodulating nitrogen-fixing bacterium isolated from *Sesbania rostrata*. **98,** 89–98 (1988).
9. Sprent, J. I., Ardley, J. & James, E. K. Biogeography of nodulated legumes and their nitrogen-fixing symbionts. *New Phytol.* (2017).
10. Giczey, G. *et al.* SymRK defines a common genetic basis for plant root endosymbioses with arbuscular mycorrhiza fungi, rhizobia, and Frankia bacteria. *Proc. Natl. Acad. Sci.* **105,** (2008).
11. Markmann, K., Giczey, G. & Parniske, M. Functional adaptation of a plant receptor-kinase paved the way for the evolution of intracellular root symbioses with bacteria. *PLoS Biol.* **6,** e68 (2008).
12. Herridge, D. F., Peoples, M. B. & Boddey, R. M. Global inputs of biological nitrogen

fixation in agricultural systems. *Plant Soil* **311,** 1–18 (2008).

13. Doyle, J. J. Phylogenetic perspectives on the origins of nodulation. *Mol. Plant-Microbe Interact.* **24,** 1289–1295 (2011).

14. Soltis, D. E. *et al.* Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 2647–2651 (1995).

15. The Legume Phylogeny Working Group (LPWG). A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny. *Taxon* **66,** 44–77 (2017).

16. Akkermans, A. D. L., Abdulkadir, S. & Trinick, M. J. Nitrogen-fixing root nodules in Ulmaceae. *Nature* **274,** 190–190 (1978).

17. Behm, J. E., Geurts, R. & Kiers, E. T. *Parasponia*: a novel system for studying mutualism stability. *Trends Plant Sci.* **19,** 757–763 (2014).

18. Werner, G. D. A., Cornwell, W. K., Sprent, J. I., Kattge, J. & Kiers, E. T. A single evolutionary innovation drives the deep evolution of symbiotic N_2-fixation in angiosperms. *Nat. Commun.* **5,** 4087 (2014).

19. Lavin, M., Herendeen, P. & Wojciechowski, M. Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the Tertiary. *Syst. Biol.* **54,** 575–594 (2005).

20. Oono, R., Schmitt, I., Sprent, J. & Denison, R. Multiple evolutionary origins of legume traits leading to extreme rhizobial differentiation. *New Phytol.* **187,** 508–520 (2010).

21. Albrecht, C., Geurts, R. & Bisseling, T. Legume nodulation and mycorrhizae formation; two extremes in host specificity meet. *EMBO J.* **18,** 281–288 (1999).

22. Oldroyd, G. E. D. & Downie, J. A. Coordinating nodule morphogenesis with rhizobial infection in legumes. *Annu. Rev. Plant Biol.* **59,** 519–46 (2008).

23. Oldroyd, G. E. D., Murray, J. D., Poole, P. S. & Downie, J. A. The rules of engagement in the legume-rhizobial symbiosis. *Annu. Rev. Genet.* **45,** 119–44 (2011).

24. Oldroyd, G. E. D. Speak, friend, and enter: signalling systems that promote beneficial

symbiotic associations in plants. *Nat. Rev. Microbiol.* **11,** 252–63 (2013).

25. Via, V. D., Zanetti, M. E. & Blanco, F. How legumes recognize rhizobia. *Plant Signal. Behav.* **11,** e1120396 (2015).

26. Venturi, V. & Keel, C. Signaling in the rhizosphere. *Trends Plant Sci.* **21,** 187–198 (2016).

27. Liu, C.-W. & Murray, J. The role of flavonoids in nodulation host-range specificity: an update. *Plants* **5,** 33 (2016).

28. Morgan, J. A. W., Bending, G. D. & White, P. J. Biological costs and benefits to plant-microbe interactions in the rhizosphere. *J. Exp. Bot.* **56,** 1729–39 (2005).

29. Denison, R. F. & Kiers, E. T. Life histories of symbiotic rhizobia and mycorrhizal fungi. *Curr. Biol.* **21,** R775-85 (2011).

30. Downie, J. A. The roles of extracellular proteins, polysaccharides and signals in the interactions of rhizobia with legume roots. *FEMS Microbiol. Rev.* **34,** 150–170 (2010).

31. Scharf, B. E., Hynes, M. F. & Alexandre, G. M. Chemotaxis signaling systems in model beneficial plant-bacteria associations. *Plant Mol. Biol.* **90,** 549–559 (2016).

32. Limpens, E., Franken, C., Smit, P. & Willemse, J. LysM domain receptor kinases regulating rhizobial Nod factor-induced infection. *Science* **302,** 630–633 (2003).

33. Radutoiu, S. *et al.* Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *Nature* **425,** 585–592 (2003).

34. Kawaharada, Y. *et al.* Receptor-mediated exopolysaccharide perception controls bacterial infection. *Nature* **523,** 308–12 (2015).

35. Roche, P. *et al.* Molecular basis of symbiotic host specificity in *Rhizobium meliloti*: *nodH* and *nodPQ* genes encode the sulfation of lipo-oligosaccharide signals. *Cell* **67,** 1131–1143 (1991).

36. Wang, D., Yang, S., Tang, F. & Zhu, H. Symbiosis specificity in the legume - rhizobial mutualism. *Cell. Microbiol.* **14,** 334–342 (2012).

37. Michiels, J. *et al. Phaseolus vulgaris* is a non-selective host for nodulation. *FEMS Microbiol. Ecol.* **26,** 193–205 (1998).

38. Yang, S., Tang, F., Gao, M., Krishnan, H. B. & Zhu, H. *R* gene-controlled host specificity in the legume-rhizobia symbiosis. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 18735–18740 (2010).

39. Crespi, M. & Frugier, F. De novo organ formation from differentiated cells: root nodule organogenesis. *Sci. Signal.* **1,** 1–9 (2011).

40. Nizampatnam, N. R., Schreier, S. J., Damodaran, S., Adhikari, S. & Subramanian, S. microRNA160 dictates stage-specific auxin and cytokinin sensitivities and directs soybean nodule development. *Plant J.* **84,** 140–153 (2015).

41. Parniske, M. Intracellular accommodation of microbes by plants: a common developmental program for symbiosis and disease? *Curr. Opin. Plant Biol.* **3,** 320–328 (2000).

42. Denison, R. F. Legume sanctions and the evolution of symbiotic cooperation by rhizobia. *Am. Nat.* **156,** 567–576 (2000).

43. Mergaert, P. *et al.* Eukaryotic control on bacterial cell cycle and differentiation in the *Rhizobium*-legume symbiosis. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 5230–5235 (2006).

44. Bonaldi, K. *et al.* Nodulation of Aeschynomene afraspera and A. indica by photosynthetic Bradyrhizobium sp. strain ORS285: the nod-dependent versus the nod-independent symbiotic interaction. *Mol. Plant-Microbe Interact.* **24,** 1359–1371 (2011).

45. Oono, R., Denison, R. & Kiers, E. Controlling the reproductive fate of rhizobia: how universal are legume sanctions? *New Phytol.* **183,** 967–979 (2009).

46. Timmers, A. C. *et al.* Saprophytic intracellular rhizobia in alfalfa nodules. *Mol. Plant. Microbe. Interact.* **13,** 1204–1213 (2000).

47. Müller, J., Wiemken, A. & Boller, T. Redifferentiation of bacteria isolated from *Lotus japonicus* root nodules colonized by *Rhizobium* sp. NGR234. *J. Exp. Bot.* **52,** 2181–2186 (2001).

48. Schumpp, O. & Deakin, W. J. How inefficient rhizobia prolong their existence within nodules. *Trends Plant Sci.* **15,** 189–95 (2010).

49. Udvardi, M. & Poole, P. S. Transport and metabolism in legume-rhizobia symbioses.

Annu. Rev. Plant Biol. **64,** 781–805 (2013).

50. Oldroyd, G. E. D., Harrison, M. J. & Udvardi, M. Peace talks and trade deals. Keys to long-term harmony in legume-microbe symbioses. *Plant Physiol.* **137,** 1205–10 (2005).

51. Baral, B., Teixeira da Silva, J. A. & Izaguirre-Mayoral, M. L. Early signaling, synthesis, transport and metabolism of ureides. *J. Plant Physiol.* **193,** 97–109 (2016).

52. Niemietz, C. M. & Tyerman, S. D. Channel-mediated permeation of ammonia gas through the peribacteroid membrane of soybean nodules. *FEBS Lett.* **465,** 110–114 (2000).

53. Masalkar, P., Wallace, I. S., Hwang, J. H. & Roberts, D. M. Interaction of cytosolic glutamine synthetase of soybean root nodules with the C-terminal domain of the symbiosome membrane nodulin 26 aquaglyceroporin. *J. Biol. Chem.* **285,** 23880–23888 (2010).

54. Tyerman, S. D., Whitehead, L. F. & Day, D. A. A channel-like transporter for $NH4^+$ on the symbiotic interface of N_2 fixing plants. *Nature* **378,** 629–632 (1995).

55. Obermeyer, G. & Tyerman, S. D. $NH4^+$ currents across the peribacteroid membrane of soybean. Macroscopic and microscopic properties, inhibition by Mg^{2+}, and temperature dependence indicate a subpicosiemens channel finely regulated by divalent cations. *Plant Physiol.* **139,** 1015–1029 (2005).

56. Sprent, J. I. Evolving ideas of legume evolution and diversity: a taxonomic perspective on the occurrence of nodulation. *New Phytol.* **174,** 11–25 (2007).

57. Ferguson, B. J. *et al.* Molecular analysis of legume nodule development and autoregulation. *J. Integr. Plant Biol.* **52,** 61–76 (2010).

58. Prell, J. *et al.* Legumes regulate *Rhizobium* bacteroid development and persistence by the supply of branched-chain amino acids. *Proc. Natl. Acad. Sci.* **106,** 12477–12482 (2009).

59. Lodwig, E., Hosie, A. & Bourdès, A. Amino-acid cycling drives nitrogen fixation in the legume–*Rhizobium* symbiosis. *Nature* **422,** (2003).

60. Prell, J. & Poole, P. Metabolic changes of rhizobia in legume nodules. *Trends Microbiol.* **14,** 161–8 (2006).

61. White, J., Prell, J., James, E. K. & Poole, P. Nutrient sharing between symbionts. *Plant Physiol.* **144,** 604–614 (2007).

62. Hosie, A., Allaway, D., Galloway, C., Dunsby, H. & Poole, P. *Rhizobium leguminosarum* has a second general amino acid permease with unusually broad substrate specificity and high similarity to branched-chain amino acid transporters (Bra/LIV) of the ABC family. *J. Bacteriol.* **184,** 4071–4080 (2002).

63. Kaiser, B. N. *et al.* The soybean NRAMP homologue, GmDMT1, is a symbiotic divalent metal transporter capable of ferrous iron transport. *Plant J.* **35,** 295–304 (2003).

64. Karunakaran, R. *et al.* Transcriptomic analysis of *Rhizobium leguminosarum* biovar *viciae* in symbiosis with host plants *Pisum sativum* and *Vicia cracca*. *J. Bacteriol.* **191,** 4002–4014 (2009).

65. Krusell, L. *et al.* The sulfate transporter SST1 is crucial for symbiotic nitrogen fixation in *Lotus japonicus* root nodules. *Plant Cell* **17,** 1625–1636 (2005).

66. Moreau, S. *et al.* GmZIP1 encodes a symbiosis-specific zinc transporter in soybean. *J. Biol. Chem.* **277,** 4738–4746 (2002).

67. Brockwell, J. & Bottomley, P. J. Recent advances in inoculant technology and prospects for the future. *Soil Biol. Biochem.* **27,** 683–697 (1995).

68. Hardarson, G. & Danso, S. K. A. Methods for measuring biological nitrogen fixation in grain legumes. *Plant Soil* **152,** 19–23 (1993).

69. Kaschuk, G., Nogueira, M. A., de Luca, M. J. & Hungria, M. Response of determinate and indeterminate soybean cultivars to basal and topdressing N fertilization compared to sole inoculation with *Bradyrhizobium*. *F. Crop. Res.* **195,** 21–27 (2016).

70. Hardy, R. W. F., Burns, R. C. & Holsten, R. D. Applications of the acetylene-ethylene assay for measurement of nitrogen fixation. *Soil Biol. Biochem.* **5,** 47–81 (1973).

71. Layzell, D., Weagle, G. & Canvin, D. A highly sensitive, flow through H_2 gas analyzer for use in nitrogen fixation studies. *Plant Physiol.* **75,** 582–585 (1984).

72. Bergersen, F. J. Some Australian studies relating to the long-term effects of the inoculation of legume seeds. *Plant Soil* **32,** 727–736 (1970).

73. Burdon, J. & Gibson, A. Variation in the effectiveness of symbiotic associations between native rhizobia and temperate Australian *Acacia*: within-species interactions. *J. Appl. Ecol.* **36,** 398–408 (1999).

74. Fening, J. O. & Danso, S. K. A. Variation in symbiotic effectiveness of cowpea bradyrhizobia indigenous to Ghanaian soils. *Appl. Soil Ecol.* **21,** 23–29 (2002).

75. Sprent, J. Mutual sanctions. *Nature* **422,** 672–674 (2003).

76. Thrall, P. H., Laine, A.-L., Broadhurst, L. M., Bagnall, D. J. & Brockwell, J. Symbiotic effectiveness of rhizobial mutualists varies in interactions with native Australian legume genera. *PLoS One* **6,** e23545 (2011).

77. Rangin, C., Brunel, B., Cleyet-Marel, J. C., Perrineau, M. M. & Béna, G. Effects of *Medicago truncatula* genetic diversity, rhizobial competition, and strain effectiveness on the diversity of a natural *Sinorhizobium* species community. *Appl. Environ. Microbiol.* **74,** 5653–5661 (2008).

78. Duranti, M. & Gius, C. Legume seeds: protein content and nutritional value. *F. Crop. Res.* **53,** 31–45 (1997).

79. Graham, P. H. & Vance, C. P. Legumes: importance and constraints to greater use. *Plant Physiol.* **131,** 872–877 (2003).

80. Voisin, A.-S., Cazenave, A.-B., Duc, G. & Salon, C. Pea nodule gradients explain C nutrition and depressed growth phenotype of hypernodulating mutants. *Agron. Sustain. Dev.* **33,** 829–838 (2013).

81. Zohary, D., Hopf, M. & Weiss, E. *Domestication of plants in the old world: the origin and spread of domesticated plants in south-west Asia, Europe and the Mediterranean basin*. (Oxford University Press, 2012).

82. Guo, J. *et al.* A single origin and moderate bottleneck during domestication of soybean (*Glycine max*): implications from microsatellites and nucleotide sequences. *Ann. Bot.* **106,** 505–514 (2010).

83. Kim, M. Y., Van, K., Kang, Y. J., Kim, K. H. & Lee, S. Tracing soybean domestication history: from nucleotide to genome. *Breed. Sci.* **452,** 445–452 (2012).

84. Kaplan, L. & Lynch, T. F. *Phaseolus* (Fabaceae) in archeaology: AMS radiocarbon dates and their significance for pre-Colombian agriculture. *Econ. Bot.* **53,** 261–272 (1999).

85. Gepts, P. Crop domestication as a long-term selection experiment. *Plant Breed.* **24,** 1–44 (2004).

86. Smil, V. Nitrogen in crop production: an account of global flows. *Global Biogeochem. Cycles* **13,** 647–662 (1999).

87. Galloway, J. N. *et al.* Transformation of the nitrogen cycle: recent trends, questions, and potential solutions. *Science* **320,** 889–892 (2008).

88. Mottet, A. *et al.* Livestock: On our plates or eating at our table? A new analysis of the feed/food debate. *Glob. Food Sec.* 1–8 (2016).

89. van der Ham, C. J. M., Koper, M. T. M. & Hetterscheid, D. G. H. Challenges in reduction of dinitrogen by proton and electron transfer. *Chem. Soc. Rev.* **43,** 5183–91 (2014).

90. Brockwell, J., Bottomley, P. J. & Thies, J. E. Manipulation of rhizobia microflora for improving legume productivity and soil fertility: a critical assessment. *Plant Soil* **174,** 143–180 (1995).

91. Sprent, J. I. & James, E. K. Legume-rhizobial symbiosis: an anorexic model? *New Phytol.* **179,** 3–5 (2008).

92. Carter, J. M., Tieman, J. S. & Gibson, A. H. Competitiveness and persistence of strains of rhizobia for faba bean in acid and alkaline soils. *Soil Biol. Biochem.* **27,** 617–623 (1995).

93. Graham, P. H. & Vance, C. P. Nitrogen fixation in perspective: an overview of research and extension needs. *F. Crop. Res.* **65,** 93–106 (2000).

94. Deaker, R., Roughley, R. J. & Kennedy, I. R. Legume seed inoculation technology - a review. *Soil Biol. Biochem.* **36,** 1275–1288 (2004).

95. Yates, R. J., Howieson, J. G., Reeve, W. G. & O'Hara, G. W. A re-appraisal of the biology and terminology describing rhizobial strain success in nodule occupancy of legumes in agriculture. *Plant Soil* **348,** 255–267 (2011).

96. Tena, W., Wolde-meskel, E. & Walley, F. Symbiotic efficiency of native and exotic *Rhizobium* strains nodulating lentil (*Lens culinaris* Medik .) in soils of Southern Ethiopia.

Agronomy **6,** 1–10 (2016).

97. Chibeba, A. M., Kyei-Boahen, S., Guimaraes, M. de F., Nogueira, M. A. & Hungria, M. Isolation, characterization and selection of indigenous *Bradyrhizobium* strains with outstanding symbiotic performance to increase soybean yields in Mozambique. *Agric. Ecosyst. Environ.* **246,** 291–305 (2017).

98. Marinho, R. de C. N. *et al.* Symbiotic and agronomic efficiency of new cowpea rhizobia from Brazilian Semi-Arid. *Bragantia* **76,** 273–281 (2017).

99. Kumar, V., Rawat, A. K. & Rao, D. L. N. Population ecology of soybean-rhizobia in diverse crop rotations in Central India. *Agric. Ecosyst. Environ.* **240,** 261–268 (2017).

100. Thilakarathna, M. S. & Raizada, M. N. A meta-analysis of the effectiveness of diverse rhizobia inoculants on soybean traits under field conditions. *Soil Biol. Biochem.* **105,** 177–196 (2017).

101. Doebeli, M. & Knowlton, N. The evolution of interspecific mutualisms. *Proc. Natl. Acad. Sci.* **95,** 8676–8680 (1998).

102. Foster, K. R. & Wenseleers, T. A general model for the evolution of mutualisms. *J. Evol. Biol.* **19,** 1283–1293 (2006).

103. West, S. A., Griffin, A. S. & Gardner, A. Evolutionary explanations for cooperation. *Curr. Biol.* **17,** 661–672 (2007).

104. Kiers, E. T. & Denison, R. F. Sanctions, cooperation, and the stability of plant-rhizosphere mutualisms. *Annu. Rev. Ecol. Evol. Syst.* **39,** 215–236 (2008).

105. Heath, K. & Tiffin, P. Stabilizing mechanisms in a legume–rhizobium mutualism. *Evolution (N. Y).* **63,** 652–662 (2009).

106. Dowling, D. N. & Broughton, W. J. Competition for nodulation of legumes. *Annu. Rev. Microbiol.* **40,** 131–157 (1986).

107. Souza, V. *et al.* Ethnomicrobiology: do agricultural practices modify the population structure of the nitrogen fixing bacteria *Rhizobium etli* biovar *phaseoli*? *J. Ethnobiol.* **17,** 249–266 (1997).

108. Hagen, M. J. & Hamrick, J. L. A hierarchical analysis of population genetic structure in

Rhizobium leguminosarum bv. *trifolii*. *Mol. Ecol.* **5,** 177–186 (1996).

109. Silva, C., Eguiarte, L. E. & Souza, V. Reticulated and epidemic population genetic structure of *Rhizobium etli* biovar *phaseoli* in a traditionally managed locality in Mexico. *Mol. Ecol.* **8,** 277–287 (1999).

110. Gubry-Rangin, C., Garcia, M. & Béna, G. Partner choice in *Medicago truncatula-Sinorhizobium* symbiosis. *Proc. Biol. Sci.* **277,** 1947–51 (2010).

111. Regus, J., Gano, K. A., Hollowell, A. C. & Sachs, J. L. Efficiency of partner choice and sanctions in *Lotus* is not altered by nitrogen fertilization. *Proc. R. Soc. B.* **281,** 20132587 (2014).

112. Hahn, M. & Studer, D. Competitiveness of a *nif*$^-$ *Bradyrhizobium japonicum* mutant against the wild-type strain. *FEMS Microbiol. Lett.* **33,** 143–148 (1986).

113. West, S. A., Kiers, E. T., Simms, E. L. & Denison, R. F. Sanctions and mutualism stability: why do rhizobia fix nitrogen? *Proc. R. Soc. B.* **269,** 685–94 (2002).

114. Kiers, E. T., Rousseau, R. A., West, S. A. & Denison, R. F. Host sanctions and the legume-rhizobium mutualism. *Nature* **425,** 78–81 (2003).

115. Kiers, T. E., Rousseau, R. A. & Denison, R. F. Measured sanctions: legume hosts detect quantitative variation in rhizobium cooperation and punish accordingly. *Evol. Ecol. Res.* **8,** 1077–1086 (2006).

116. Oono, R., Anderson, C. G. & Denison, R. F. Failure to fix nitrogen by non-reproductive symbiotic rhizobia triggers host sanctions that reduce fitness of their reproductive clonemates. *Proc. R. Soc. B.* **278,** 2698–703 (2011).

117. Singleton, W. & Stockinger, K. R. Compensation against ineffective nodulation in soybean. *Crop Sci.* **23,** 69–72 (1983).

118. Sessitsch, A., Hardarson, G., De Vos, W. M. & Wilson, K. J. Use of marker genes in competition studies of *Rhizobium*. *Plant Soil* **204,** 35–45 (1998).

119. Kaschuk, G. *et al.* Photosynthetic adaptation of soybean due to varying effectiveness of N_2 fixation by two distinct *Bradyrhizobium japonicum* strains. *Environ. Exp. Bot.* **76,** 1–6 (2012).

120. Voisin, A. S., Salon, C., Jeudy, C. & Warembourg, F. R. Symbiotic N_2 fixation activity in relation to C economy of *Pisum sativum* L. as a function of plant phenology. *J. Exp. Bot.* **54,** 2733–44 (2003).

121. Kaschuk, G., Kuyper, T. W., Leffelaar, P. A., Hungria, M. & Giller, K. E. Are the rates of photosynthesis stimulated by the carbon sink strength of rhizobial and arbuscular mycorrhizal symbioses? *Soil Biol. Biochem.* **41,** 1233–1244 (2009).

122. Kaschuk, G., Hungria, M., Leffelaar, P. a, Giller, K. E. & Kuyper, T. W. Differences in photosynthetic behaviour and leaf senescence of soybean (*Glycine max* [L.] Merrill) dependent on N_2 fixation or nitrate supply. *Plant Biol.* **12,** 60–69 (2010).

123. Ghoul, M., Griffin, A. S. & West, S. A. Toward an evolutionary definition of cheating. *Evolution* **68,** 318–331 (2013).

124. Sakakibara, H., Takei, K. & Hirose, N. Interactions between nitrogen and cytokinin in the regulation of metabolism and development. *Trends Plant Sci.* **11,** 440–448 (2006).

125. Sulieman, S. Does GABA increase the efficiency of symbiotic N_2 fixation in legumes? *Plant Signal. Behav.* **6,** 32–36 (2011).

126. Kaur, B., Kaur, G. & Asthir, B. Biochemical aspects of nitrogen use efficiency: an overview. *J. Plant Nutr.* **40,** 506–523 (2017).

127. Scheible, W. R. Nitrate acts as a signal to induce organic acid metabolism and repress starch metabolism in tobacco. *Plant Cell Online* **9,** 783–798 (1997).

128. Hsieh, M. H., Lam, H. M., van de Loo, F. J. & Coruzzi, G. A PII-like protein in *Arabidopsis*: putative role in nitrogen sensing. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 13965–13970 (1998).

129. Coruzzi, G. M. & Zhou, L. Carbon and nitrogen sensing and signaling in plants: emerging 'matrix effects'. *Curr. Opin. Plant Biol.* **4,** 247–253 (2001).

130. Stitt, M. *et al.* Steps towards an integrated view of nitrogen metabolism. *J. Exp. Bot.* **53,** 959–970 (2002).

131. Walch-Liu, P., Filleur, S., Gan, Y. & Forde, B. G. Signaling mechanisms integrating root and shoot responses to changes in the nitrogen supply. *Photosynth. Res.* **83,** 239–250

(2005).

132. Ruffel, S., Gojon, A. & Lejay, L. Signal interactions in the regulation of root nitrate uptake. *J. Exp. Bot.* **65,** 5509–5517 (2014).

133. Lu, Y. *et al.* ABI1 regulates carbon/nitrogen-nutrient signal transduction independent of ABA biosynthesis and canonical ABA signalling pathways in *Arabidopsis*. *J. Exp. Bot.* **66,** 2763–2771 (2015).

134. Otori, K. *et al.* Enhanced photosynthetic capacity increases nitrogen metabolism through the coordinated regulation of carbon and nitrogen assimilation in *Arabidopsis thaliana*. *J. Plant Res.* **0,** 0 (2017).

135. Lebauer, D. S. & Treseder, K. K. Nitrogen limitation of net primary productivity in terrestrial ecosystems is globally distritubed. *Ecology* **89,** 371–379 (2008).

136. Gifford, M. L., Dean, A., Gutierrez, R. A., Coruzzi, G. M. & Birnbaum, K. D. Cell-specific nitrogen responses mediate developmental plasticity. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 803–8 (2008).

137. Coskun, D., Britto, D. T. & Kronzucker, H. J. Nutrient constraints on terrestrial carbon fixation: the role of nitrogen. *J. Plant Physiol.* **203,** 95–109 (2016).

138. Thomsen, H. C., Eriksson, D., Møller, I. S. & Schjoerring, J. K. Cytosolic glutamine synthetase: a target for improvement of crop nitrogen use efficiency? *Trends Plant Sci.* **19,** 656–663 (2014).

139. Curci, P. L. *et al.* Transcriptomic response of durum wheat to nitrogen starvation. *Sci. Rep.* **7,** 1176 (2017).

140. Vidal, E. A., Tamayo, K. P. & A, G. R. Gene networks for N-sensing, signaling and response in *Arabidopsis thaliana*. *Rev Syst Biol Med* **2,** 683–693 (2010).

141. Forde, B. G., Cutler, S. R., Zaman, N. & Krysan, P. J. Glutamate signalling via a MEKK1 kinase-dependent pathway induces changes in *Arabidopsis* root architecture. *Plant J.* **75,** 1–10 (2013).

142. Von Wirén, N., Gazzarrini, S., Gojon, A. & Frommer, W. B. The molecular physiology of ammonium uptake and retrieval. *Curr. Opin. Plant Biol.* **3,** 254–261 (2000).

143. Andrews, M., Raven, J. A. & Lea, P. J. Do plants need nitrate? The mechanisms by which nitrogen form affects plants. *Ann. Appl. Biol.* **163,** 174–199 (2013).

144. Forde, B. G. Nitrogen signalling pathways shaping root system architecture: an update. *Curr. Opin. Plant Biol.* **21,** 30–36 (2014).

145. Ruffel, S. *et al.* Systemic signaling of the plant nitrogen status triggers specific transcriptome responses depending on the nitrogen source in *Medicago truncatula*. *Plant Physiol.* **146,** 2020–35 (2008).

146. Temple, S. J., Vance, C. P. & Gantt, S. J. Glutamate synthase and nitrogen assimilation. *Trends Plant Sci.* **3,** 6 (1998).

147. Miao, G. H., Hirel, B., Marsolier, M. C., Ridge, R. W. & Verma, D. P. S. Ammonia-regulated expression of a soybean gene encoding cytosolic glutamine synthetase in transgenic *Lotus corniculatus*. *Plant Cell* **3,** 11–22 (1991).

148. Schoenbeck, M. A. *et al.* Decreased NADH glutamate synthase activity in nodules and flowers of alfalfa (*Medicago sativa* L.) transformed with an antisense glutamate synthase transgene. *J Exp Bot* **51,** 29–39 (2000).

149. Lea, P. J. & Miflin, B. J. Glutamate synthase and the synthesis of glutamate in plants. *Plant Physiol. Biochem.* **41,** 555–564 (2003).

150. Zhang, H., Jennings, A., Barlow, P. W. & Forde, B. G. Dual pathways for regulation of root branching by nitrate. *Plant Biol.* **96,** 6529–6534 (1999).

151. Krouk, G. *et al.* A framework integrating plant growth with hormones and nutrients. *Trends Plant Sci.* **16,** 178–182 (2011).

152. Alvarez, J. M., Vidal, E. A. & Gutiérrez, R. A. Integration of local and systemic signaling pathways for plant N responses. *Curr. Opin. Plant Biol.* **15,** 185–191 (2012).

153. Bisseling, T. & Scheres, B. Nutrient computation for root architecture. *Science* **346,** 300–301 (2014).

154. Forde, B. G. Local and long-range signaling pathways regulating plant responses to nitrate. *Annu. Rev. Plant Biol.* **53,** 203–24 (2002).

155. Li, Y., Krouk, G., Coruzzi, G. M. & Ruffel, S. Finding a nitrogen niche: a systems

integration of local and systemic nitrogen signalling in plants. *J. Exp. Bot.* **65,** 5601–5610 (2014).

156. Xuan, Y. H. *et al.* Indeterminate domain 10 regulates ammonium-mediated gene expression in rice roots. *New Phytol.* **197,** 791–804 (2013).

157. Chen, X. *et al.* Shoot-to-root mobile transcription factor HY5 coordinates plant carbon and nitrogen acquisition. *Curr. Biol.* **26,** 640–646 (2016).

158. Eyles, R. P. *et al.* microRNA profiling of root tissues and root forming explant cultures in *Medicago truncatula*. *Planta* **238,** 91–105 (2013).

159. Kinkema, M., Scott, P. T. & Gresshoff, P. M. Legume nodulation: successful symbiosis through short- and long-distance signalling. *Funct. Plant Biol.* **33,** 707–721 (2006).

160. Mortier, V., Holsters, M. & Goormachtig, S. Never too many? How legumes control nodule numbers. *Plant. Cell Environ.* **35,** 245–58 (2012).

161. Mohd-Radzman, N. A., Djordjevic, M. A. & Imin, N. Nitrogen modulation of legume root architecture signaling pathways involves phytohormones and small regulatory molecules. *Front. Plant Sci.* **4,** 385 (2013).

162. Imin, N., Mohd-Radzman, N. A., Ogilvie, H. A. & Djordjevic, M. A. The peptide-encoding *CEP1* gene modulates lateral root and nodule numbers in *Medicago truncatula*. *J. Exp. Bot.* **64,** 5395–409 (2013).

163. Streeter, J. & Wong, P. P. Inhibition of legume nodule formation and N_2 fixation by nitrate. *CRC. Crit. Rev. Plant Sci.* **7,** 1–23 (1988).

164. Parsons, R., Stanforth, A., Raven, J. a. & Sprent, J. I. Nodule growth and activity may be regulated by a feedback mechanism involving phloem nitrogen. *Plant, Cell Environ.* **16,** 125–136 (1993).

165. Caballero Mellado, J. & Martinez Romero, E. Soil fertilization limits the genetic diversity of *Rhizobium* in bean nodules. *Svmbiosis* **26,** 111–121 (1999).

166. Brennan, R. D., de Wit, C. T., Williams, W. A. & Quattrin, E. V. The utility of a digital simulation language for ecological modeling. *Oecologia* **4,** 113–132 (1970).

167. Leffelaar, P. A. *On systems analysis and simulation of ecological processes with*

examples in CSMP and FORTRAN. (Springer Science & Business Media, 1993).

168. Van Ittersum, M. K. *et al.* On approaches and applications of the Wageningen crop models. *Eur. J. Agron.* **18,** 201–234 (2003).

169. De Wit, C. T., Penning de Vries, F. & Lake, J. Predictive models in agricultural production. *Philos. Trans. R. Soc. B Biol. Sci.* **310,** 309–315 (1985).

170. Silver, N. *The signal and the noise: The art and science of prediction*. (Penguin Books, 2012).

171. Bouman, B. A. M., van Keulen, H., van Laar, H. H. & Rabbinge, R. The school-of-de-Wit crop growth simulation models: a pedigree and historical overview. *Agric. Syst.* **52,** 171–198 (1996).

172. Passioura, J. B. Sense and nonsense in crop simulation. *J. Aust. Inst. Agric. Sci.* 181–183 (1973).

173. Loomis, R. S., Rabbinge, R. & Ng, E. Explanatory models in crop physiology. *Annu. Rev. Plant Biol* **30,** 339–67 (1979).

174. Van Ittersum, M. K. *et al.* Developments in modelling crop growth, cropping systems and production systems in the Wageningen school. *NJAS - Wageningen J. Life Sci.* **50,** 239–247 (2003).

175. Aber, J. D. & Federer, C. A. A generalized, lumped-parameter model of photosynthesis, evapotranspiration and net primary production in temperate and boreal forest ecosystems. *Oecologia* **92,** 463–474 (1992).

176. Klepper, O. & Rouse, D. I. A procedure to reduce parameter uncertainty for complex models by comparison with real system output illustrated on a potato growth model. *Agric. Syst.* **36,** 375–395 (1991).

177. Johnston, A. W. & Beringer, J. E. Identification of the rhizobium strains in pea root nodules using genetic markers. *J. Gen. Microbiol.* **87,** 343–350 (1975).

178. Young, J. P. W. *et al.* The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome Biol.* **7,** R34 (2006).

179. Lodwig, E. M. *et al.* Role of polyhydroxybutyrate and glycogen as carbon storage

compounds in pea and bean bacteroids. *Mol. Plant. Microbe. Interact.* **18,** 67–74 (2005).

180. Mulley, G. *et al.* Pyruvate is synthesized by two pathways in pea bacteroids with different efficiencies for nitrogen fixation. **192,** 4944–4953 (2010).

181. Garcia-Fraile, P. *et al.* Arabinose and protocatechuate catabolism genes are important for growth of *Rhizobium leguminosarum* biovar *viciae* in the pea rhizosphere. *Plant Soil* **390,** 251–264 (2015).

182. Westhoek, A. *et al.* Policing the legume-Rhizobium symbiosis: a critical test of partner choice. *Sci. Rep.* 1–10 (2017).

2 General materials and methods

'The value and utility of any experiment are determined by the fitness of the material to the purpose for which it is used […].'

Gregor Mendel, Experiments in plant hybridization, 1865

2.1 Overview

This chapter describes the general methods for the experimental work described in this thesis. Experimental designs and specific methods are described in each chapter. General methods of growing plants (2.2) and culturing bacteria (2.3) are described first, and then the various post-harvest procedures for data collection (2.4).

2.2 Plant growth

2.2.1 Seed sterilisation and germination

Pea (*Pisum sativum* L. cv Avola) seeds were surface sterilised by immersing them in 95% ethanol in a sterile flask for 30 seconds, washed with sterile distilled water once, immersed in 2% bleach (sodium hypochlorite, NaClO) for five to six minutes and then rinsed with sterile distilled water five to ten times to remove all traces of bleach, transferring the seeds to a clean sterile flask after around half of the washes. Under sterile conditions, peas were then transferred to 0.8% distilled water agar plates. Around 8 peas were placed on a single plate, plates were stacked in groups of around five, wrapped in aluminium foil and stored at room temperature. After around five days, peas that had germinated and which displayed emergent radicles were used for planting.

2.2.2 Nitrogen free nutrient solution

At planting, plants were supplied with 75 mL of nutrient solution containing all essential nutrients except nitrogen (unless otherwise stated), so that rhizobial nitrogen fixation was the sole nitrogen source for plant growth. The formulation is 2.67 times more concentrated as in Poole *et al*., 1994[1] to allow for decreased water holding capacity of a different substrate, and consisted of: 2.67 mM $CaCl_2 \cdot 2\ H_2O$, 276 µM KCl, 2.13 mM $MgSO_4 \cdot 7\ H_2O$, 26.67 µM Fe EDTA, 93.33 µM H_3BO_3, 24 µM $MnCl_2 \cdot 4\ H_2O$, 2.13 µM $ZnCl_2$, 1.33 µM $Na_2MoO_4 \cdot 2\ H_2O$, 0.8 µM $CuSO_4 \cdot 5\ H_2O$, 1.33 g/L KH_2PO_4 and 1.52 g/L Na_2HPO_4.

2.2.3 Planting and growth conditions

Planting was done in the fume hood under sterile conditions. Sterile nitrogen free nutrient solution (75 mL per plant) was mixed with 1 mL of the appropriate rhizobial

inoculum (or sterile water for water controls), and poured over 500 mL pots (Azlon, Fisher) containing a mixture of silver sand and fine vermiculite (equal volumes) which had been autoclaved. Using sterile tweezers, a pre-germinated pea was then placed in the centre of the pot, the substrate just covering the pea. Pots were labelled, covered with cling-film, and placed in the growth room (21 °C, 16 hour photoperiod). For each experiment, pots were placed in a single block and randomised within that block (**Figure 2-1**). After two to five days, the cling-film was slit with a sterile knife to allow the emerging seedling to grow through the hole in the cling-film. Plants were grown for three to six weeks. From two weeks onwards, plants were watered with distilled sterile water as necessary.

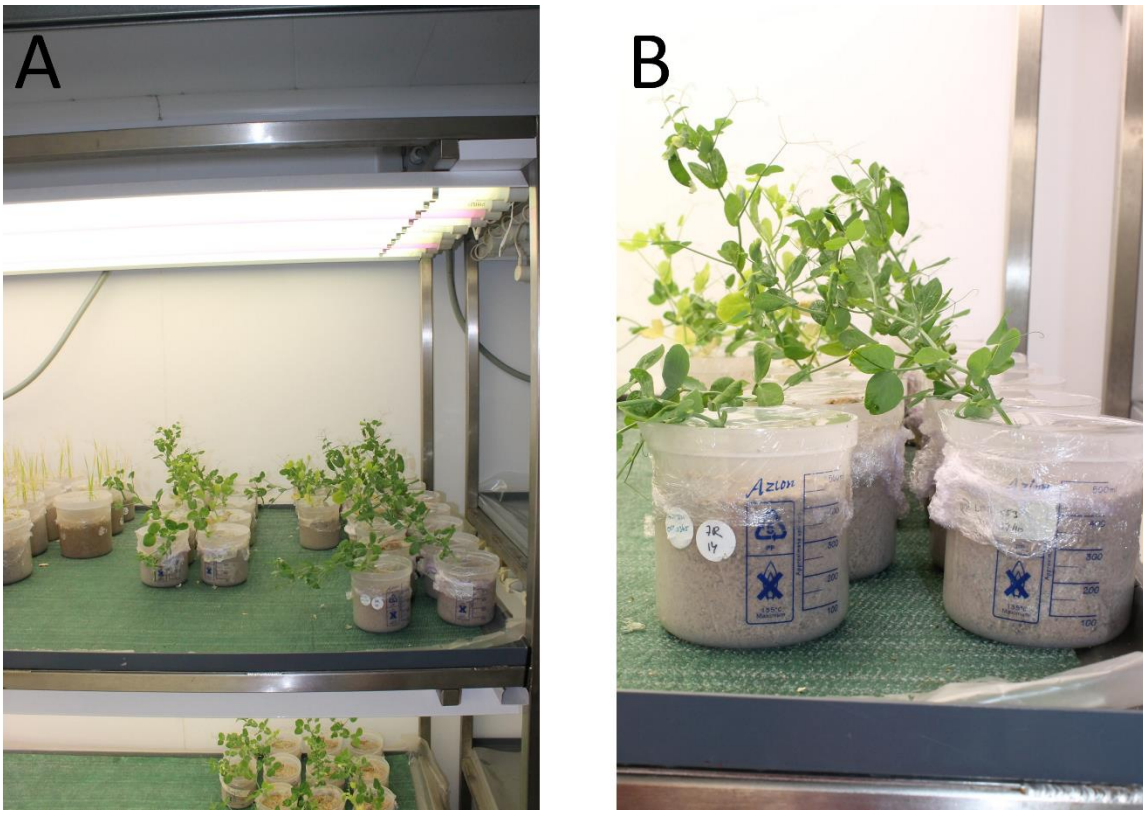


Figure 2-1. Pea plants growing in growth room. (**a**) Three experiments can be seen in three separate blocks. (**b**) Close-up of plants in the growth room. The cling-film covering the pots protects against cross-contamination of rhizobia between pots.

2.2.4 Harvest

At harvest, plants were removed from the growth room and taken out of the pot by removing the cling-film and gently pouring the substrate out of the pot while lightly shaking the plant. Unless acetylene reductions were carried out, roots were gently washed with water to remove sand and vermiculite from the root systems. Depending on the further procedures, shoots remained attached to the roots, or shoots were snipped off just above the pea, leaving the pea attached to the root system unless it fell off.

2.3 Bacterial strains and culture conditions

2.3.1 Bacterial strains

Rhizobium leguminosarum bv. viciae 3841 (Rlv3841) is the parent strain from which all other strains used here are derived. Rlv3841 is a spontaneous streptomycin resistant mutant of UK field isolate 300[2–4]. It has been an often used lab strain in the past four decades, nodulating both pea (*Pisum sativum*) and broad bean (*Vicia faba*), and its genome was published in 2006[4]. Over the years, mutations in the genome or addition of plasmids via conjugations has led to many derivative strains of Rlv3841. The strains used here are presented in **Table 2-1**.

Table 2-1. Rhizobial strains used in this thesis.

Strain	**Antibiotic resistance[a]**	**Description**	**Reference**	**Used in chapter**
Rlv3841	St	St^r derivative of *R.leguminosarum* bv. *viciae* strain 300	[2]	3
Rlv3841 *gusA*	St	Rlv3841 with *gusA* marker	[5]	3, 5
Rlv3841 *celB*	St	Rlv3841 with *celB* marker	[5]	3
RU3940	St, Sp	Rlv3841 *nifH::Ω* Sp^r	This work[6]	3
OPS0365	St, Sp	3841 *gusA nifH::Ω* Sp^r	This work[6]	3, 5
OPS0366	St, Sp	3841 *celB nifH::Ω* Sp^r	This work[6]	3, 5

LMB827	St, Sp	*Ω* Sp^r resistant insertion in *fixC* of Rlv3841, reduced fixation phenotype.	[7]	5
LMB542	St, Tet	Rlv3841[pIJ11268]. Rlv3841 with *lux* genes but no promoter.	[8]	5
LMB553	St, Tet	Rlv3841[pLMB640] (p*nifH* in pIJ11268). Rlv3841 with *lux* genes under *nifH* promoter.	Ramakrishnan Karunakaran (Poole lab), [9]	5

[a]Neo = neomycin (40 μg ml^{-1}), Sp = spectinomycin (100 μg ml^{-1}), St = streptomycin (500 μg ml^{-1}), Tet = tetracyclin (5 μg ml^{-1}).

2.3.2 Culture conditions

Rhizobium strains were grown at 28 °C on Tryptone Yeast (TY) medium (5 g/L tryptone, 3 g/L yeast extract, 6 mM $CaCl_2$)[10], usually solidified with agar (1.75% w/v). Strains were grown on TY agar plates (petri dish) for three to four days to prepare for further use, to count colonies, or to carry out selections. Strains were grown on TY agar slopes (universal containing 10 mL agar solidified at an angle to increase surface area) for two to three days to grow up bacterial cultures for planting. TY agar was supplemented with antibiotics as was appropriate for the strain (see **Table 2-1**). If colonies needed to be distinguished via *gusA* marker genes, the TY agar was supplemented with 400 μg/mL X-glcA or magenta-glcA (Sigma-Aldrich) (see 2.4.3 for further explanation on GusA staining).

2.3.3 Preparation of rhizobial inocula

To prepare rhizobial inocula, strains were grown on TY agar slopes for two or three days at 28°C. Slopes were then washed with 5 mL sterile distilled water, and the liquid now containing the rhizobia transferred to a clean Falcon tube. A sample of 1 mL was used to establish rhizobial density using spectrophotometry at a wavelength of 600 nm (Thermo Scientific Helios Gamma). Slopes were then diluted down with sterile water to OD_{600} = 0.1 (corresponding to approximately 1 x 10^8 cells ml^{-1}), and from there to the appropriate final inoculation density (1 x 10^2 – 1 x 10^8 cells ml^{-1}). For mixed inoculations, rhizobial inocula were mixed in the appropriate ratio.

2.3.4 Storage

Rhizobial strains were kept on TY agar plates containing the appropriate antibiotics in the cold room (7 °C) for short-term storage (up to a few weeks). For long term storage, strains were kept at -80 °C in TY medium with 15-20% glycerol.

2.4 Post-harvest procedures

2.4.1 Shoot dry weight

To measure shoot dry weight, shoots were placed in paper envelopes and incubated in a 70°C oven for at least 24 hours. Dried shoots were then taken out of the envelope and weighed using an Acculab Sartorius Group scale model ALC-80-40.

2.4.2 Measuring nitrogen fixation by acetylene reduction

Acetylene reduction rapidly became an often used assay of nitrogen fixation rates after it was developed at the end of 1960's[11]. The method has also been extensively criticised, the main concern being that reduction rates are affected during the assay[12] and by disturbing the root system[13]. The assay is also only suitable for measuring rates of nitrogen fixation at a specific point in time, not over the lifetime of a plant[14]. Alternative methods include measuring nitrogen accumulation in plants grown with rhizobia as the sole nitrogen source, and stable isotope analysis in which ^{15}N in plant tissue is measured after incubation in a ^{15}N-enriched atmosphere (virtually all atmospheric nitrogen is ^{14}N)[14]. These methods also have their problems, and are considerably more time-consuming. Acetylene reductions have therefore remained a standard way of measuring nitrogen fixation rates.

Acetylene reductions provide a measure of nitrogen fixation rates because the nitrogenase enzyme that normally reduces atmospheric nitrogen (N_2) to ammonia (NH_3) (Reaction 2-1), also reduces acetylene (C_2H_2) to ethylene (C_2H_4) (Reaction 2-2)[15].

$$N{\equiv}N + 8H^+ + 8e^- \xrightarrow{nitrogenase} 2NH_3 + H_2 \qquad \text{Reaction 2-1}$$

$$HC{\equiv}CH + 2H^+ + 2e^- \xrightarrow{nitrogenase} H_2C{=}CH_2 \qquad \text{Reaction 2-2}$$

The nitrogenase activity is greater for acetylene reduction than for nitrogen fixation[16], so a measure of nitrogen fixation rates can be obtained by incubating nodulated roots in an atmosphere with acetylene and measuring the amount of ethylene produced by gas chromatography[17].

Roots were placed on a damp piece of tissue paper in 250 mL Schott bottles (total volume 320 mL) with rubber seal lids, taking care not to damage any part of the root or attached shoot. With a syringe, 8 mL of air was removed from the Schott bottle and replaced by 6.4 mL of acetylene gas, resulting in a concentration of 2% acetylene inside the bottle. After incubation for one hour, three replicate samples of 1 mL were taken from the bottles. The samples were analysed by gas chromatography (Clarus® 480, Perkin-Elmer®).

To calculate the amount of ethylene produced, the fraction of acetylene converted was calculated as % ethylene / (% ethylene + % acetylene), based on the percentages ethylene and acetylene measured in the samples. The initial amount of acetylene added was 261.5982 μmol (based on 6.4 mL of gas added and 24.465 L mol^{-1}). Multiplying the fraction of acetylene converted by the amount of acetylene initially added, yields the amount of ethylene produced.

2.4.3 Distinguishing strains based on GusA and CelB staining

The *gusA* and *celB* marker genes were used to visually distinguish between rhizobial strains after nodule formation on plants. The *gusA* gene, derived from *E. coli*, encodes for β-glucuronidase (GUS)[18]. Providing strains which have the *gusA* gene with glucuronide substrates such as X-glcA (5-bromo-4-chloro-3-indolyl-β-D-glucuronide) or magenta-glcA (5-bromo-6-chloro-3-indolyl-β-D-glucuronic acid), results in precipitates staining nodules containing the strain blue or magenta, respectively[19]. The gene is not

naturally present in plants, nor in *Rhizobium* strains, and therefore very useful as a marker gene[18]. The *celB* gene used as a marker comes from the hyperthermophilic archaeon *Pyrococcus furiosus*. It encodes a thermostable β-glucosidase with a high β-galactosidase activity, which is active up to 100 °C[19]. Plants and *Rhizobium* naturally have high β-galactosidase activity, but this is not thermostable and can easily be inactivated with heat treatment[19]. Thus, heat treating nodules before supplying the X-gal substrate (5-bromo-4-chloro-3-indolyl-β-D-galactoside) renders only nodules containing strains with the thermostable *celB* marker gene blue[19]. Staining for GusA and CelB is fatal for rhizobia, so the method is incompatible with, for example, counting viable rhizobia.

To stain nodules based on *gusA* and *celB* marker genes, a post-harvest protocol (as previously described[19]) was followed. Roots were submerged upside down in 40 mL phosphate buffer (NaH_2PO_4 7 g/L, Na_2HPO_4 7.2 g/L, EDTA 1 mM (pH 8), Sarkosyl 1%, Triton 1 mL/L), in 50 mL Falcon tubes covered with aluminium foil, as the substrates are light sensitive. For GusA staining, X-glcA or magenta-glcA (Sigma-Aldrich) was added to a final concentration of 200 μg mL^{-1} and the roots were incubated overnight at 28 °C. For CelB staining, roots were submerged in preheated phosphate buffer and incubated at 70 °C for at least one hour. Then, X-gal (Sigma-Aldrich) was added to a final concentration of 250 μg mL^{-1} and the roots were incubated overnight at 37 °C. If just a single strain was marked, the *gusA* marker gene was used and nodules were stained either blue (X-glcA) or magenta (magenta-glcA). If two strains were marked, GusA staining (using magenta-glcA) and CelB staining were done sequentially, transferring the roots to fresh (preheated) phosphate buffer after the GusA staining. Stocks of X-glcA, magenta-glcA and X-gal solutions were made at 20 mg mL^{-1} in dimethylformamide (DMF), and usually made fresh for each use.

2.4.4 Photographing root systems

Root systems of harvested plants were usually photographed for later data collection and preservation. Roots were gently washed to remove sand and vermiculite if this had not already been done. They were carefully dried with tissue paper, and then spread out on 120 mm x 120 mm square petri dishes, making sure that all nodules were visible. The root systems were photographed by John Baker (Department of Plant Sciences, University of Oxford). Photographs were taken against a black background (black cloth) and against a white background (light box underneath the plate). The lids of the petri dishes were removed before taking the photograph and roots were illuminated from two sides. Photographs were taken using Nikon D300 or D700, with an exposure time as needed and no flash.

2.4.5 Microscopy

Microscopy was used to identify mixed nodules, and to capture nodule phenotypes of both mixed and singly infected nodules. Photographs were taken using a dissecting microscope (Leica M165 FC) with accompanying software (LAS v4.5).

2.4.6 Counting nodules and measuring nodule area

The photographed root systems were used to count nodules and measure their area. Nodules were counted while marking them in Microsoft Paint®, to make it easier to keep track of which nodules were counted and allowing for zooming in as necessary. To measure nodule area, the ten largest nodules of each strain were selected by eye and marked in Microsoft Paint®. To reduce variability within treatments, the ten largest, rather than randomly selected, nodules were measured to ensure nodules of similar developmental stage were compared (younger nodules are smaller). Then, each

photograph was opened in ImageJ v1.49v[20]. The scale was set using 10 mm of the scale bar, and for each nodule, the contour around it was drawn so that the ImageJ software could measure the area inside the drawn shape. If nodules of different strains were measured, the picture file was not closed between measuring the different strains, so that the same scale bar settings were used for all nodules on a given plant. Photographs of root systems were processed in this way in random order within each experiment.

2.5 References

1. Poole, P. S., Blyth, A., Reid, C. J. & Walters, K. myo-Inositol catabolism and catabolite regulation in *Rhizobium leguminosarum* bv. *viciae*. *Microbiology* **140,** 2787–2795 (1994).
2. Johnston, A. W. & Beringer, J. E. Identification of the rhizobium strains in pea root nodules using genetic markers. *J. Gen. Microbiol.* **87,** 343–350 (1975).
3. Brewin, N. J., Beringer, J. E. & Johnston, a. W. B. Plasmid-mediated transfer of host-range specificity between two strains of *Rhizobium leguminosarum*. *Microbiology* **120,** 413–420 (1980).
4. Young, J. P. W. *et al.* The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome Biol.* **7,** R34 (2006).
5. Sánchez-Cañizares, C. & Palacios, J. Construction of a marker system for the evaluation of competitiveness for legume nodulation in *Rhizobium* strains. *J. Microbiol. Methods* **92,** 246–249 (2013).
6. Westhoek, A. *et al.* Policing the legume-Rhizobium symbiosis: a critical test of partner choice. *Sci. Rep.* 1–10 (2017).
7. Webb, I. U. C. Characterisation of the *fixABCX* operon in symbiotic nitrogen fixation. (University of East Anglia, 2016).
8. Pini, F. *et al.* Lux bacterial biosensors for in vivo spatiotemporal mapping of root secretion. *Plant Physiol.* 1302 (2017). doi:10.1104/pp.16.01302
9. Frederix, M. *et al.* Mutation of *praR* in *Rhizobium leguminosarum* enhances root biofilms, improving nodulation competitiveness by increased expression of attachment proteins. *Mol. Microbiol.* **93,** 464–478 (2014).
10. Beringer, J. E. R factor transfer in *Rhizobium leguminosarum*. *J. Gen. Microbiol.* **84,** 188–198 (1974).
11. Hardy, R. W. F., Burns, R. C. & Holsten, R. D. Applications of the acetylene-ethylene assay for measurement of nitrogen fixation. *Soil Biol. Biochem.* **5,** 47–81 (1973).
12. Minchin, Witty, Sheehy & Müller. A major error in the acetylene reduction assay:

decreases in nodular nitrogenase activity under assay conditions. *J. Exp. Bot.* **34,** 641–649 (1983).

13. Minchin, F. R., Sheehy, J. E. & Witty, J. F. Further errors in the acetylene reduction assay: effects of plant disturbance. *J. Exp. Bot.* **37,** 1581–1591 (1986).
14. Hardarson, G. & Danso, S. K. A. Methods for measuring biological nitrogen fixation in grain legumes. *Plant Soil* **152,** 19–23 (1993).
15. Giller, K. E. in *Nitrogen fixation in tropical cropping systems.* 72–92 (CABI Publishing, 2001).
16. Schöllhorn, R. & Burris, R. H. Acetylene as a competitive inhibitor of N_2 fixation. *Proc. Natl. Acad. Sci. U. S. A.* **58,** 213–216 (1967).
17. Trinick, M. J., Dilworth, M. J. & Grounds, M. Factors affecting the reduction of acetylene by root nodules of *Lupinus* species. *New Phytologist* **77,** 359–370 (1976).
18. Wilson, K. J. *et al. β*-glucuronidase (GUS) transposons for ecological and genetic studies of rhizobia and other gram-negative bacteria. *Microbiology* **141,** 1691–1705 (1995).
19. Sessitsch, A., Wilson, K. J., Akkermans, A. D. L. & De Vos, W. M. Simultaneous detection of different *Rhizobium* strains marked with either the *Escherichia coli gusA* gene or the *Pyrococcus furiosus celB* gene. *Appl. Environ. Microbiol.* **62,** 4191–4194 (1996).
20. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9,** 671–675 (2012).

3 Partner choice in the legume-rhizobia symbiosis

'[…] *in nature nothing exists alone.'*

Rachel Carson, Silent spring, 1962

3.1 Overview

This chapter addresses the question of whether legumes can exert partner choice and preferentially establish symbioses with more effective rhizobial strains. The chapter was published as:

Policing the legume-Rhizobium symbiosis: a critical test of partner choice.

Annet Westhoek, Elsa Field, Finn Rehling, Geraldine Mulley, Isabel Webb, Philip S. Poole & Lindsay A. Turnbull.

Scientific Reports 7, Article number: 1419 (2017). doi:10.1038/s41598-017-01634-2

Here, the work is presented, formatted and edited for consistency with the remainder of the thesis. The published manuscript can be found in Appendix A. Supplementary materials are in Appendix B. The data that support the findings of this study are available in the Oxford Research Archive (ORA-Data) repository, DOI: 10.5287/bodleian:BeXp1PrbB.

3.2 Abstract

In legume-rhizobia symbioses, specialised soil bacteria fix atmospheric nitrogen in return for carbon. However, ineffective strains can arise, making discrimination essential. Discrimination can occur via partner choice, where legumes prevent ineffective strains from entering, or via sanctioning, where plants provide fewer resources. Several studies

have inferred that legumes exercise partner choice, but the rhizobia compared were not otherwise isogenic. To test when and how plants discriminate ineffective strains we developed sets of fixing and non-fixing strains that differed only in the expression of *nifH* – essential for nitrogen fixation – and could be visualised using marker genes. We show that the plant is unable to select against the non-fixing strain at the point of entry, but that non-fixing nodules are sanctioned. We also used the technique to characterise mixed nodules (containing both a fixing and a non-fixing strain), whose frequency could be predicted using a simple diffusion model. We discuss that sanctioning is likely to evolve in preference to partner choice in any symbiosis where partner quality cannot be adequately assessed until goods or services are actively exchanged.

3.3 Introduction

Across the globe primary productivity is nitrogen limited[1]. This limitation has been overcome for plants in the family Fabaceae (commonly known as legumes) through a mutualistic association with nitrogen-fixing bacteria collectively called rhizobia[2]. The nitrogen provided through this symbiosis makes legumes rich in protein and important crops in human diets[3]. But, as ineffective strains will inevitably arise through mutation, there is the potential for the relationship to break down. Ineffective strains are known to be common, at least in some situations, which for agricultural legumes means poor yields and reduced nutritional quality[4,5].

Theory predicts that ineffective strains could be successful within legume – rhizobia symbioses for two reasons[6,7]. First, rhizobia are not transmitted directly from parent plant to offspring. Instead, plants acquire rhizobia from the soil through an intricate signalling process in which bacteria enter specialised root nodules, where they fix nitrogen in return for plant-derived carbon[8]. This horizontal transmission means that rhizobial fitness is not perfectly aligned with the fitness of the host plant[7]. Second, although each nodule is usually occupied by the clonal descendants of a single rhizobial bacterium[9,10], a plant is usually infected by multiple rhizobial strains[11,12]. Thus, a non-fixing strain can potentially thrive by taking plant resources while leaving the costly process of nitrogen fixation to others[7,13]. To prevent losing resources to ineffective rhizobial strains that provide little or no nitrogen, legumes have two options: partner choice or sanctions[7].

Partner choice is usually defined as any mechanism that allows detection of suitable partners *before* a mutualistic relationship is established[7,13–15], while sanctioning is a mechanism to discriminate against low-quality partners once the relationship is underway[6,7,13–16] (although confusingly ‘partner choice’ has also been used to describe a

broader concept which includes sanctioning[13,17,18]). Partner choice might seem to be the more attractive option as resources are not wasted setting up a relationship that is doomed to fail. But crucially, effective partner choice requires accurate assessment of the quality of partners in advance[13]. This is likely to be problematical for any symbiosis in which key traits are not manifested prior to the relationship being established. For example, in the legume-rhizobia symbiosis, nitrogen fixation does not begin until the bacteria have entered the roots and nodule formation is sufficiently advanced for rhizobia to have differentiated into nitrogen-fixing bacteroids[8]. Once nodules are established they can be sanctioned, if they prove to be ineffective, by cutting off their supply of carbon, oxygen or other nutrients and this has been demonstrated empirically using argon gas to force nodules to fix less nitrogen[19–21].

Despite the empirical support for sanctioning and its apparent advantages, there are nevertheless several studies that claim evidence for partner choice[7,14,15]. However, the interpretation of these studies is problematic because the tested strains are rarely isogenic – meaning that strains differ in several traits, and not just in how much nitrogen they provide. Most importantly, strains are likely to differ in their competitiveness in colonizing plant roots and forming nodules. A range of traits affects competitiveness: examples including motility[22], production of antibiotics[23] and the secretion of proteins and polysaccharides involved in biofilm formation and root attachment[24]. Such differences in competitiveness explain why poorly-fixing strains can also end up occupying a higher proportion of nodules – a problem that is often encountered when developing effective strains for use in agricultural settings[4,25,26]. Thus comparing the nodulation success of naturally occurring strains is difficult to interpret as a test of partner choice.

To test whether plants can directly assess the effectiveness of potential rhizobia prior to nodulation we created a non-fixing mutant from a fixing strain and compared their success in colonising pea plant nodules. There are several key genes involved in nitrogen fixation[8], any of which could undergo a mutation that would render the gene non-functional and hence transform the fixing into a non-fixing strain. We chose the *nifH* gene in *Rhizobium leguminosarum* bv. viciae (*Rlv*) 3841, and created a non-fixing mutant that was otherwise identical to its fixing parent strain. We then assessed when and how the plant discriminated between the two strains. To identify strains, they were marked with *gusA* or *celB* marker genes, rendering strains magenta or blue (respectively) following the application of a simple post-harvest staining protocol. Insertion of marker genes solves a secondary problem as it is usually extremely time-consuming to identify different strains using antibiotic markers and this limits the number of nodules that can be assessed.

One possible complication is that a non-fixing strain can potentially thrive via mixed nodules (where two different bacteria have entered and colonised). If a non-fixing strain can take advantage of mixed nodules to increase its fitness at the expense of the fixing strain, then this would provide a route by which non-fixing strains could increase in frequency. Currently little is known about the frequency at which mixed nodules occur, and the relative fitness of strains within mixed nodules. The staining protocol rendered mixed nodules easily visible, so we assessed the frequency of mixed nodules under different inoculation densities. We discuss mixed nodules, partner choice and sanctioning in the context of the evolutionary stability of the legume-rhizobia symbiosis.

3.4 Results

In single inoculations, the non-fixing mutant strain reduced shoot mass and formed small white nodules, while the fixing parent strain formed larger pink nodules (indicating the presence of leghaemoglobin) (**Figure 3-1**).

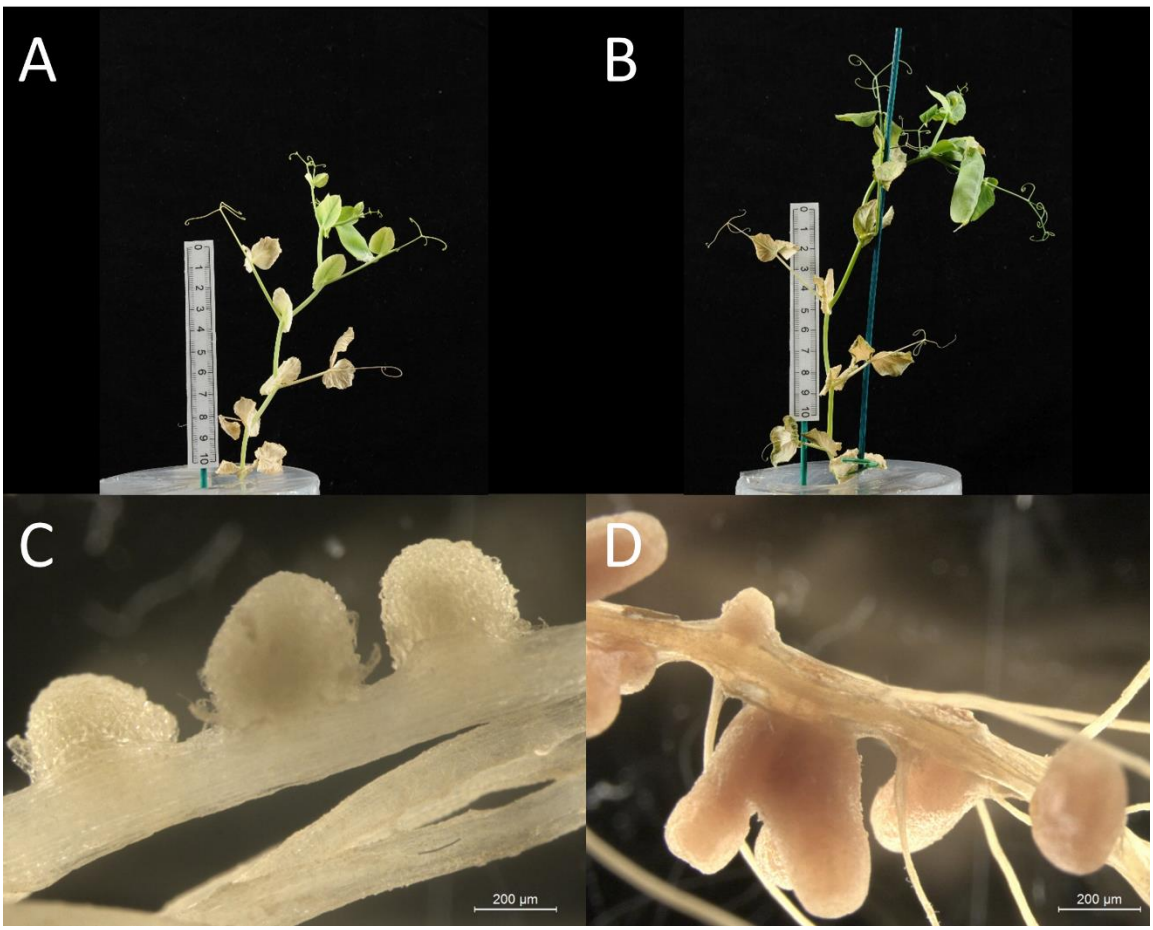


Figure 3-1. Comparison of the non-fixing strain RU3940, (**a, c**) and the otherwise isogenic fixing strain Rlv3841 (**b, d**). After five weeks of growth, plants inoculated with the non-fixing strain (**a**) had reduced biomass and produced fewer peas than those inoculated with the fixing strain (**b**). The non-fixing strain formed white (**c**) rather than pink (**d**), nodules, indicating a lack of leghaemoglobin. Scale bars indicate 200 µm.

3.4.1 Partner choice

To test for partner choice, we varied the ratio of the fixing: non-fixing strains in the inoculum (1:10, 1:1 and 10:1) and counted the number of fixing and non-fixing nodules after three weeks. In the absence of partner choice, we expect the percentage of nodules containing the fixing strain to reflect the percentage in the inoculum. In other words, we expect a 1:1 relationship between the percentage of fixing nodules and the percentage of the fixing strain in the inoculum (a slope of 1.0 through the origin). We found that the percentage of nodules containing only the fixing strain exactly reflected the percentage of the fixing strain in the inoculum (**Figure 3-2**). The slope of the regression line was 0.99 ± 0.02 (95% CI [0.95, 1.02]) and this is not significantly different from 1.0 (t_{39} = -0.835, p = 0.41). The intercept was not significantly different from zero (t_{39} = -1.338, p = 0.189). Whether the fixing strain was marked with *gusA* or *celB* had no impact on the proportion of fixing nodules formed (t_{37} = -0.177, p = 0.861) or on the interaction with the inoculum ratio (t_{37} = -0.189, p = 0.851). Thus, pea plants do not discriminate between fixing and non-fixing strains prior to nodule formation and are therefore unable to exercise partner choice against the mutant strain.

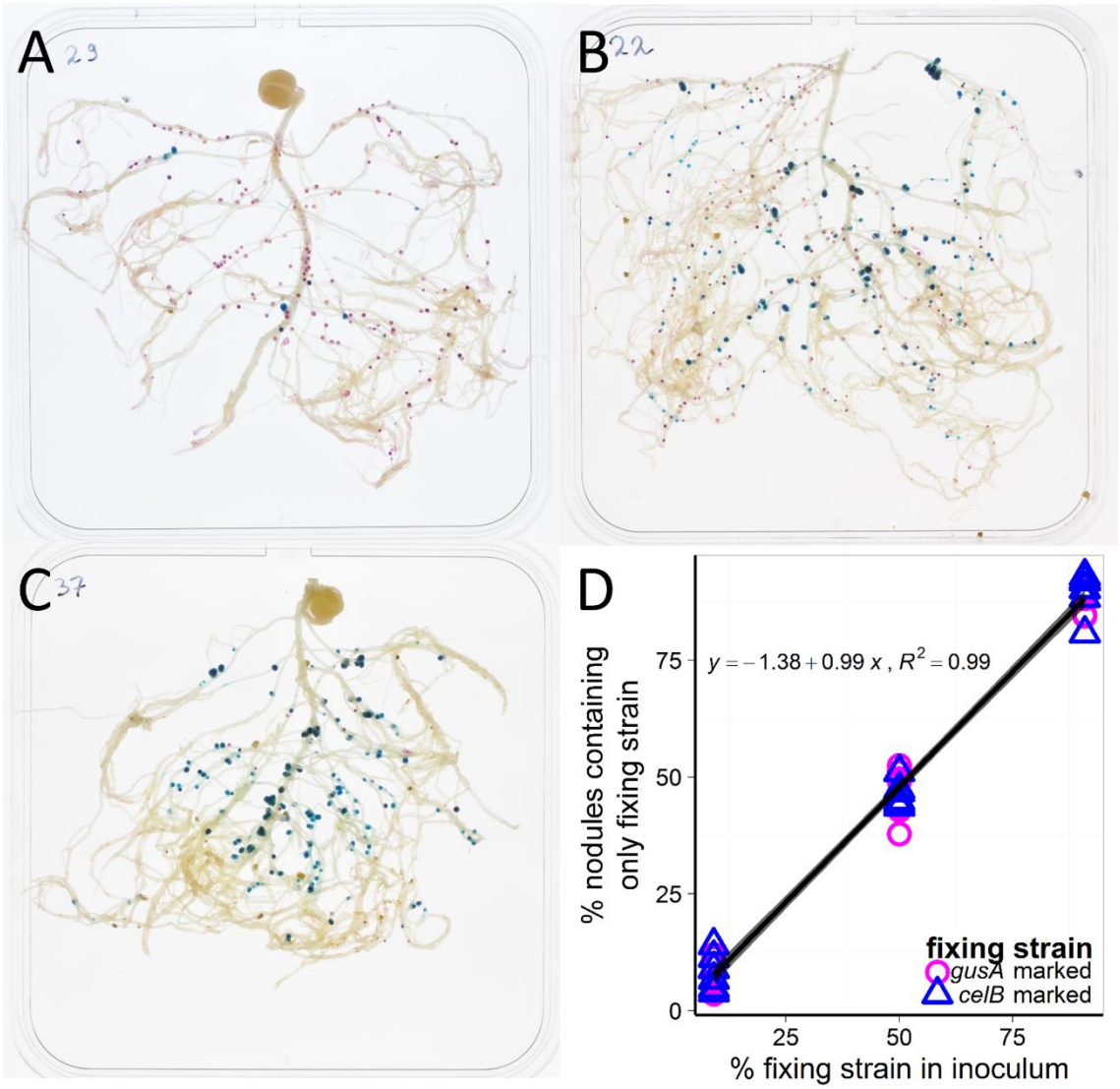


Figure 3-2. We found no evidence of partner choice. (**a**) – (**c**) show pea root systems with stained nodules following inoculation with different ratios of a fixing *celB* (blue) marked strain (3841 *celB*) and a non-fixing *gusA* (magenta) marked mutant strain (OPS0365): (**a**) 1:10 (fixing:non-fixing); (**b**) 1:1 (fixing:non-fixing); (**c**) 10:1 (fixing:non-fixing). (**d**) The percentage of nodules containing only the fixing strain exactly reflected the percentage of the fixing strain in the inoculum. The slope of the regression was 0.99 ± 0.02, 95% CI [0.95, 1.02] and this is not significantly different from one (t_{39} = -0.835, p = 0.41).

3.4.2 Nodule size

In the absence of partner choice, sanctions remain the only option for discriminating against non-fixing strains. If pea plants are able to sanction and differentially allocate resources on the basis of the nitrogen provided, we expect nodules containing a fixing strain to be larger than nodules containing a non-fixing strain. We therefore measured the sizes of a sample of nodules colonised by the different strains. Nodules containing the nitrogen-fixing strain were significantly larger than nodules containing the non-fixing strain (paired t-test, t_{13} = 7.7176, p = 3.307 x 10^{-6}), indicating that pea plants preferentially allocated resources to fixing nodules, supporting sanctioning. Fixing nodules had an average area of 1.20 ± 0.07 mm^2 (mean ± s.e.), and non-fixing nodules were 0.68 mm^2 (95% CI [0.49; 0.87]) smaller – less than half of the area of fixing nodules (**Figure 3-3**).

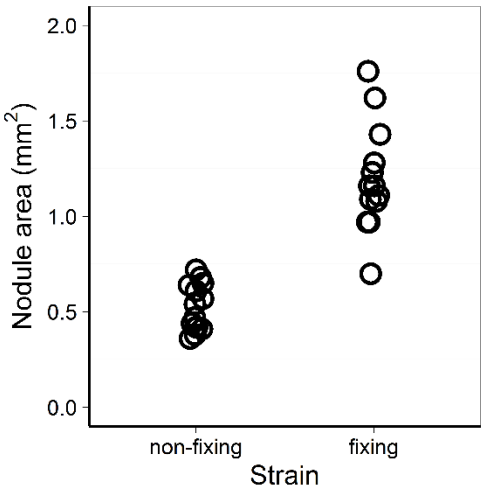


Figure 3-3. Legumes sanction non-fixing nodules. Nodules containing the fixing strain were significantly larger than nodules containing the fixing strain (paired t-test, t_{13} = 7.7176, p = 3.307 x 10^{-6}), indicating that pea plants preferentially allocated resources to fixing nodules.

3.4.3 Total number of nodules

Each pea plant formed 183 ± 12 nodules (mean ± s.e.) on average. The total number of nodules (**Figure 3-4**) decreased as the percentage of nitrogen-fixing nodules increased (slope = − 0.88 ± 0.34, t_{39} = -2.565, p = 0.0143), probably because the nodulation

process is inhibited once fixing nodules have been successfully established. A ten-fold increase in the percentage of nitrogen-fixing nodules resulted in 72 fewer nodules in total (95% CI [9, 135]) – a decrease of about 30%.

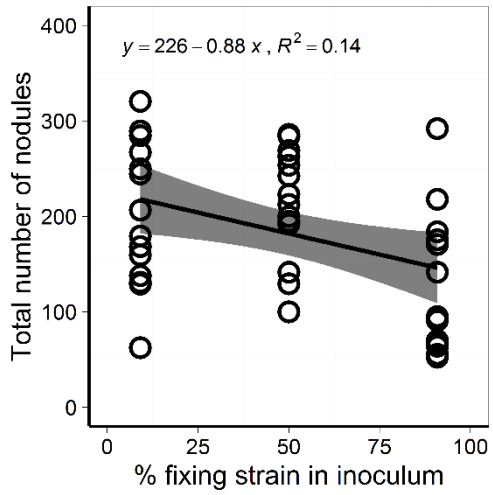


Figure 3-4. Total nodule number. The total number of nodules decreased as the percentage of fixing nodules increased (slope = − 0.88 ± 0.34, t_{39} = -2.565, p = 0.0143). Regression with 95% confidence interval shown.

3.4.4 Shoot mass

After three weeks of growth we found no differences in shoot dry mass among treatments, including water controls, ($F_{3, 44}$ = 2.1921, p = 0.1024), presumably because the pea seeds contain large reserves of nitrogen. However, after five weeks of growth, independently grown plants inoculated with non-fixing RU3940 weighed 0.32 ± 0.04 grams (mean ± s.e.) while plants inoculated with the fixing strain Rlv3841 weighed 0.68 ± 0.04 grams, a significant difference of 0.37 grams (95% CI [0.23, 0.51]).

3.4.5 Frequency of mixed nodules

The staining technique allowed clear visualization of mixed nodules, which occasionally appeared to result from multiple independent infections (**Figure 3-5A**), but normally consisted of two – a fixing and a non-fixing strain **(Figure 3-5B, C)**. Mixed nodules

occurred at an average frequency of 2.02% ± 0.35% (mean ± s.e.), and the frequency depended on the inoculum ratio (**Figure 3-5D**). According to a simple diffusion model (in which we assume that the two rhizobial strains in the soil are well mixed) the predicted frequency of mixed nodules (F) is simply given by $F = pq\varepsilon$, where p is the frequency of the fixing strain, q is the frequency of the non-fixing strain and ε is the unknown probability with which two bacteria simultaneously enter the same nodule; hence: $F \propto pq$. We regressed F against pq ('encounter rate', **Figure 3-5D**) and found that the frequency of mixed nodules indeed increased in proportion to the increase in encounter rate: a ± three-fold increase in encounter rate increased the odds of a mixed infection 3.5 times (95% CI [2.2, 5.9], $n = 41$). In an additional experiment, we varied the total inoculation density, as the diffusion model also predicts that mixed nodules occur more often at higher densities of rhizobia. As predicted, we found that the percentage of mixed nodules increased with inoculation density (slope = 1.87 ± 0.41, $t_{18} = 4.545$, $p = 0.000251$). A 10-fold increase in inoculation density, increased the percentage of mixed nodules by almost 2%.

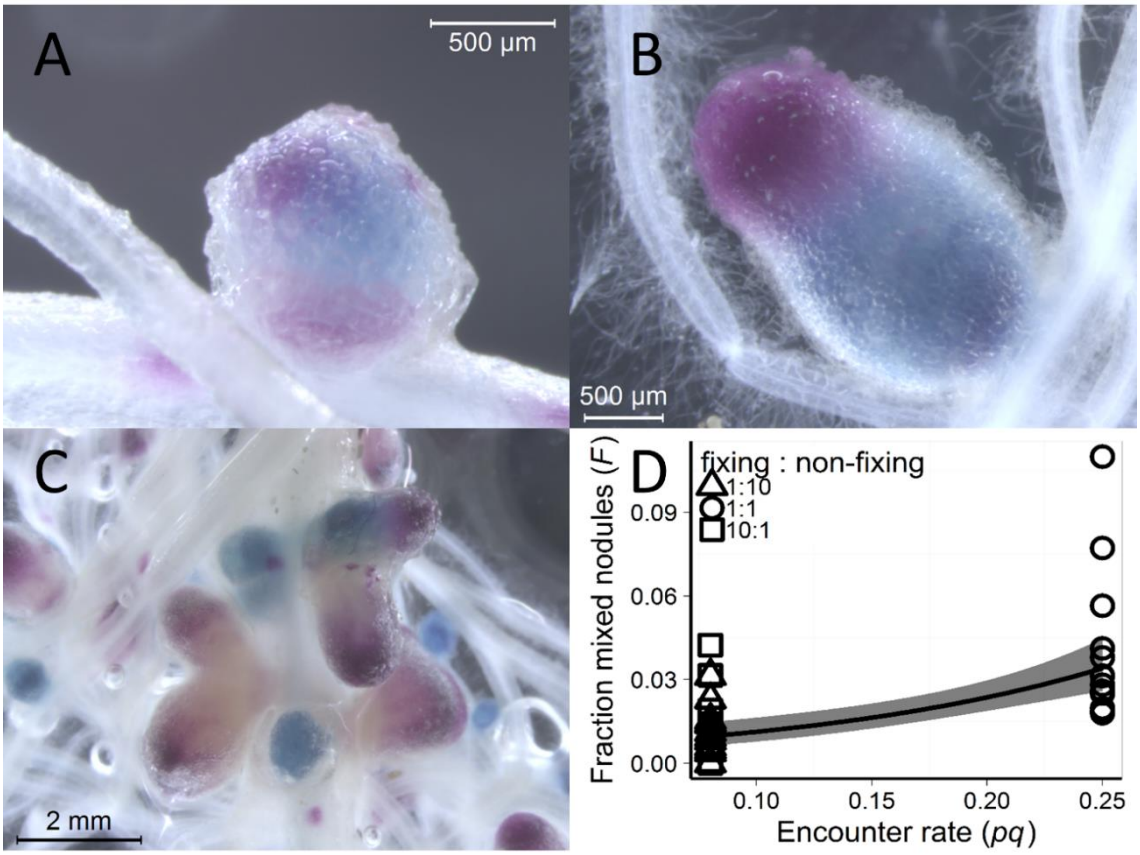


Figure 3-5. Mixed nodules. (**a**) Microscopic image of a mixed nodule on a pea root, which seems to have been infected by more than two rhizobia. The nitrogen-fixing strain is *celB* (blue) marked (3841 *celB*), and the non-fixing mutant strain is *gusA* (magenta) marked (OPS0365). (**b**) Microscopic image of a mixed nodule on a pea root infected by two rhizobia. The *gusA* marked nitrogen-fixing strain (3841 *gusA*) is magenta, and the *celB* marked non-fixing mutant strain (OPS0366) is blue. (**c**) Nodules, including mixed nodules, on a pea root. The *gusA* marked nitrogen-fixing strain (3841 *gusA*) is magenta, and the *celB* marked non-fixing mutant strain (OPS0366) is blue. (**d**) Mixed nodules occur more often when the probability that different strains encounter each other during nodule formation is higher (t_{39} = 4.913, p = 1.65 x 10-5). Generalised linear model with quasibinomial distribution and logit link function shown with 95% confidence interval.

3.4.6 Fitness of rhizobial strains within mixed nodules

To assess the relative fitness of fixing versus non-fixing strains within mixed nodules, we measured the area occupied by each strain as indicated by the area stained blue versus magenta (or vice versa). We found that the percentage of nodule area occupied by the fixing strain ranged from 1.1 to 99 %. Unfortunately, although the staining technique is valuable in visualizing mixed nodules, we conclude that it cannot be reliably used to assess the relative fitness of different strains within mixed nodules. The percentage of the nodule occupied by the fixing strain depended highly on the marker used to identify the fixing strain ($t_{218} = 14.49$, $p < 2 \times 10^{-16}$). If the fixing strain was marked with *gusA*, it appeared to occupy 31% of the nodule (95% CI [27.5; 33.6%]), but if the fixing strain was marked with *celB*, it appeared to occupy 75% of the nodule (95% CI [68.7; 80.0%].

3.5 Discussion

We created a non-fixing but otherwise isogenic mutant to mimic a process that might occur in nature, where a mutation arises in a fixing rhizobial strain, rendering it ineffective. We found that pea plants could not discriminate between these fixing and non-fixing strains prior to nodule formation. Pea plants were therefore unable to detect whether the potential partner was effective at fixing nitrogen and could not prevent the formation of non-fixing nodules. Our results are supported by an early study using similar isogenic strains[27], but this study was severely limited in sample size and has therefore been overlooked.

Our results indicate that partner choice is not a robust mechanism against ineffective strains as pea plants were unable to prevent non-fixing strains from entering. It could be argued that legumes may use genes other than *nifH* to assess the nitrogen fixation capacity of fixing strains before nodule formation; however we believe that this is unlikely. A mutation rendering a strain less effective can arise in any gene and effective partner choice would then require a mutation in the plant genome to detect this change. If the new mutation stops the cheat from entering, then it will spread through the plant population; however, given that rhizobial generation times are much faster than host plant generation times[28], it seems that the host plants will be locked in an evolutionary arms race that they are doomed to lose; hence partner choice seems to be an ineffective way to stabilise the mutualism in the long term[21]. Furthermore, partner choice is susceptible to dishonest signals[13].

In contrast, we found that pea plants did discriminate against ineffective strains via sanctioning, which has been previously reported using argon gas to replace atmospheric nitrogen[19–21]. In our experiment, nodules containing the non-fixing strain were roughly half the size of fixing nodules, indicating reduced plant resources. In

contrast to the case of partner choice, sanctioning can stabilise the mutualism in the long-term. If a mutation arises that allows a plant to detect and sanction a partner that is not delivering the goods it would be effective against a wide variety of future ineffective strains. Thus, sanctioning allows an instantaneous response to ineffective strains and does not require specific recognition genes or rely on honest signals. Sanctioning is therefore a more robust[13] mechanism against ineffective strains and can provide long-term stability to legume-rhizobia mutualisms[16,29].

However, any discussion of sanctioning should take into account both plant and rhizobial fitness. Sanctioning can only be selected for when it saves plant resources and thus increases plant fitness. The reduced nodule size that we and others[20,21] have seen indicates that plants allocate fewer resources to non-fixing nodules. Whether this reduced resource allocation also reduces rhizobial fitness and thus stabilises the mutualism on evolutionary time-scales is more difficult to establish. This may depend on whether or not the nitrogen-fixing bacteroids can still reproduce (usually determinate nodules), or are terminally differentiated and unable to reproduce (indeterminate nodules)[30]. In studies using argon gas, reduced rhizobial fitness has been shown in both determinate[19,20] and indeterminate[21] non-fixing nodules. However, in a study using a non-fixing isogenic strain, the fitness of non-fixing rhizobia was not reduced in determinate soybean nodules up to five weeks old[31]. Because effects on rhizobial viability may emerge later in the sanctioning process, perhaps the best test would be a multigenerational experiment, where in the first generation plants are inoculated with both fixing and non-fixing strains, and new plants are then repeatedly grown in the same soil for several generations to see how quickly the non-fixing strain is eliminated. It would be of special interest to perform such an experiment with both indeterminate and determinate species.

Evidence from other mutualisms suggests that whether partner choice evolves in preference to sanctioning critically depends on how well partners can assess quality prior to establishment of the mutualistic relationship including the potential for dishonest signalling. For example, clients of the cleaner fish *Labroides dimidiatus* have evolved partner choice to counter cheating by individuals that take healthy tissue while removing parasites[32]. Partner choice is highly effective in this mutualism because the quality of service is known from previous experience and there are repeated interactions between individuals. In contrast, sanctioning has evolved in mutualisms between yucca moths and fig wasps and their respective plant hosts. In both cases the insects deposit seed-eating larvae in the flowers of host plants in return for pollination. In these mutualisms, plants cannot prevent eggs being laid nor assess partner quality, hence sanctioning has evolved: flowers containing too many eggs[33], or too little pollen[33,34] are selectively aborted.

While our study shows that partner choice is not a robust mechanism to exclude ineffective strains, legumes do not form symbioses with all potential rhizobial strains. Instead, an extensive signalling process[35–37] between legumes and their rhizobial partners can impose a high degree of selectivity on the relationship[36,38,39], although the degree of selectivity varies greatly among hosts[39]. There are two explanations for this selectivity that are commonly proposed and are not mutually exclusive. First, specificity may arise in order to prevent the entry of pathogenic bacteria which utilise similar signalling pathways to gain access to host roots[38,40,41]. Second, by fine-tuning signalling pathways to target rhizobia that are particularly effective for a specific host, legume species might achieve greater nitrogen-fixation efficiency[42,43]. This is likely to be true if host environments are sufficiently different that specialization by rhizobia is selected for. Support for specialization comes from the observation that a single rhizobial strain can vary greatly in its effectiveness among hosts[42]. This type of co-evolutionary process is

separate from the need to avoid non-fixing rhizobia, which can arise by mutation at any time, in any strain, even those that are usually highly effective. That these two processes are indeed separate is supported by the fact that the genes involved in nitrogen fixation (*nif* and *fix* genes) are only expressed once the symbiosis has been established[44] and are different from the signalling genes involved in infection (*nod* genes)[45]. Currently, genomic analyses are shedding more light on the selective pressures affecting both legume and rhizobial genes[46–48].

Sanctioning is a robust mechanism against ineffective strains, but requires hosts to monitor partner quality and provide resources accordingly. Currently, little is known about the exact mechanism behind sanctioning in legume-rhizobia symbioses, and whether it only takes place at the nodule level, or also occurs within nodules[13,16]. If sanctioning takes place at the nodule level, mixed nodules could be a way for ineffective strains to avoid sanctions[6,13,49]. Indeed certain endosymbionts, even those belonging to different genera and lacking any genes for nitrogen fixation, have been shown to co-infect nodules by "piggybacking" on the genuine symbionts as they infect the root hairs[50]. Whether mixed nodules allow ineffective strains to persist depends on the frequency of mixed nodules, and on the relative fitness of fixing and non-fixing strains within mixed nodules. Estimates of the frequency of mixed nodules in the literature range from 2% to 74%[6,51]. Our findings at least partly explain this variability as the frequencies we found could be adequately represented by a simple diffusion model, which predicts that more mixed nodules are expected: (1) at high rhizobial densities; (2) when the proportions of different strains are similar; and (3) when rhizobia diffuse more easily, which might occur, for example, under wet conditions. Although the staining technique is valuable in identifying strains and characterising mixed nodules, it could not be used reliably to assess fitness of rhizobial strains within mixed nodules. Further work

on the mechanism of sanctioning, how it is affected by external conditions such as soil nitrogen, and how it affects rhizobial fitness will help illuminate how the legume-rhizobia mutualism has persisted for much longer than humans have been around to reap its benefits.

3.6 Methods

3.6.1 Bacterial strains and culture conditions

We used *Rhizobium leguminosarum* bv. viciae (*Rlv*) 3841 strains labelled at identical positions in the chromosome with either a *gusA* or a *celB* marker gene[52] (Table 1). Non-fixing mutants of these strains were made by replacing the *nifH* gene with a *nifH* gene disrupted by a spectinomycin resistance cassette (*nifH::ΩSpc*). *Rhizobium leguminosarum* only has a single copy of the *nifH* gene[53], which is essential for nitrogen fixation[54] . First, an unmarked non-fixing strain was made, then the non-functional *nifH* gene was transduced into the marked fixing strains. To make the unmarked non-fixing strain, the *nifH* gene was amplified by PCR (2,660 bp product) using primers p950 and p951 that contain a SacI and SpeI restriction site and cloned into pCR2.1 in Dam- *E.coli* SCS110 (Stratagene) (pRU1907). A *Ω* spectinomycin cassette from pHP45Ω-Sp was cloned between two NruI sites within *nifH* resulting in a 166 bp deletion, and the *nifH::ΩSpc* fragment was cloned into the suicide vector pJQ200SK (pRU1908). Plasmid pRU1908 was conjugated into Rlv3841, and cells were plated on TY agar with gentamicin (20 μg ml^{-1}) and then on AMS agar supplemented with 10% sucrose, 10 mM NH_4Cl, and spectinomycin (100 μg ml^{-1}) to select for gene replacement. Strain RU3940 was found to contain the correct *nifH::ΩSpc* mutation by PCR mapping using pOT forward with either p1002 or p1118 in separate reactions. To confirm differences in the nitrogen-fixing capacity of the two strains, we performed acetylene reductions on plants inoculated with either RU3940 or Rlv3841. RU3940 does not fix nitrogen (0.05 ± 0.02 (mean ± s.e, $n = 3$) μmol ethylene per plant per hour, which is not different from water control plants (t_6 = -0.025, p = 0.98)), whereas the parent strain Rlv3841 does fix nitrogen (2.58 ± 0.33 (mean ± s.e, $n = 3$) μmol ethylene per plant per hour).

The *nifH::ΩSpc* cassette from non-fixing strain RU3940 was transduced into the *gusA* or *celB* marked strains using phage RL38[55], yielding non-fixing *gusA* (OPS0365) and *celB* (OPS0366) marked strains which are otherwise isogenic to their fixing *gusA* and *celB* marked parent strains. Correct insertion of the *nifH::ΩSpc* cassette from RU3940 was confirmed by PCR mapping using pOT forward with either oxp460 or opx461 in separate reactions (Supplementary Fig. S1 online). Enzyme assays on free living cultures confirmed conservation of the *gusA* and *celB* marker genes in the non-fixing mutants (Supplementary Fig. S2 and Supplementary Fig. S3 respectively). Bacterial cultures were maintained on TY agar[56] with 500 μg ml^{-1} streptomycin (all strains) and 100 μg ml^{-1} spectinomycin (RU3940, OPS0365 and OPS0366 only).

3.6.2 Plant growth

Pea (*Pisum sativum* cv Avola) seeds were germinated in the dark for five days on agar plates at room temperature and then transferred to 500 ml pots containing a mixture of silver sand - fine vermiculite (1:1 v/v) substrate, 75 ml nitrogen free nutrient solution (as in Poole *et al.*[57] but 2.67 times more concentrated) and 1 ml rhizobial inoculum. Peas were grown in the growth room at 21 °C with a 16 hour photoperiod.

3.6.3 Experimental designs

To test for partner choice, we applied rhizobial inocula consisting of 1:10, 1:1 and 10:1 ratios of fixing to non-fixing strains. To exclude any effect of the *gusA* or *celB* marker genes, both combinations were tested: fixing *gusA* (3841 *gusA*) with non-fixing *celB* (OPS0366) strains and fixing *celB* (3841 *celB*) with non-fixing *gusA* (OPS0365) strains. Total inoculum density for all treatments was 1 x 10^4 cells per pot. Rhizobial cultures for the inocula were spread-plated to confirm numbers of rhizobia in the inocula. In a fully

randomised design, we grew 49 plants in total: six treatments plus a water control (no rhizobia), all replicated seven times. Plants were grown for 21 days without additional watering.

To assess the effect of inoculation density on the frequency of mixed nodules, plants were inoculated with 1:1 ratios of fixing *gusA* (3841 *gusA*) to non-fixing *celB* (OPS0366) strains at total densities of 1 x 10^2, 1 x 10^3, 1 x 10^6 and 1x 10^8 cells per pot. In a fully randomised design, we grew five plants at each inoculum density plus two water controls. Plants were grown for 38 days and watered with sterilised water as needed after 21 days.

To test for differences in markers on the percentage of the nodule area occupied by the fixing strain, six plants were grown in an additional independent experiment, inoculated with 1:1 ratios of fixing *celB* (3841 *celB*) to non-fixing *gusA* strains (OPS0365) at a total density of 1 x 10^8 cells per pot. Plants were grown for 37 days and watered with sterilised water as needs after 21 days.

To assess the phenotype of non-fixing mutant strain RU3940, plants were inoculated with either fixing strain Rlv3841 or non-fixing strain RU3940 and grown for four or five weeks in independent experiments.

3.6.4 Harvest

At harvest, roots were gently washed and then stained for *gusA* and *celB* marker genes[58]. In aluminium covered tubes, roots were incubated overnight at 28 °C submerged in phosphate buffer (7 g L^{-1} NaH_2PO_4, 7.2 g L^{-1} Na_2HPO_4, 1 mM EDTA (pH 8), 1% Sarkosyl, 1 ml L^{-1} Triton) supplemented with 0.2 mg ml^{-1} Magenta-glc (5-bromo-6-chloro-3-indolyl-β-D-glucuronide). In fresh phosphate buffer, roots were then incubated at 70 °C for 1 hour 45 min. After cooling down, X-gal (5-bromo-4-chloro-3-indolyl-β-D-

galactopyranoside) was added to a final concentration of 0.25 mg ml^{-1} and roots were incubated overnight at 37 °C. This yielded visibly magenta (*gusA* marked strains), blue (*celB* marked strains) or magenta and blue (mixed) nodules (Fig. 2 A-C**)**. Shoots were dried at 70 degrees Celsius for 24-72 hours and then weighed.

3.6.5 Data collection

Roots were carefully laid out so that all nodules were visible and then photographed. Two people independently counted blue and magenta nodules from photographs taken of whole roots, zooming in as necessary, and the average was taken as the final count. Mixed nodules were counted and photographed using a dissecting microscope (Leica M165 FC) with accompanying software (LAS v4.5). To estimate size of nodules containing fixing versus non-fixing strains, we measured the area of nodules from the photographs taken of the whole root, using ImageJ v1.49v[59], which allowed for measuring areas of irregular shapes. Ten fixing and ten non-fixing nodules were randomly selected for each of the 14 plants inoculated with a 1:1 ratio of fixing to non-fixing strains (7 of which had the *gusA* marked fixing strain and 7 had the *celB* marked fixing strain). To estimate the percentage of mixed nodules that was occupied by the fixing strain, we measured the areas of blue and magenta in all mixed nodules for which this was possible (distinct enough areas of blue and magenta), using ImageJ. Additional microscopy was done on whole nodules (five-week-old nodules). Images were taken with a dissecting microscope (Leica M165 FC) with accompanying software (LAS v4.5).

3.6.6 Statistical analyses

We tested for partner choice by regressing the percentage of nodules containing nitrogen-fixing rhizobia against the percentage of nitrogen-fixing rhizobia in the inoculum. If

plants do not exert partner choice, we expect a 1:1 relationship (a slope of 1.0). Water controls and mixed nodules were excluded from this analysis. To test whether the presence of *gusA* versus *celB* marker genes affected competitive ability, we carried out an ANCOVA with the marker gene as the categorical variable. Results presented are from the regression model without this interaction, as marker genes did not affect the slope ($t_{37} = -0.189$, $p = 0.851$) or intercept ($t_{37} = -0.177$, $p = 0.861$) of the regression line. The intended inoculation ratios were used as the explanatory variable, as it was confirmed using colony counts that the actual ratios in inocula did not deviate by more than 11% from intended inoculation ratios ($n = 6$, mean ± s.e. 3.78 ± 1.49 %, Supplementary Table S1 online). Accounting for this deviation did not change results, and if anything would lead to a decrease of the estimate of the slope (whereas for partner choice we would expect an increase of the estimate of the slope).

To test for differences in size between fixing and non-fixing nodules, we performed a paired t-test on the average size of ten nodules, pairing fixing and non-fixing nodules of each plant. To test whether the total number of nodules formed depended on the ratio of fixing to non-fixing strains in the inoculum, we regressed the total number of nodules against the percentage of fixing strain in the inoculum. Shoot mass was also regressed against the percentage of the fixing strain in the inoculum. To assess the frequency of mixed nodules, we analysed how the occurrence of mixed nodules depended on the inoculation ratio of fixing to non-fixing strains with a generalised linear model using a quasibinomial distribution and a logit link function (because data were expressed as proportions and there was significant overdispersion). We also regressed the percentage of mixed nodules against the log of the total inoculation density. Whether the fixing strain was marked with *gusA* or *celB* had no impact on the proportion of mixed nodules detected ($t_{37} = 0.494$, $p = 0.62493$) or on the interaction with the encounter rate

(t_{37} = 1.137, p = 0.26301), so we present results from analyses which do not include this interaction. To assess fitness of fixing to non-fixing strains within mixed nodules, we used a linear model testing for the effect of the marker (*gusA* or *celB*) used to mark the fixing strain. Individual nodules were seen as the unit of replication, and this was confirmed to be a valid assumption as the plant from which nodules came accounted for only 9% of the variation in the data. All analyses were performed in R version 3.0.2[60].

Table 3-1. Strains, plasmids and primers.

Strain, plasmid, or primer	**Genotype or sequence**	**Reference of source**
Strains		
Rlv3841	St^r derivative of *R. leguminosarum* bv. *viciae* strain 300	[61]
RU3940	Rlv3841 *nifH::Ω* Sp^r	This work
3841 *gusA*	Rlv3841 with *gusA* marker	[52]
3841 *celB*	Rlv3841 with *celB* marker	[52]
OPS0365	3841 *gusA nifH::Ω* Sp^r	This work
OPS0366	3841 *celB nifH::Ω* Sp^r	This work
Plasmids		
pRU1907	*nifH* cloned in pCR2.1	This work
pHP45Ω-Sp	Plasmid containing Ω spec cassette, Sp^r	[62]
pJQ200SK	pACYC derivative, P15A origin of replication, Gm^r	[63]
pRU1908	*nifH::Ω*, Sp^r in pJQ200SK	This work

Primers	
pOTforward	CGGTTTACAAGCATAAAGC
p1002	TTCCTCCATGTGCCTGGAGA
p1118	GGTTCTTCGGAGTTTCTAT
oxp0460	GCTTGATCATCGCCGGAAAC
oxp0461	TGTCACCGCCGAAAACGATG

3.7 References

1. Lebauer, D. S. & Treseder, K. K. Nitrogen limitation of net primary productivity in terrestrial ecosystems is globally distritubed. *Ecology* **89,** 371–379 (2008).
2. Sprent, J. I. Evolving ideas of legume evolution and diversity: a taxonomic perspective on the occurrence of nodulation. *New Phytol.* **174,** 11–25 (2007).
3. Graham, P. H. & Vance, C. P. Legumes: importance and constraints to greater use. *Plant Physiol.* **131,** 872–877 (2003).
4. Hafeez, F., Hameed, S., Ahmad, T. & Malik, K. A. Competition between effective and less effective strains of *Bradyrhizobium* spp. for nodulation on *Vigna radiata*. *Biol. Fertil. Soils* **33,** 382–386 (2001).
5. Kiers, E. T., Hutton, M. G. & Denison, R. F. Human selection and the relaxation of legume defences against ineffective rhizobia. *Proc. R. Soc. B.* **274,** 3119–26 (2007).
6. Denison, R. F. Legume sanctions and the evolution of symbiotic cooperation by rhizobia. *Am. Nat.* **156,** 567–576 (2000).
7. Heath, K. & Tiffin, P. Stabilizing mechanisms in a legume–rhizobium mutualism. *Evolution (N. Y).* **63,** 652–662 (2009).
8. Udvardi, M. & Poole, P. S. Transport and metabolism in legume-rhizobia symbioses. *Annu. Rev. Plant Biol.* **64,** 781–805 (2013).
9. Gage, D. J. Analysis of infection thread development using Gfp- and DsRed-expressing *Sinorhizobium meliloti*. *J. Bacteriol.* **184,** 7042–7046 (2002).
10. Oldroyd, G. E. D. & Downie, J. A. Coordinating nodule morphogenesis with rhizobial infection in legumes. *Annu. Rev. Plant Biol.* **59,** 519–46 (2008).
11. Hagen, M. & Hamrick, J. Population level processes in *Rhizobium leguminosarum* bv. *trifolii*: the role of founder effects. *Mol. Ecol.* **5,** 707–714 (1996).
12. Silva, C., Eguiarte, L. E. & Souza, V. Reticulated and epidemic population genetic structure of *Rhizobium etli* biovar *phaseoli* in a traditionally managed locality in Mexico. *Mol. Ecol.* **8,** 277–287 (1999).

13. Kiers, E. T. & Denison, R. F. Sanctions, cooperation, and the stability of plant-rhizosphere mutualisms. *Annu. Rev. Ecol. Evol. Syst.* **39,** 215–236 (2008).

14. Gubry-Rangin, C., Garcia, M. & Béna, G. Partner choice in *Medicago truncatula-Sinorhizobium* symbiosis. *Proc. Biol. Sci.* **277,** 1947–51 (2010).

15. Regus, J., Gano, K. A., Hollowell, A. C. & Sachs, J. L. Efficiency of partner choice and sanctions in *Lotus* is not altered by nitrogen fertilization. *Proc. R. Soc. B.* **281,** 20132587 (2014).

16. West, S. A., Kiers, E. T., Pen, I. & Denison, R. F. Sanctions and mutualism stability: when should less beneficial mutualists be tolerated? *J. Evol. Biol.* **15,** 830–837 (2002).

17. Simms, E. & Taylor, D. Partner choice in nitrogen-fixation mutualisms of legumes and rhizobia. *Integr. Comp. Biol.* **42,** 369–380 (2002).

18. Simms, E. L. *et al.* An empirical test of partner choice mechanisms in a wild legume-rhizobium interaction. *Proc. Biol. Sci.* **273,** 77–81 (2006).

19. Kiers, E. T., Rousseau, R. A., West, S. A. & Denison, R. F. Host sanctions and the legume-rhizobium mutualism. *Nature* **425,** 78–81 (2003).

20. Kiers, T. E., Rousseau, R. A. & Denison, R. F. Measured sanctions: legume hosts detect quantitative variation in rhizobium cooperation and punish accordingly. *Evol. Ecol. Res.* **8,** 1077–1086 (2006).

21. Oono, R., Anderson, C. G. & Denison, R. F. Failure to fix nitrogen by non-reproductive symbiotic rhizobia triggers host sanctions that reduce fitness of their reproductive clonemates. *Proc. R. Soc. B.* **278,** 2698–703 (2011).

22. Malek, W. The role of motility in the efficiency of nodulation by *Rhizobium meliloti*. *Arch. Microbiol.* **158,** 26–28 (1992).

23. Robleto, E. A., Kmiecik, K., Oplinger, E. S., Nienhuis, J. & Triplett, E. W. Trifolitoxin production increases nodulation competitiveness of *Rhizobium etli* CE3 under agricultural conditions. *Appl. Environ. Microbiol.* **64,** 2630–2633 (1998).

24. Frederix, M. *et al.* Mutation of *praR* in *Rhizobium leguminosarum* enhances root biofilms, improving nodulation competitiveness by increased expression of attachment proteins.

Mol. Microbiol. **93,** 464–478 (2014).

25. Triplett, E. W. & Sadowsky, M. Genetics of competition for nodulation of legumes. *Annu. Rev. Microbiol.* **46,** 399–428 (1992).

26. Friesen, M. L. Widespread fitness alignment in the legume – rhizobium symbiosis. *New Phytol.* **194,** 1096–1111 (2012).

27. Hahn, M. & Studer, D. Competitiveness of a *nif*⁻ *Bradyrhizobium japonicum* mutant against the wild-type strain. *FEMS Microbiol. Lett.* **33,** 143–148 (1986).

28. Gordon, B. R. *et al.* Decoupled genomic elements and the evolution of partner quality in nitrogen-fixing rhizobia. *Ecol. Evol.* **6(5),** 1317–1327 (2016).

29. West, S. A., Kiers, E. T., Simms, E. L. & Denison, R. F. Sanctions and mutualism stability: why do rhizobia fix nitrogen? *Proc. R. Soc. B.* **269,** 685–94 (2002).

30. Oono, R., Denison, R. & Kiers, E. Controlling the reproductive fate of rhizobia: how universal are legume sanctions? *New Phytol.* **183,** 967–979 (2009).

31. Marco, D. E. *et al.* An experimental and modelling exploration of the host-sanction hypothesis in legume-rhizobia mutualism. *J. Theor. Biol.* **259,** 423–33 (2009).

32. Bshary, R. & Schäffer, D. Choosy reef fish select cleaner fish that provide high-quality service. *Anim. Behav.* **63,** 557–564 (2002).

33. Pellmyr, O. & Huth, C. J. Evolutionary stability of mutualism between yuccas and yucca moths. *Nature* **372,** 257–260 (1994).

34. Tarachai, Y., Compton, S. G. & Trisonthi, C. The benefits of pollination for a fig wasp. *Symbiosis* **45,** 29–32 (2008).

35. Oldroyd, G. E. D. & Downie, J. A. Calcium, kinases and nodulation signalling in legumes. *Nat. Rev. Mol. Cell Biol.* **5,** 566–576 (2004).

36. Garg, N. Symbiotic nitrogen fixation in legume nodules: process and signaling. A review. *Agron. Sustain. Dev.* **27,** 59–68 (2007).

37. Murray, J. D. Invasion by invitation: rhizobial infection in legumes. *Mol. Plant. Microbe. Interact.* **24,** 631–639 (2011).

38. Yang, S., Tang, F., Gao, M., Krishnan, H. B. & Zhu, H. *R* gene-controlled host specificity

in the legume-rhizobia symbiosis. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 18735–18740 (2010).

39. Lira, M. A., Nascimento, L. R. S. & Fracetto, G. G. M. Legume-rhizobia signal exchange: promiscuity and environmental effects. *Front. Microbiol.* **6,** 1–9 (2015).

40. Jones, K. M. *et al.* Differential response of the plant *Medicago truncatula* to its symbiont *Sinorhizobium meliloti* or an exopolysaccharide-deficient mutant. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 704–709 (2008).

41. Via, V. D., Zanetti, M. E. & Blanco, F. How legumes recognize rhizobia. *Plant Signal. Behav.* **11,** e1120396 (2015).

42. Young, J. P. & Johnston, A. W. The evolution of specificity in the legume-rhizobium symbiosis. *Trends Ecol. Evol.* **4,** 341–349 (1989).

43. Barrett, L. G., Zee, P. C., Bever, J. D., Miller, J. T. & Thrall, P. H. Evolutionary history shapes patterns of mutualistic benefit in *Acacia*-rhizobial interactions. *Evolution (N. Y).* 1473–1485 (2016).

44. Terpolilli, J. J., Hood, G. A. & Poole, P. S. What determines the efficiency of N_2-fixing *Rhizobium*-legume symbioses? *Adv. Microb. Physiol.* **60,** 325–89 (2012).

45. Downie, J. A. The roles of extracellular proteins, polysaccharides and signals in the interactions of rhizobia with legume roots. *FEMS Microbiol. Rev.* **34,** 150–170 (2010).

46. Batstone, R. T., Dutton, E. M., Wang, D., Yang, M. & Frederickson, M. E. The evolution of symbiont preference traits in the model legume *Medicago truncatula*. *New Phytol.* (2016).

47. Grillo, M. A., Mita, S. De, Burke, P. V, Sol, K. L. S. & Heath, K. D. Intrapopulation genomics in a model mutualist: population structure and candidate symbiosis genes under selection in *Medicago truncatula*. *Evolution (N. Y).* **70,** 2704–2717 (2016).

48. Yoder, J. B. Understanding the coevolutionary dynamics of mutualism with population genomics. *Am. J. Bot.* **103,** 1742–1752 (2016).

49. Steidinger, B. S. & Bever, J. D. Host discrimination in modular mutualisms: a theoretical framework for meta-populations of mutualists and exploiters. *Proc. R. Soc. B.* **283,**

20152428 (2016).

50. Zgadzaj, R., James, E. K., Kelly, S., Kawaharada, Y. & De, N. Infection of nodules by symbiotic and endophytic bacteria. *PLoS Genet.* **11,** 1–21 (2015).

51. Moawad, H. A., Ellis, W. R. & Schmidt, E. L. Rhizosphere response as a factor in competition among three serogroups of indigenous *Rhizobium japonicum* for nodulation of field-grown soybeans. *Appl. Environ. Microbiol.* **47,** 607–612 (1984).

52. Sánchez-Cañizares, C. & Palacios, J. Construction of a marker system for the evaluation of competitiveness for legume nodulation in *Rhizobium* strains. *J. Microbiol. Methods* **92,** 246–249 (2013).

53. Young, J. P. W. *et al.* The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome Biol.* **7,** R34 (2006).

54. Fischer, H. M. Genetic regulation of nitrogen fixation in rhizobia. *Microbiol. Rev.* **58,** 352–386 (1994).

55. Buchanan-Wollaston, V. Generalized transduction in *Rhizobium leguminosarum*. *J. Gen. Microbiol.* **112,** 135–142 (1979).

56. Beringer, J. E. R factor transfer in *Rhizobium leguminosarum*. *J. Gen. Microbiol.* **84,** 188–198 (1974).

57. Poole, P. S., Blyth, A., Reid, C. J. & Walters, K. myo-Inositol catabolism and catabolite regulation in *Rhizobium leguminosarum* bv. *viciae*. *Microbiology* **140,** 2787–2795 (1994).

58. Sessitsch, A., Wilson, K. J., Akkermans, A. D. L. & De Vos, W. M. Simultaneous detection of different *Rhizobium* strains marked with either the *Escherichia coli gusA* gene or the *Pyrococcus furiosus celB* gene. *Appl. Environ. Microbiol.* **62,** 4191–4194 (1996).

59. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9,** 671–675 (2012).

60. R Core Team. R: A language and environment for statistical computing. (2013).

61. Johnston, A. W. & Beringer, J. E. Identification of the rhizobium strains in pea root nodules using genetic markers. *J. Gen. Microbiol.* **87,** 343–350 (1975).

62. Prentki, P. & Krisch, H. M. In vitro insertional mutagenesis with a selectable DNA fragment. *Gene* **29,** 303–313 (1984).

63. Quandt, J. & Hynes, M. F. Versatile suicide vectors which allow direct selection for gene replacement in Gram-negative bacteria. *Gene* **127,** 15–21 (1993).

4 Mathematical modelling of the legume-rhizobia symbiosis

'[…] it is always advisable to perceive clearly our ignorance.'

Charles Darwin, The expression of the emotions in man and animals, 1872

4.1 Overview

This chapter describes a mathematical model of plant growth which was used to investigate plant resource allocation to rhizobia in the case of (*i*) a single strain and (*ii*) multiple strains, under variable soil nitrogen conditions. The aim of the modelling is to formulate hypotheses, which are then tested experimentally (**chapter 5**). In contrast to other chapters, methods are described before the results, because a description of the model is needed for the results to make sense. Mathematical derivations are described in numbered boxes, but full verbal explanations are given in the main text as well. The results are to be published together with the results of **chapter 5**. An excerpt of the R code used to implement the model, showing the approach followed, can be found in Appendix C. The full version of the R code is available in the Oxford Research Archive (ORA-Data) repository, DOI: 10.5287/bodleian:BDVypApax.

4.2 Abstract

Plants of the legume family can obtain nitrogen via both root uptake and rhizobial nitrogen fixation. Rhizobial strains vary in how much nitrogen they provide (effectiveness) and in how efficiently they do so, both of which may affect plant growth.

It has been suggested that plants sanction strains providing little benefit by reducing resource allocation. Sanctioning can be selected for if plant growth is enhanced by differentially allocating resources to rhizobial strains. But it is unclear how plants should allocate resources among multiple strains to optimise growth, and whether optimal resource allocation depends only on how much nitrogen strains provide or also on how much carbon they demand.

We used mathematical modelling to establish how plants should allocate resources to maximise growth, and therefore fitness and yield. The model is based on the assumption that plant growth is either carbon- or nitrogen-limited. Analytical solutions were derived for optimal resource allocation to (*i*) a single rhizobial strain, and (*ii*) multiple rhizobial strains, under variable soil nitrogen conditions. The model explicitly describes easily observable outcomes – the dry mass of different plant tissues – and can therefore be used to formulate testable hypotheses.

We found that strains with high carbon demands should only be invested in by strongly nitrogen-limited plants, so that a threshold soil nitrogen concentration can be defined for each rhizobial strain, beyond which they should no longer receive resources. The value of this threshold concentration depends on both the effectiveness and the efficiency of the strain. We further predict that it is always optimal to allocate all resources to the single best available rhizobial strain. Dividing resources among multiple strains is always sub-optimal, but different strains may be optimal under different conditions. We describe experiments which can test these hypotheses.

4.3 Introduction

The symbiosis with nitrogen-fixing rhizobia provides plants of the legume family with an alternative source of nitrogen besides root uptake of soil nitrogen. The symbiosis can enhance plant growth, as nitrogen is the limiting nutrient in many soils[1]. But the nitrogen from the symbiosis comes at a carbon cost, which may be considerable, as up to 14% of newly sequestered carbon is allocated to nodules[2]. Rhizobial strains vary in how much nitrogen they provide (effectiveness)[3–7] and in the carbon cost per gram nitrogen fixed (efficiency)[2,8,9], both of which may affect plant growth. The effects on plant growth of the variability in rhizobial strains has been studied from an evolutionary biology perspective[10–14] and from an applied, agricultural perspective[15–19]. Here we address outstanding questions shared by both fields.

Evolutionary biologists have studied the legume-rhizobia symbiosis asking how the mutualism has persisted over time, and predicted that 'cheating' rhizobial strains are likely to arise which increase their fitness by (partially) forgoing nitrogen fixation and instead diverting carbon obtained from the plant into storage compounds or the reproduction of non-fixing bacteroids[10]. It was suggested that plants sanction such cheating strains by providing fewer resources, thus decreasing their fitness and (as a side-effect) preventing their spread through the rhizobial population[10]. Previous modelling has shown that sanctioning could be selected for in plants[11] and that this in turn leads to the evolution of increased levels of nitrogen fixation in rhizobial strains[12]. Experiments using argon gas[13,20,14], and *nifH* mutant strains which are non-fixing (**chapter 3**[21]) indicated that effectively fixing nodules are larger, and thus received more resources, than less effective nodules. But although one study included intermediate rates of nitrogen fixation[20], it remains unclear how plants would allocate resources among the multiple strains they are usually infected with[22–25]. In addition, it is generally assumed

that cheating strains provide less nitrogen, since part of the carbon obtained from the plant is directed into increasing rhizobial fitness rather than into nitrogen fixation. However, it has been argued that plants are not usually carbon limited[26], so alternatively plants may provide more carbon to cheating strains to keep up the rate of nitrogen fixation, especially when nitrogen is limiting plant growth. Therefore, another question is whether sanctioning is based just on how much nitrogen is provided, or also on carbon costs per gram nitrogen fixed.

From an agricultural perspective, the varied effects of rhizobial strains on yields have led to a search for elite inoculant strains[18]. Again, the distinction between effective and efficient strains is not usually made, as measures such as yield differences confound both traits. Although it has been acknowledged that plants may differentially allocate resources to nodules and that this may affect yields[27], more emphasis has been placed on finding strains effective at enhancing yields which are also competitive and form sufficient numbers of nodules[18]. On the other hand, it is well established that plants regulate resource allocation to the shoot and different parts of the root in response to nutrient availability[28–34]. Carbon and nitrogen are the main constituents of plant tissue and their uptake and use are tightly regulated to maximise plant growth[32,35–42]. For example, more resources are allocated to root growth when nitrogen is limiting[28,31]. From a plant perspective, sanctioning is just an extension of this regulation of carbon and nitrogen uptake. Differential resource allocation to nodules based on how much nitrogen they provide and/or how much carbon they cost can only be selected for in plants if it enhances plant fitness, irrespective of the effects on rhizobia. Therefore, understanding how plants sanction and allocate resources to different strains may help in understanding which rhizobial strains positively affect yields and under which conditions. For example, soil nitrogen is known to affect resource allocation to nodules[8,43–46] and if different

rhizobial strains are affected in different ways, this may inform whether effective inoculant strains should be selected under low or high soil nitrogen conditions.

Here, we use mathematical modelling to ask how plants should allocate resources to roots and nodules in order to maximise plant growth. In particular, we ask (*i*) how resources should be allocated among multiple rhizobial strains, (*ii*) whether resources should be allocated solely on the basis of nitrogen provided, or whether the plant should take account of the carbon costs per gram nitrogen fixed and (*iii*) how soil nitrogen affects resource allocation to different rhizobial strains. We use mathematical modelling because even though the questions asked are simple, the outcomes are counterintuitive. A mathematical model can be used to make quantitative predictions about the nature of optimal allocation, and thus generate hypotheses, which can be tested using experiments. In contrast to previous models of the symbiosis[11,12], we explicitly model plant growth and growth of different tissues, so that predictions are in the form of observable quantities, such as nodule mass.

4.4 Methods

4.4.1 Model description

4.4.1.1 Basic model: constant soil nitrogen and a single rhizobial strain

The model describes the growth of a single plant both above and belowground. The aboveground tissue fixes carbon, while the belowground tissue is allocated to two pools: (1) *roots*, which take up nitrogen directly from the soil; and (2) *nodules*, which house rhizobia that can fix atmospheric nitrogen. New plant tissue of any kind requires a fixed ratio of carbon to nitrogen, so that plant growth is either carbon- or nitrogen-limited[47]. The plant obtains carbon via photosynthesis, and nitrogen via roots and/or rhizobial strains. All tissues respire carbon for maintenance, although the model allows this to occur at different rates. For example, nodules may have higher carbon demands than roots as nitrogen fixation is more energy intensive than direct uptake from the soil (estimated averages of 6 g C per gram N[5,46] versus 2 g C per g N[9,48,49] respectively). We use the model to determine the best plant growth strategy by assuming: (*i*) that maximising the instantaneous growth rate will lead to maximum total final plant size; and (*ii*) maximum total final plant size will be selected because larger plants have higher fitness (from an evolutionary perspective) and higher yields (from an agricultural perspective). For simplicity, flowering is not modelled and nodules are not modelled individually. Each rhizobial strain is represented by a single (total) nodule mass.

Mathematically, the model is based on a previous model considering plant growth and allocation in the absence of rhizobia[47]. The carbon-limited growth rate (G_C) at a given point in time depends on the rate of photosynthesis (γ) and the current biomass of aboveground tissue (M_{leaf}) minus the carbon demands of existing shoot (d_{leaf}), root (d_{root}) and nodules (d_{nod}) (*Eq. 1a*). Thus, carbon demands of existing structures are met before resources can be allocated to new growth and the model does not allow for

senescence and redistribution of resources. The nitrogen-limited growth rate (G_N) is given by the sum of root nitrogen uptake and nitrogen fixation by nodules. Root nitrogen uptake depends on current root biomass (M_{root}), soil nitrogen concentration (c_N), and root uptake rate (θ), while nitrogen fixation is given by nodule mass (M_{nod}) and the nitrogen fixation rate (m). The terms in the nitrogen-limited growth rate are multiplied by the tissue C:N ratio (ρ) so that both growth rates are expressed as mg carbon day^{-1} (*Eq. 1b*):

$$G_C(t) = \gamma M_{leaf}(t) - d_{leaf}M_{leaf}(t) - d_{root}M_{root}(t) - d_{nod}M_{nod}(t) \quad (1a)$$

$$G_N(t) = \rho\theta c_N(t)M_{root}(t) + \rho m M_{nod}(t) \quad (1b)$$

The actual growth rate achieved is the minimum of the carbon-limited and nitrogen-limited growth rates (*Eq. 1c*):

$$G(t) = \min(G_C(t), G_N(t)) \quad (1c)$$

Growth is then distributed over shoot, root and nodules (*Eq. 2*) (**Figure 4-1**). A fraction (f) of the total growth is allocated to aboveground biomass (*Eq. 2a*), and the remainder ($1 - f$) to belowground biomass. The growth belowground is split between roots and nodules by allocating a fraction (a) to nodules (*Eq. 2c*) and the remainder ($1 - a$) to roots (*Eq. 2b*).

$$\frac{dM_{leaf}}{dt} = f(t)G(t) \quad (2a)$$

$$\frac{dM_{root}}{dt} = \big(1 - f(t)\big)\,(1 - a(t))\,G(t) \quad (2b)$$

$$\frac{dM_{nod}}{dt} = \big(1 - f(t)\big)\,a(t)\,G(t) \quad (2c)$$

The values of the allocation parameters (f and a) determine how resources are allocated to shoots, roots and nodules. Values of allocation parameters leading to maximum

instantaneous growth rates could be analytically derived and were numerically confirmed.

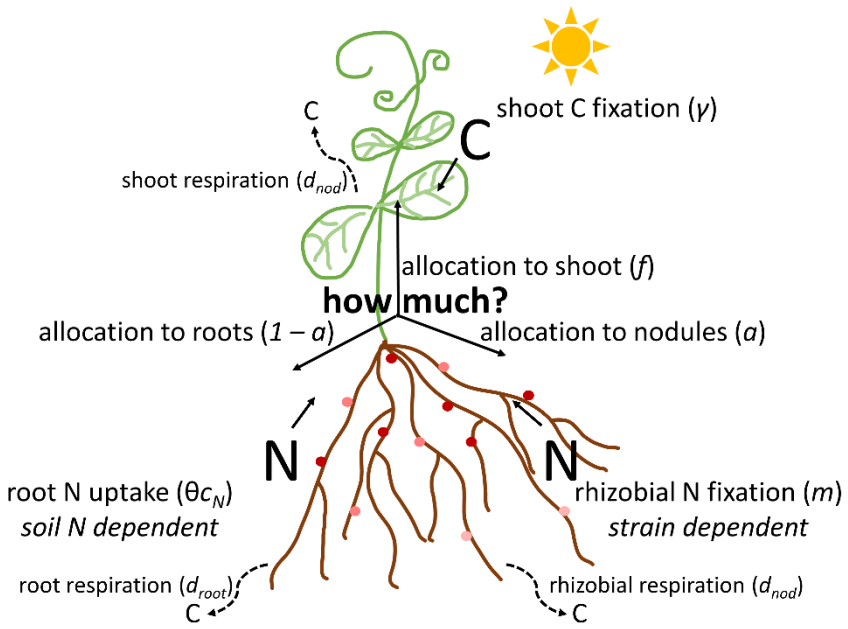


Figure 4-1. Schematic representation of the mathematical model. The model assumes that plant growth is either carbon- or nitrogen-limited and is used to ask how resources should be allocated to shoot, root and nodules for optimal growth rates. For simplicity, nodules are not modelled individually and each rhizobial strain is represented by a single (total) nodule mass.

4.4.1.2 Extension with variable soil nitrogen concentration

The pool of soil nitrogen is unlikely to be a fixed quantity. To test whether allocation to rhizobial strains changes when soil nitrogen supply rates change over time (which is likely in reality), the model was extended with the option of a dynamic ($c_N(t)$) rather than a fixed (c_N) soil nitrogen concentration. Soil nitrogen concentrations were made

dynamic by replacing the parameter for the soil nitrogen concentration with an equation describing the change in concentration over time as a function of fertilisation (or remineralisation) (φ) and root nitrogen uptake, taking into account the volume of soil available to the root (V) (*Eq. 3*).

$$\frac{dc_N}{dt} = \varphi(t) - \frac{\theta c_N(t) M_{root}(t)}{V} \qquad (3)$$

4.4.1.3 Extension with multiple rhizobial strains

In order to see how a plant should allocate resources between two strains which differ both in their effectiveness (how much nitrogen they fix) and their efficiency (how much carbon they require), we included a second rhizobial strain. This requires adding terms and equations to the model, which can indeed be extended to include any number of strains using the same methodology. The carbon-limited growth rate is now affected by the carbon demands of the two strains (*Eq. 4a*), while the nitrogen-limited growth rate now depends on nitrogen provision by two rhizobial strains (*Eq. 4b*).

$$G_C(t) = \gamma M_{leaf}(t) - d_{leaf} M_{leaf}(t) - d_{root} M_{root}(t) - d_{nod,A} M_{nod,A}(t) - d_{nod,B} M_{nod,B}(t) \qquad (4a)$$

$$G_N(t) = \rho\theta c_N(t) M_{root}(t) + \rho m M_{nod,A}(t) + \rho m M_{nod,B}(t) \qquad (4b)$$

The actual growth rate achieved remains the smaller of the carbon-limited and nitrogen-limited growth rates (*Eq. 1c*). Growth is again allocated to aboveground and belowground tissue as determined by the fraction invested in shoot growth (f). Growth belowground is allocated to the two different strains (a_A and a_B, with $a_A + a_B \leq 1$), and the remainder to roots ($1 - a_A - a_B$) (*Eq. 5*).

$$\frac{dM_{leaf}}{dt} = f(t) G(t) \qquad (5a)$$

$$\frac{dM_{root}}{dt} = \left(1 - f(t)\right)\left(1 - a_A(t) - a_B(t)\right) G(t) \quad (5b)$$

$$\frac{dM_{nod,A}}{dt} = \left(1 - f(t)\right) a_A(t)\, G(t) \quad (5c)$$

$$\frac{dM_{nod,B}}{dt} = \left(1 - f(t)\right) a_B(t)\, G(t) \quad (5d)$$

4.4.2 Model implementation

All results are based on analytical solutions for optimal values of allocation parameters f (allocation to aboveground biomass) and a (allocation to nodules within belowground biomass $(1 - f)$). Results are therefore not dependent on specific parameter values used in the model implementation. The model was implemented with specific parameter values to check analytical solutions and to generate graphs. Implementation was in R version 3.3.2[50] and all code is available as supplementary material (Appendix C). Parameter values used in this implementation are presented in **Table 4-1** and were chosen using three sources: (1) those used in Guilbaud *et al.*, 2014[47], who chose parameters suitable for an annual plant with a growth period of 30-200 days, (2) experimental conditions used in this work (**chapter 5**), and (3) reasonable estimates sourced from the literature or in relation to already defined parameters. For example, although the model allows for any rate of nodule respiration (d_{nod}), it was taken to be a linear function of nitrogen fixation with an efficiency (e) between 2 and 10 mg carbon per mg nitrogen fixed[2,8,9,51,52].

Table 4-1. Parameter values used for model implementation.

Parameter	Description	Value and/or range	Unit	Reference
γ	C uptake rate (photosynthesis)	0.5	mg C mg leaf^{-1} d^{-1}	[47,53]
ρ	C:N ratio	7.0	dimension-less	[47]
θ	N uptake rate root	2	cm^3 mg^{-1} d^{-1}	[47,54]
c_N	N concentration in soil	0.05 (0 – 0.08)	mg cm^{-3}	[47] and experimental conditions.
m	N fixation rate	0 – 1	mg N mg nodule^{-1} d^{-1}	Up to ten times as much nitrogen per mg tissue as roots.
d_{leaf}	Shoot respiration rate	0.2	mg C mg leaf^{-1} d^{-1}	Similar to root respiration[53].
d_{root}	Root respiration rate	0.2	mg C mg root^{-1} d^{-1}	± 2 mg C per mg N uptake[9,48,49].
d_{nod}	Nodule respiration rate	0 – 10 ($d_{nod} = em$)	mg C mg nodule^{-1} d^{-1}	Up to efficiency (e) of ten mg C per mg N fixation (estimated average ± 6[51,52])

V	Pot volume	500	cm^3	Experimental conditions
φ	Fertilisation rate	0	mg N cm^{-3} d^{-1}	[47]
M_{seed}	Seed mass	200	mg	Experimental conditions
f	Fraction of growth allocated to shoot	0 – 1	dimension-less	Analytically derived / numerically tested
a	Fraction of belowground growth allocated to nodules	0 – 1	dimension-less	Analytically derived / numerically tested

The optimal value of allocation to aboveground biomass (f) was found analytically, following the approach used in Guilbaud *et al*., 2014[47]. By defining the growth rate at time $t + \Delta t$ and assuming that maximum growth is achieved when carbon and nitrogen are equally limiting at all times ($G_C(t) = G_N(t)$ and $G_C(t + \Delta t) = G_N(t + \Delta t)$), it is possible to solve for the optimal value of allocation to aboveground biomass (f) (**Box 4-1**). Numerical simulations were used to confirm that the analytical solution did indeed maximise growth rates, and resulted in a root:shoot ratio of about 0.3 in the absence of rhizobia, which is comparable with values in the literature for soybean[55]. This value for growth allocation to shoots (f) was therefore used in all subsequent analyses.

Box 4-1. Derivation of optimal allocation of new growth to aboveground tissue (*f*)

The derivation of the optimal allocation of growth to the shoot is based on the assumption that optimal growth is achieved when carbon and nitrogen are equally limiting, following the approach of Guilbaud *et al*., 2014[47]:

$$G_C(t) = G_N(t) \quad \textit{(B1.1a)}$$

$$G_C(t + \Delta t) = G_N(t + \Delta t) \quad \textit{(B1.1b)}$$

where growth at time $t + \Delta t$ (in the case of a single rhizobial strain) is defined as:

$$G_C(t + \Delta t) = \gamma M_{leaf}(t + \Delta t) - d_{leaf} M_{leaf}(t + \Delta t) - d_{root} M_{root}(t + \Delta t) - d_{nod} M_{nod}(t + \Delta t) \quad \textit{(B1. 2a)}$$

$$G_N(t + \Delta t) = \rho\theta M_{root}(t + \Delta t) c_N + \rho m M_{nod}(t + \Delta t) \quad \textit{(B1. 2b)}$$

in which

$$M_{leaf}(t + \Delta t) = M_{leaf}(t) + f(t) G(t) \Delta t \quad \textit{(B1. 3a)}$$

$$M_{root}(t + \Delta t) = M_{root}(t) + (1 - f(t))(1 - a(t)) G(t) \Delta t \quad \textit{(B1. 3b)}$$

$$M_{nod}(t + \Delta t) = M_{nod}(t) + (1 - f(t)) a(t) G(t) \Delta t \quad \textit{(B1. 3c)}$$

Substituting equation B1.3 into equation B1.2 into equation B1.1, cancelling out $G_C(t) = G_N(t)$ and dividing by $G(t)\Delta t$ leads to:

$$\gamma f(t) - d_{leaf} f(t) - d_{root}\big(1 - a(t) - f(t) - a(t) f(t)\big) - a(t) d_{nod} + a(t) f(t) d_{nod} = \rho\theta c_N (1 - f(t) - a(t) + a(t) f(t)) + \rho m a(t) - \rho m f(t) a(t) \quad \textit{(B1. 4)}$$

Rearranging results in an equation for optimal allocation to shoot growth:

$$f = \frac{d_{root} + \rho\theta c_N + a(d_{nod} - d_{root} + \rho(m - \theta c_N))}{\gamma - d_{leaf} + d_{root} + \rho\theta c_N + a(d_{nod} - d_{root} + \rho(m - \theta c_N))} \quad \textit{(B1. 5)}$$

Following the same approach for the case of two rhizobial strains leads to:

$$f = \frac{d_{root} + \rho\theta c_N + a_A(d_{nod,A} - d_{root} + \rho m_A - \rho\theta c_N) + a_B(d_{nod,B} - d_{root} + \rho m_B - \rho\theta c_N)}{\gamma - d_{leaf} + a_A(d_{nod,A} - d_{root} + \rho m_A - \rho\theta c_N) + a_B(d_{nod,B} - d_{root} + \rho m_B - \rho\theta c_N)} \quad \textit{(B1. 6)}$$

To run the model, initial values for shoot, root and nodule masses had to be chosen. To use the analytical solution for optimal allocation to shoot (f), carbon and nitrogen must also equally limit growth at time $t = 0$. This was achieved by assuming that there are initially no nodules and that the initial allocation to root and shoot mass immediately following germination are such that carbon- and nitrogen-limited growth rates are equal ($G_C = G_N$) (**Box 4-2**).

Box 4-2. Derivation of initial conditions.

Derivation of the initial shoot and root mass follows the approach of Guilbaud *et al.*, 2014[47] and is based on the assumptions that (*i*) carbon and nitrogen are equally limiting growth at time $t = 0$, (*ii*) initially there are no nodules, and (*iii*) the combined initial root and shoot mass must equal the seed mass. Therefore, at time $t = 0$:

$$G_C = G_N \qquad \textit{(B2. 1)}$$

$$M_{seed} = M^0_{leaf} + M^0_{root} \qquad \textit{(B2. 2)}$$

This means that:

$$\gamma M^0_{leaf} - d_{leaf} M^0_{leaf} - d_{root} M^0_{root} = \rho\theta M^0_{root} c^0_N \qquad \textit{(B2. 3)}$$

Rearranging gives:

$$M^0_{leaf} = \frac{M^0_{root}(\rho\theta c_N + d_{root})}{\gamma - d_{leaf}} \qquad \textit{(B2. 4a)}$$

$$M^0_{root} = \frac{M^0_{leaf}(\gamma - d_{leaf})}{\rho\theta c_N + d_{root}} \qquad \textit{(B2. 4b)}$$

Substituting $M^0_{root} = M_{seed} - M^0_{leaf}$ into equation B2. 4a and $M^0_{leaf} = M_{seed} - M^0_{root}$ into equation B2. 4b and rearranging leads to the initial conditions:

$$M^0_{leaf} = \frac{\rho\theta c_{N_0} + d_{root}}{\rho\theta c_{N_0} + \gamma + d_{root} - d_{leaf}} \cdot M_{seed} \qquad \textit{(B2. 5a)}$$

$$M^0_{root} = \frac{\gamma - d_{leaf}}{\rho\theta c_{N_0} + \gamma + d_{root} - d_{leaf}} \cdot M_{seed} \qquad \textit{(B2. 5b)}$$

4.5 Results

4.5.1 Allocation to a rhizobial strain depends on its effectiveness and efficiency and on plant nitrogen status

The analytically-derived optimal proportion of belowground growth allocated to nodules (a) (**Box 4-3**) depends on how much carbon a strain respires relative to how much nitrogen it provides, and the carbon and nitrogen uptake rates of the plant itself. Given a rate of nitrogen fixation (m), there is a threshold value of rhizobial carbon use (d_{nod}) above which the plant should not invest in the nodule (*Eq. 6*).

Invest in nodules ($a = 1$) if:

$$d_{nod} < \; d_{leaf} - \gamma + m \cdot \frac{d_{root} + \gamma - d_{leaf}}{\theta c_N} \qquad (6)$$

The threshold value above which the plant should cease investing in nodules is higher when the plant is more strongly nitrogen limited. In other words, strongly nitrogen-limited plants should still invest in rhizobial strains with high carbon demands, while plants which are only weakly nitrogen limited should not invest substantial carbon in nitrogen fixation (**Figure 4-2**). When the rate of nitrogen fixation in nodules is equal to the rate of nitrogen uptake by roots ($m = \theta c_N$), the plant should only invest in nodules if they respire less carbon than the roots ($d_{nod} < d_{root}$), which is unlikely. For a given rhizobial strain the model thus predicts that the plant should switch between investing all belowground growth in nodules when the external nitrogen concentration is low, to investing all belowground resources in roots when some critical value of the external nitrogen concentration is reached: a plant with both roots and nodules is always sub-optimal. A plant without roots is viable because the model makes the simplifying assumption that the only role of roots is to take up nitrogen, and ignores other root functions such as anchorage and uptake of water and other nutrients.

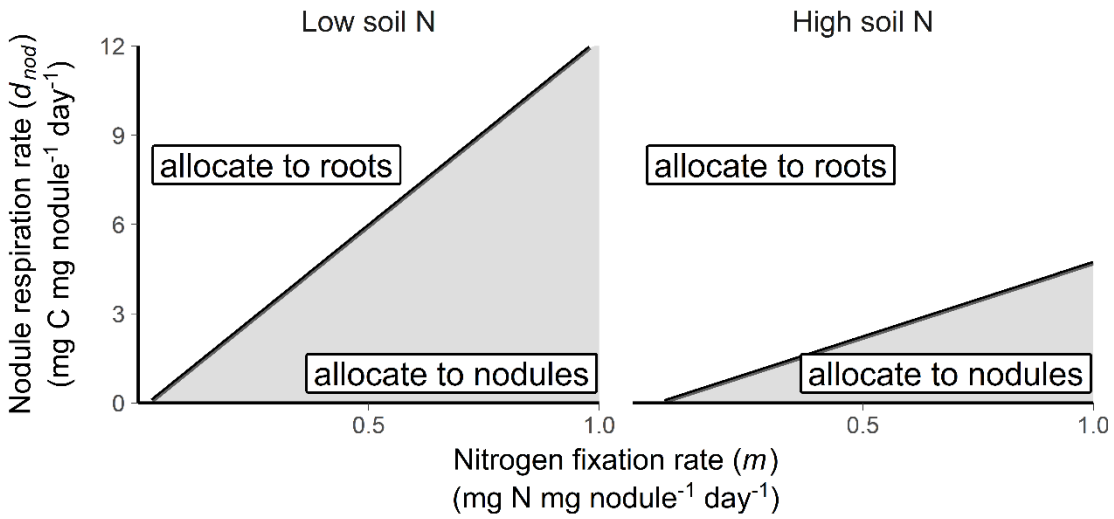


Figure 4-2. Optimal allocation to nodules is all or nothing and depends on rhizobial nitrogen fixation and carbon use, and how nitrogen limited the plant is. Only nitrogen-limited plants should invest in strains with high carbon demands.

From a rhizobial perspective, the threshold soil nitrogen concentration at which the plant would cease to invest in them increases when they are more effective (larger m) or more efficient (smaller e, with $e = d_{nod}/m$), which can be seen by rewriting Equation 6 as a function of the threshold soil nitrogen concentration (**Figure 4-3**):

Invest in nodules ($a = 1$) if:

$$c_N < \frac{m\,(d_{root} + \gamma - d_{leaf})}{\theta(em + \gamma - d_{leaf})} \qquad (7)$$

The relative importance of providing more nitrogen (effectiveness) versus doing so at a lower carbon cost (efficiency) depends on net plant carbon fixation rates: The threshold value is more strongly affected by efficiency when carbon fixation rates are small, and more strongly affected by effectiveness when carbon fixation rates are higher and the plant is more nitrogen-limited.

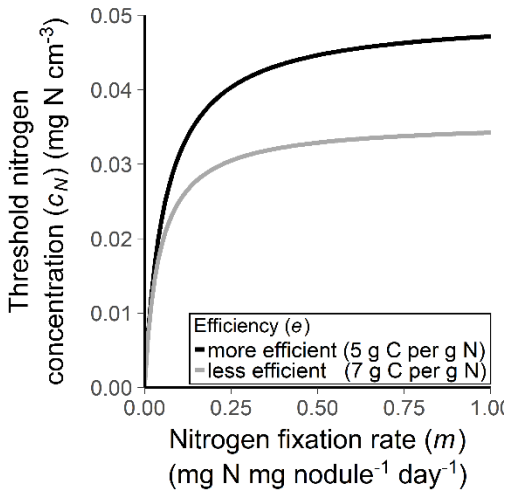


Figure 4-3. The threshold concentration of soil nitrogen above which plants should cease to invest in nodules is higher for more effective (larger m) and more efficient (smaller e) rhizobial strains.

Box 4-3. Derivation of optimal allocation to nodules (a).

Optimal allocation to nodules (the value of parameter a) was found by considering the carbon-limited (G_C) and nitrogen-limited (G_N) growth rates at time $t + \Delta t$. Growth at time $t + \Delta t$ (in the case of a single rhizobial strain) is defined as:

$$G_C(t + \Delta t) = \gamma M_{leaf}(t + \Delta t) - d_{leaf} M_{leaf}(t + \Delta t) - d_{root} M_{root}(t + \Delta t) - d_{nod} M_{nod}(t + \Delta t) \quad \textit{(B3. 1a)}$$

$$G_N(t + \Delta t) = \rho\theta M_{root}(t + \Delta t) c_N + \rho m M_{nod}(t + \Delta t) \quad \textit{(B3. 1b)}$$

in which

$$M_{leaf}(t + \Delta t) = M_{leaf}(t) + f(t) G(t) \Delta t \quad \textit{(B3. 2a)}$$

$$M_{root}(t + \Delta t) = M_{root}(t) + (1 - f(t))(1 - a(t)) G(t) \Delta t \quad \textit{(B3. 2b)}$$

$$M_{nod}(t + \Delta t) = M_{nod}(t) + (1 - f(t)) a(t) G(t) \Delta t \quad \textit{(B3. 2c)}$$

Both the carbon-limited and the nitrogen-limited growth rates are linear functions of how much growth is allocated to the shoot (f), but non-linear functions of allocation to nodules (a) (since f is a non-linear function of a). Therefore, if roots and nodules provide equal amounts of nitrogen and use equal amounts of carbon per gram tissue ($m = \theta c_N$ and $d_{nod} = d_{root}$), both growth rates are independent of allocation to nodules (a) and a linear function of allocation to the shoot (f). The carbon-limited growth rate increases as more resources are allocated to the shoot (larger f), while the nitrogen-limited growth rate decreases (**Figure 4-4a**). The maximum growth rate is achieved when carbon- and nitrogen-limited growth rates are equal, and this yields the optimal value of allocation to the shoot (f) (**Box 4-1**).

If nodules provide more nitrogen than roots without respiring more carbon, increasing allocation to nodules (a) increases the nitrogen-limited growth rate while the carbon-limited growth rate is unaffected. The maximum growth rate is achieved

when carbon and nitrogen are equally limiting, and is highest when all resources are allocated to nodules ($a = 1$) (**Figure 4-4b**). Of course, providing more nitrogen than roots without respiring more carbon is unlikely. Instead, increasing allocation to nodules (a) increases the nitrogen-limited growth rate, but decreases the carbon-limited growth rate, as nodules respire more carbon than roots (**Figure 4-4c**). As long as the increase in the nitrogen-limited growth rate increases the overall growth rate more than the decrease in the carbon-limited growth rate, optimal allocation remains allocating all resources to nodules ($a = 1$). But there comes a point where the carbon-limited growth rate is decreased so much by allocating to nodules, that the maximum growth rate is achieved by ceasing allocation to nodules ($a = 0$) (**Figure 4-4d**). The maximum growth rate always occurs at the extreme ends of allocation to nodules ($a = 0$ or $a = 1$), so that optimal allocation to nodules is all or nothing. We can express the point where plants should switch from allocating to nodules ($a = 1$) to allocating to roots ($a = 0$) in mathematical terms:

Invest in nodules if:

$$G_N(t + \Delta t)_{a=1} > G_N(t + \Delta t)_{a=0} \qquad \textit{(B3. 3)}$$

(Since the optimal value for allocation to the shoot (f) is already known, carbon-limited and nitrogen-limited growth rates are equal ($G_C = G_N$) at all times, so either the carbon-limited or the nitrogen-limited growth rate can be considered for further derivations.)

Substituting equation B3.2 into B3.1 into B3.3 gives:

Invest in nodules if:

$$m(1 - f_{a=1}) > \theta C_N(1 - f_{a=0}) \qquad \textit{(B3. 4)}$$

Substituting for allocation to shoots (f) (see **Box 4-1**) and rearranging yields the optimal allocation to nodules (a):

Invest in nodules ($a = 1$) if:

$$d_{nod} < \; d_{leaf} - \gamma + m \cdot \frac{d_{root} + \gamma - d_{leaf}}{\theta c_N} \qquad (B3.\ 5)$$

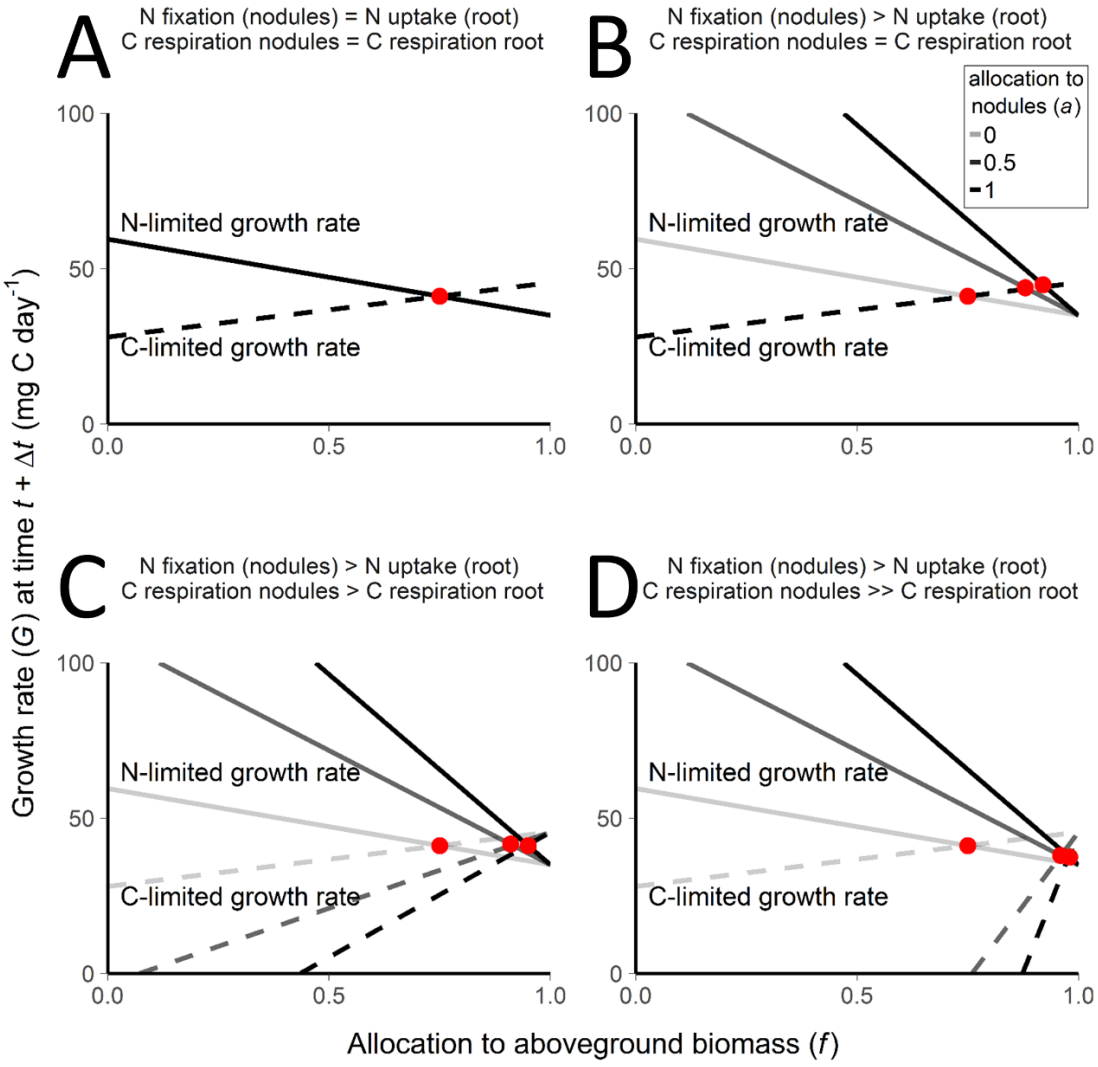


Figure 4-4. Allocation to nodules is always 'all or nothing'. The overall growth rate is defined to be the minimum of carbon- and nitrogen-limited growth rates and is highest where carbon and nitrogen are equally limiting growth (red dots), at the optimal value of

allocation to the shoot (f). (**a**) Nodules and roots provide equal amounts of nitrogen ($m = \theta c_N$) and respire equal amounts of carbon ($d_{nod} = d_{root}$). The carbon-limited growth rate (dashed line) increases as more growth is allocated to the shoot (f), while the nitrogen-limited growth rate (solid line) decreases. Roots and nodules are essentially identical; hence growth rates are independent of allocation to nodules (a). (**b**) Nodules provide more nitrogen than roots ($m > \theta c_N$) and respire the same amount of carbon ($d_{nod} = d_{root}$). For a given allocation to aboveground biomass (f), the nitrogen-limited growth rate is higher if more resources are allocated to nodules (larger a, darker lines), while the carbon-limited growth rate is unaffected. Maximum growth rates for different values of allocation to nodules (a) are indicated by red dots, and the growth rate is highest when all resources are allocated to nodules ($a = 1$). (**c**) Nodules provide more nitrogen than roots ($m > \theta c_N$) but also respire more carbon ($d_{nod} > d_{root}$). Increased allocation to nodules (larger a, darker lines) increases the nitrogen-limited growth rate but decreases the carbon-limited growth rate. The overall maximum growth rate (red dots) is (just) increased more by the increase in the nitrogen-limited growth rate than it is decreased by the decrease in the carbon-limited growth rate. Hence, the growth rate is again maximised when all resources are allocated to nodules ($a = 1$). (**d**) Nodules provide more nitrogen than roots ($m > \theta c_N$) but also respire substantially more carbon ($d_{nod} \gg d_{root}$). The maximum growth rate (red dots) is decreased more by the decrease in the carbon-limited growth rate than it is increased by the increase in the nitrogen-limited growth rate. The growth rate is highest when no resources are allocated to nodules ($a = 0$). Since the highest growth rates are always at either $a = 0$ or at $a = 1$, optimal allocation to nodules is all or nothing.

4.5.2 Allocation to a rhizobial strain depends on other strains present on the same plant

In the case of two or more rhizobial strains, optimal allocation was found to be similarly 'all or nothing', and the plant should allocate to the single – best available – strain (**Box**

4-4). If one of the two strains respires too much carbon for the nitrogen it provides (given the plant nitrogen status), then obviously resources should be allocated to the strain whose carbon demand does not exceed the threshold value (*Eq. 6*). But even if both strains would be a good option and preferred over the root, one of the two will be the better option. Strains providing a lot of nitrogen for little carbon (bottom right corner) are a better option than strains providing less nitrogen and/or using more carbon (strains closer to the threshold line), so that a second threshold line can be drawn for choosing between strain A and strain B (**Figure 4-5**), the slope of which depends on nitrogen provision and carbon use of strain A and net plant carbon fixation (*Eq. 8)*.

Invest in strain B ($a_A = 0$ and $a_B = 1$) if:

$$d_{nod,B} < d_{leaf} - \gamma + m_B \cdot \frac{d_{nod,A} + \gamma - d_{leaf}}{m_A} \qquad (8)$$

Consider for example a strain (A) which is both effective and efficient. A second strain (B) would only be better if it provides even more nitrogen or does so even more efficiently (**Figure 4-5**), while it might still be preferred over the root even if it less effective or efficient than strain A. This means that investment in a rhizobial strain will depend on the other strains present on the same plant. A plant might invest in a given strain, should it be the only one present, and yet cease to invest in the same strain when a better one is available. Since the same argument can be made for comparing a strain C with strain A, the prediction is that optimal resource allocation entails investment in a single – the best available – rhizobial strain.

Which strain is the best available strain depends on conditions, as the relative importance of higher nitrogen fixation versus more efficient fixation depends on the degree of carbon limitation experienced by the plant (*Eq. 8*). When net plant carbon fixation is low ($\gamma - d_{leaf}$ is small), the slope of the threshold line is close to the efficiency of strain A (d_{nod}/m), so that strain B will be chosen over strain A when it is

more efficient than strain A. At higher net carbon fixation rates ($\gamma - d_{leaf}$ is large), the slope of the threshold line increases (and the intercept becomes more negative), so that strain B will be chosen over strain A when it fixes more nitrogen, even when fixation is somewhat less efficient. Because the model assumes a rootless plant is viable, no resources are allocated to the root when the plants invests in rhizobia, so that the soil nitrogen concentration does not affect the choice between strain A and B.

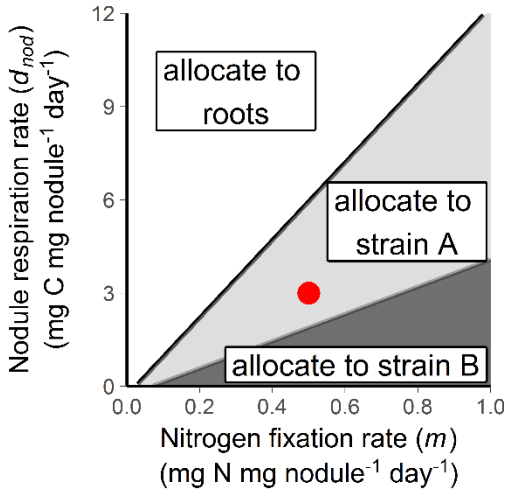


Figure 4-5. Optimal allocation entails investing in a single – best available – rhizobial strain. Given a strain A, the plant should only allocate resources to a more effective or efficient strain B (dark grey area). Therefore, if strain B is preferred over the root, but not over strain A (red dot), it will be invested in only in the absence of strain A.

Box 4-4. Derivation of optimal allocation to two or more rhizobial strains.

A plant will only invest in a rhizobial strain if it provides more nitrogen or does so more efficiently than the root, and thus meets the condition in *Equation 6*. If two different strains meet this requirement, then the plant could choose to invest in one or both (therefore, $a_A + a_B = 1$, so that $a_A = 1 - a_B$). Optimal allocation to two rhizobial strains (the values of parameter a_A and a_B) is found by considering the carbon-limited (G_C) and nitrogen-limited (G_N) growth rates at time $t = \Delta t$. Growth at time $t = \Delta t$ in the case of two rhizobial strains is defined as:

$$G_C(t + \Delta t) = \gamma M_{leaf}(t + \Delta t) - d_{leaf} M_{leaf}(t + \Delta t) - d_{root} M_{root}(t + \Delta t) - d_{nod,A} M_{nod,A}(t + \Delta t) - d_{nod,B} M_{nod,B}(t + \Delta t) \quad \textit{(B4. 1a)}$$

$$G_N(t + \Delta t) = \rho\theta M_{root}(t + \Delta t) c_N + \rho m_A M_{nod,A}(t + \Delta t) + \rho m_B M_{nod,B}(t + \Delta t) \quad \textit{(B4. 1b)}$$

in which

$$M_{leaf}(t + \Delta t) = M_{leaf}(t) + f(t) G(t) \Delta t \quad \textit{(B4. 2a)}$$

$$M_{root}(t + \Delta t) = M_{root}(t) + (1 - f(t))(1 - a_A - a_B)(t)) G(t) \Delta t \quad \textit{(B4. 2b)}$$

$$M_{nod,A}(t + \Delta t) = M_{nod,A}(t) + (1 - f(t)) a_A(t) G(t) \Delta t \quad \textit{(B4. 2c)}$$

$$M_{nod,B}(t + \Delta t) = M_{nod,B}(t) + (1 - f(t)) a_B(t) G(t) \Delta t \quad \textit{(B4. 2d)}$$

For a given strain (A) with nitrogen fixation rate (m_A) and efficiency ($e_A = d_{nod,A}/m_A$), the growth rate at time $t = \Delta t$ is a non-linear function of the nitrogen fixation and respiration rates of strain B, and allocation to strain B (**Figure 4-6**). When strain A and strain B respire equal amounts of carbon per gram nitrogen fixed, overall growth rate is highest when allocating all resources to strain A (if strain A fixes more nitrogen), or all resources to strain B (when strain B fixes more nitrogen). The overall

growth rate is independent of resource allocation when the strains provide equal amounts of nitrogen (strains are indistinguishable) (**Figure 4-6A**). If strain B respires more carbon per gram nitrogen fixed than strain A, then allocation to strain B only results in higher growth rates if strain B provides considerably more nitrogen than strain A (**Figure 4-6B**). If the difference in efficiency of nitrogen fixation becomes too large, growth rates remain highest by allocating to strain A, even if strain A provides less nitrogen (**Figure 4-6C**). Again, growth rates are highest at extreme ends of allocation, so by either allocating all resources to strain A, or to strain B. Analogously to comparing roots with nodules, this yields a criterion to choose between strain A and strain B.

Invest in strain B if:

$$G_N(t + \Delta t)_{a_A=0 \,\&\, a_B=1} > G_N(t + \Delta t)_{a_A=1 \,\&\, a_B=0} \qquad \textit{(B4. 3)}$$

Substituting equation B4.2 into equation B4.1 into equation B4.3 leads to:

Invest in strain B if ($a_A = 0$ and $a_B = 1$):

$$m_B(1 - f_{a_A=0 \,\&\, a_B=1}) > m_A(1 - f_{a_A=1 \,\&\, a_B=0}) \qquad \textit{(B4. 4)}$$

Substituting for allocation to shoots (f) (see **Box 4-1**) and rearranging yields the optimal allocation to nodules strain B:

Invest in strain B if ($a_A = 0$ and $a_B = 1$):

$$d_{nod,B} < d_{leaf} - \gamma + m_B \cdot \frac{d_{nod,A} + \gamma - d_{leaf}}{m_A} \qquad \textit{(B4. 5)}$$

Optimal allocation means that the plant allocates all resources to either strain A or to strain B. If strain B is the better choice, the same argument can be made for comparing strain B with strain C. Therefore, given a soil nitrogen concentration, there will always be a single rhizobial strain to which all resources should be allocated for maximum growth.

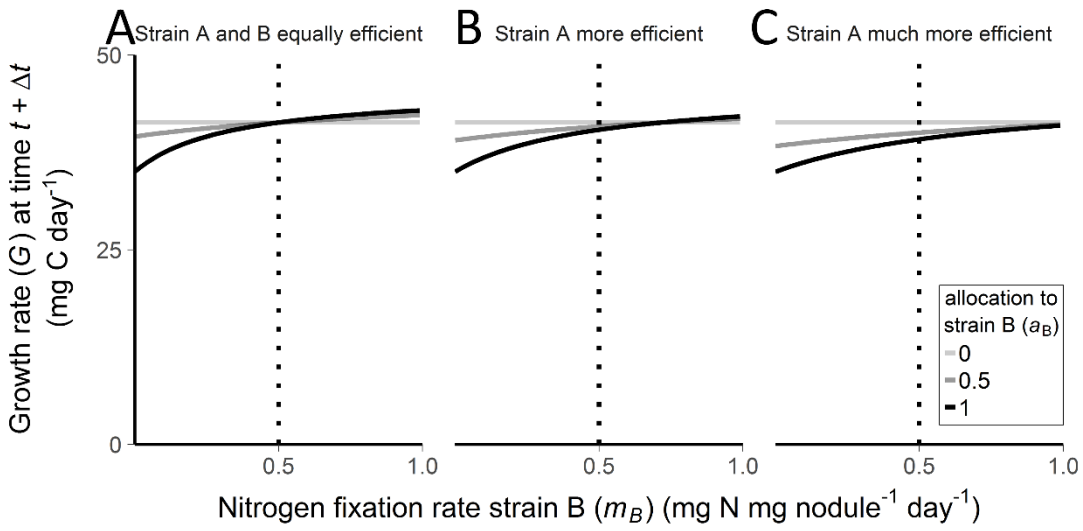


Figure 4-6. Optimal allocation to two rhizobial strains. The dotted vertical line indicates the rate of nitrogen fixation of strain A (m_A). All belowground growth is allocated to nodules ($a_A + a_B = 1$), since a choice between two strains only needs to be made if both strains are preferred over the root. (**a**) If strain A and strain B respire equal amounts of carbon per gram nitrogen fixed ($d_{nod,A} = d_{nod,B}$), investing in strain B leads to higher growth rates when strain B fixes more nitrogen than strain A ($m_B > m_A$). (**b**) If strain B respires more carbon than strain A ($d_{nod,B} > d_{nod,A}$), investing in strain B only leads to higher growth rates when strain B provides considerably more nitrogen than strain A ($m_B \gg m_A$). (**c**) If strain B respires much more carbon than strain A ($d_{nod,B} \gg d_{nod,A}$), allocating all resources to strain A leads to the highest growth rates, even if strain B provides more nitrogen.

4.6 Discussion

We developed a model to establish how plants can maximise their growth rate, and thus fitness and yield, by differentially allocating resources to roots and to nodules differing in how much nitrogen they provide and how much carbon they demand. We found that: (*i*) strongly nitrogen-limited plants should invest in rhizobial strains with high carbon demands, while weakly nitrogen-limited plants should not. Thus, a plant should cease to invest in a rhizobial strain above a threshold value of soil nitrogen concentration. This threshold concentration is higher for both more effective and more efficient strains, and the relative importance of the two depends on plant carbon fixation rates; (*ii*) Optimal allocation entails investing in either roots, or a single – best available – rhizobial strain. Dividing resources among multiple strains is sub-optimal. Therefore, a strain might receive resources when present on its own but not when a better alternative is available to the plant. Which strain is the best strain may be different under different conditions as plants should invest more carbon in obtaining nitrogen when they have higher carbon fixation rates. Model outcomes can be compared to observations.

The model predicts all-or-nothing investment in either the root or nodules, so that a given rhizobial strain only receives investment up to a threshold soil nitrogen concentration, which is lower for less effective or less efficient strains. In reality of course roots are always formed, as they are needed for purposes other than nitrogen uptake. Nevertheless, the effect of soil nitrogen, especially in the form of nitrate, on both the number and size of nodules is well established. At higher soil nitrogen concentrations, generally fewer and smaller nodules are formed[44,45,56]. The effects of nitrate have been reported to be strain dependent[56–58], but it remains unclear whether there is a correlation with how much nitrogen strains provide or how efficiently they do so. Based on the model predictions, it would be expected that more effective and

efficient strains are found in higher soil nitrogen environments. Indeed, others have suggested that the Inverted Repeat Lacking Clade (IRLC) group of legumes evolved increased control over the symbiosis, rendering it more beneficial, because they evolved in fertile, formerly glacial, areas[59].

The model predicts all-or-nothing resource allocation to different strains, so that all resources are invested in the single – best available – strain. Effective strains have been shown to form larger nodules than ineffective strains[13,14,20,27,60,61] (and **chapter 3**[21]), but this observation does not give insight into whether the plant actually invests all resources in the single, best available strain. For example, different nodule sizes could also arise from the plant allocating resources to strains in proportion to nitrogen received, which could be regulated locally. But the all-or-nothing allocation predicted by the model, in which resource allocation depends on other strains present on the same plant, would require integration of information across the root network and could not arise through local regulation alone. There is some evidence that advanced regulatory mechanisms to achieve such integration exist. For example, integration is required to regulate compensatory root growth in nutrient-rich patches when part of the root system is exposed to nutrient poor conditions[62]. But, even if regulatory mechanisms for optimal resource allocation to nodules are feasible, resource allocation may still deviate from our model predictions because of other constraints, such as the number of nodules formed by different strains.

The model prediction of investment in a single strain is not necessarily at odds with the observation that legumes tend to be infected by multiple strains[22–25]. Resource allocation to multiple strains could arise because the best available strain only forms small numbers of nodules, or forms nodules late in the growing season, and cannot meet the nitrogen requirements of the plant. Such constraints are not currently included in the

model. Furthermore, the model might predict investment in multiple strains through time, as the best strain to invest in depends on external conditions. In the model, the choice between strains is only affected by net carbon fixation rates and not by soil nitrogen, since a rootless plant is assumed to be viable and when resources are allocated to rhizobia no resources are allocated to the root. In reality, soil nitrogen is likely to affect the strength of carbon limitation and thus to affect the choice between strains. It remains to be seen how likely carbon limitation is under field conditions and thus whether the efficiency of nitrogen fixation plays a role in choosing between strains. It seems plausible that carbon fixation could be limited by availability of nutrients other than nitrogen and that this in turn may affect resource allocation to nodules. Phosphorus deficiency for example has been shown to affect carbon allocation to nodules because more carbon was allocated to the root[63].

The model could be further developed to address some of the outstanding questions, for example by including other nutrients (thereby also preventing rootless plants) and by more explicitly modelling the nodulation process, so that uneven nodule numbers and sequential inoculation could be taken into account. But the model has already provided hypotheses, which can be tested in experiments, and guide further model development. In particular, the model predicts that: (*i*) a rhizobial strain with reduced nitrogen fixation or increased carbon demands will form large nodules when singly inoculated on a nitrogen-limited plant, but small nodules when co-inoculated with a more effective or efficient strain, and; (*ii*) nodules cease to receive resources above a certain threshold level of soil nitrogen, and the concentration at which this happens is lower for strains providing less nitrogen. These hypotheses can be tested in growth room experiments using strains that differ in how much nitrogen they provide and how much

carbon they demand, but are otherwise similar and thus have equal chances of nodulation. **Chapter 5** provides tests of these hypotheses.

4.7 References

1. Lebauer, D. S. & Treseder, K. K. Nitrogen limitation of net primary productivity in terrestrial ecosystems is globally distritubed. *Ecology* **89,** 371–379 (2008).
2. Kaschuk, G., Kuyper, T. W., Leffelaar, P. A., Hungria, M. & Giller, K. E. Are the rates of photosynthesis stimulated by the carbon sink strength of rhizobial and arbuscular mycorrhizal symbioses? *Soil Biol. Biochem.* **41,** 1233–1244 (2009).
3. Bergersen, F. J. Some Australian studies relating to the long-term effects of the inoculation of legume seeds. *Plant Soil* **32,** 727–736 (1970).
4. Burdon, J. & Gibson, A. Variation in the effectiveness of symbiotic associations between native rhizobia and temperate Australian *Acacia*: within-species interactions. *J. Appl. Ecol.* **36,** 398–408 (1999).
5. Fening, J. O. & Danso, S. K. A. Variation in symbiotic effectiveness of cowpea bradyrhizobia indigenous to Ghanaian soils. *Appl. Soil Ecol.* **21,** 23–29 (2002).
6. Sprent, J. Mutual sanctions. *Nature* **422,** 672–674 (2003).
7. Thrall, P. H., Laine, A.-L., Broadhurst, L. M., Bagnall, D. J. & Brockwell, J. Symbiotic effectiveness of rhizobial mutualists varies in interactions with native Australian legume genera. *PLoS One* **6,** e23545 (2011).
8. Voisin, A. S., Salon, C., Jeudy, C. & Warembourg, F. R. Symbiotic N_2 fixation activity in relation to C economy of *Pisum sativum* L. as a function of plant phenology. *J. Exp. Bot.* **54,** 2733–44 (2003).
9. Kaschuk, G., Hungria, M., Leffelaar, P. a, Giller, K. E. & Kuyper, T. W. Differences in photosynthetic behaviour and leaf senescence of soybean (*Glycine max* [L.] Merrill) dependent on N_2 fixation or nitrate supply. *Plant Biol.* **12,** 60–69 (2010).
10. Denison, R. F. Legume sanctions and the evolution of symbiotic cooperation by rhizobia. *Am. Nat.* **156,** 567–576 (2000).
11. West, S. A., Kiers, E. T., Pen, I. & Denison, R. F. Sanctions and mutualism stability: when should less beneficial mutualists be tolerated? *J. Evol. Biol.* **15,** 830–837 (2002).

12. West, S. A., Kiers, E. T., Simms, E. L. & Denison, R. F. Sanctions and mutualism stability: why do rhizobia fix nitrogen? *Proc. R. Soc. B.* **269,** 685–94 (2002).

13. Kiers, E. T., Rousseau, R. A., West, S. A. & Denison, R. F. Host sanctions and the legume-rhizobium mutualism. *Nature* **425,** 78–81 (2003).

14. Oono, R., Anderson, C. G. & Denison, R. F. Failure to fix nitrogen by non-reproductive symbiotic rhizobia triggers host sanctions that reduce fitness of their reproductive clonemates. *Proc. R. Soc. B.* **278,** 2698–703 (2011).

15. Howieson, J. G., Malden, J., Yates, R. J. & O'Hara, G. W. Techniques for the selection and development of elite inoculant strains of Rhizobium leguminosarum in southern Australia. *Symbiosis (Rehovot)* **28,** 33–48 (2000).

16. Slattery, J. & Pearce, D. Development of elite inoculant *Rhizobium* strains in Southeastern Australia. *ACIAR Proc. 109e* (2002).

17. Yates, R. J. *et al.* Host-strain mediated selection for an effective nitrogen-fixing symbiosis between *Trifolium* spp. and *Rhizobium leguminosarum* biovar *trifolii*. *Soil Biol. Biochem.* **40,** 822–833 (2008).

18. Yates, R. J., Howieson, J. G., Reeve, W. G. & O'Hara, G. W. A re-appraisal of the biology and terminology describing rhizobial strain success in nodule occupancy of legumes in agriculture. *Plant Soil* **348,** 255–267 (2011).

19. Thilakarathna, M. S. & Raizada, M. N. A meta-analysis of the effectiveness of diverse rhizobia inoculants on soybean traits under field conditions. *Soil Biol. Biochem.* **105,** 177–196 (2017).

20. Kiers, T. E., Rousseau, R. A. & Denison, R. F. Measured sanctions: legume hosts detect quantitative variation in rhizobium cooperation and punish accordingly. *Evol. Ecol. Res.* **8,** 1077–1086 (2006).

21. Westhoek, A. *et al.* Policing the legume-Rhizobium symbiosis: a critical test of partner choice. *Sci. Rep.* 1–10 (2017).

22. Dowling, D. N. & Broughton, W. J. Competition for nodulation of legumes. *Annu. Rev. Microbiol.* **40,** 131–157 (1986).

23. Souza, V. *et al.* Ethnomicrobiology: do agricultural practices modify the population structure of the nitrogen fixing bacteria *Rhizobium etli* biovar *phaseoli*? *J. Ethnobiol.* **17,** 249–266 (1997).

24. Hagen, M. J. & Hamrick, J. L. A hierarchical analysis of population genetic structure in *Rhizobium leguminosarum* bv. *trifolii*. *Mol. Ecol.* **5,** 177–186 (1996).

25. Silva, C., Eguiarte, L. E. & Souza, V. Reticulated and epidemic population genetic structure of *Rhizobium etli* biovar *phaseoli* in a traditionally managed locality in Mexico. *Mol. Ecol.* **8,** 277–287 (1999).

26. Kaschuk, G. *et al.* Responses of legumes to rhizobia and arbuscular mycorrhizal fungi: a meta-analysis of potential photosynthate limitation of symbioses. *Soil Biol. Biochem.* **42,** 125–127 (2010).

27. Singleton, W. & Stockinger, K. R. Compensation against ineffective nodulation in soybean. *Crop Sci.* **23,** 69–72 (1983).

28. Zhang, H., Jennings, A., Barlow, P. W. & Forde, B. G. Dual pathways for regulation of root branching by nitrate. *Plant Biol.* **96,** 6529–6534 (1999).

29. Krouk, G. *et al.* A framework integrating plant growth with hormones and nutrients. *Trends Plant Sci.* **16,** 178–182 (2011).

30. Alvarez, J. M., Vidal, E. A. & Gutiérrez, R. A. Integration of local and systemic signaling pathways for plant N responses. *Curr. Opin. Plant Biol.* **15,** 185–191 (2012).

31. Andrews, M., Raven, J. A. & Lea, P. J. Do plants need nitrate? The mechanisms by which nitrogen form affects plants. *Ann. Appl. Biol.* **163,** 174–199 (2013).

32. Ruffel, S., Gojon, A. & Lejay, L. Signal interactions in the regulation of root nitrate uptake. *J. Exp. Bot.* **65,** 5509–5517 (2014).

33. Forde, B. G. Nitrogen signalling pathways shaping root system architecture: an update. *Curr. Opin. Plant Biol.* **21,** 30–36 (2014).

34. Bisseling, T. & Scheres, B. Nutrient computation for root architecture. *Science* **346,** 300–301 (2014).

35. Scheible, W. R. Nitrate acts as a signal to induce organic acid metabolism and repress

starch metabolism in tobacco. *Plant Cell Online* **9,** 783–798 (1997).

36. Hsieh, M. H., Lam, H. M., van de Loo, F. J. & Coruzzi, G. A PII-like protein in *Arabidopsis*: putative role in nitrogen sensing. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 13965–13970 (1998).

37. Coruzzi, G. M. & Zhou, L. Carbon and nitrogen sensing and signaling in plants: emerging 'matrix effects'. *Curr. Opin. Plant Biol.* **4,** 247–253 (2001).

38. Stitt, M. *et al.* Steps towards an integrated view of nitrogen metabolism. *J. Exp. Bot.* **53,** 959–970 (2002).

39. Walch-Liu, P., Filleur, S., Gan, Y. & Forde, B. G. Signaling mechanisms integrating root and shoot responses to changes in the nitrogen supply. *Photosynth. Res.* **83,** 239–250 (2005).

40. Lu, Y. *et al.* ABI1 regulates carbon/nitrogen-nutrient signal transduction independent of ABA biosynthesis and canonical ABA signalling pathways in *Arabidopsis*. *J. Exp. Bot.* **66,** 2763–2771 (2015).

41. Kaur, B., Kaur, G. & Asthir, B. Biochemical aspects of nitrogen use efficiency: an overview. *J. Plant Nutr.* **40,** 506–523 (2017).

42. Otori, K. *et al.* Enhanced photosynthetic capacity increases nitrogen metabolism through the coordinated regulation of carbon and nitrogen assimilation in *Arabidopsis thaliana*. *J. Plant Res.* **0,** 0 (2017).

43. Wery, J., Turc, O. & Salsac, L. Relationship between growth, nitrogen fixation and assimilation in a legume (*Medicago sativa* L.). *Plant Soil* **96,** 17–29 (1986).

44. Streeter, J. & Wong, P. P. Inhibition of legume nodule formation and N_2 fixation by nitrate. *CRC. Crit. Rev. Plant Sci.* **7,** 1–23 (1988).

45. Parsons, R., Stanforth, A., Raven, J. a. & Sprent, J. I. Nodule growth and activity may be regulated by a feedback mechanism involving phloem nitrogen. *Plant, Cell Environ.* **16,** 125–136 (1993).

46. Kinkema, M., Scott, P. T. & Gresshoff, P. M. Legume nodulation: successful symbiosis through short- and long-distance signalling. *Funct. Plant Biol.* **33,** 707–721 (2006).

47. Guilbaud, C. S. E., Dalchau, N., Purves, D. W. & Turnbull, L. A. Is 'peak N' key to understanding the timing of flowering in annual plants? *New Phytol.* **205,** 918–927 (2014).

48. Cooper, H. D. & Clarkson, D. T. Cycling of amino-nitrogen and other nutrients between shoots and roots in cereals - a possible mechanism integrating shoot and root in the regulation of nutrient uptake. *J. Exp. Bot.* **40,** 753–762 (1989).

49. Shi, M., Fisher, J. B., Brzostek, E. R. & Phillips, R. P. Carbon cost of plant nitrogen acquisition: global carbon cycle impact from an improved plant nitrogen cycle in the community land model. *Glob. Chang. Biol.* 1299–1314 (2016). doi:10.1111/gcb.13131

50. R Core Team. A language and environment for statistical computing. (2016).

51. Vance, C. P. & Heichel, G. H. Carbon in N_2 fixation: limitation or exquisite adaptation. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **42,** 373–390 (1991).

52. Udvardi, M. & Poole, P. S. Transport and metabolism in legume-rhizobia symbioses. *Annu. Rev. Plant Biol.* **64,** 781–805 (2013).

53. Poorter, H., Remkes, C. & Lambers, H. Carbon and nitrogen economy of 24 wild species differing in relative growth rate. *Plant Physiol.* **94,** 621–627 (1990).

54. Tolley-Henry, L., Raper, C. D. & Granato, T. C. Cyclic variations in nitrogen uptake rate of soybean plants: effects of external nitrate concentration. *J. Exp. Bot.* **39,** 613–622 (1988).

55. Rogers, H. H., Peterson, C. M., Mccrimmon, J. N. & Cure, J. D. Response of plant roots to elevated atmospheric carbon dioxide. *Plant, Cell Environ.* **15,** 749–752 (1992).

56. Caballero Mellado, J. & Martinez Romero, E. Soil fertilization limits the genetic diversity of *Rhizobium* in bean nodules. *Svmbiosis* **26,** 111–121 (1999).

57. McNeil, D. L. Variations in ability of *Rhizobium japonicum* strains to nodulate soybeans and maintain fixation in the presence of nitrate. *Appl. Environ. Microbiol.* **44,** 647–652 (1982).

58. Boivin, C. *et al.* The *Sesbania* root symbionts *Sinorhizobium saheli* and *S. teranga* bv. sesbaniae can form stem nodules on *Sesbania rostrata*, although they are less adapted to

stem nodulation than *Azorhizobium caulinodans*. *Appl. Environ. Microbiol.* **63,** 1040–1047 (1997).

59. Sprent, J. I., Ardley, J. K. & James, E. K. From North to South: A latitudinal look at legume nodulation processes. *South African J. Bot.* **89,** 31–41 (2013).

60. Simms, E. L. *et al.* An empirical test of partner choice mechanisms in a wild legume-rhizobium interaction. *Proc. Biol. Sci.* **273,** 77–81 (2006).

61. Gubry-Rangin, C., Garcia, M. & Béna, G. Partner choice in *Medicago truncatula-Sinorhizobium* symbiosis. *Proc. Biol. Sci.* **277,** 1947–51 (2010).

62. Ruffel, S. & Krouk, G. Nitrogen economics of root foraging: transitive closure of the nitrate–cytokinin relay and distinct systemic signaling for N supply vs. demand. *Proc. Natl. Acad. Sci.* **108,** 18524–18529 (2011).

63. Kleinert, A., Venter, M., Kossmann, J. & Valentine, A. The reallocation of carbon in P deficient lupins affects biological nitrogen fixation. *J. Plant Physiol.* **171,** 1619–1624 (2014).

5 Resource allocation in the legume-rhizobia symbiosis

'In all things of nature there is something of the marvellous.'

Aristotle, 384−322 BC

5.1 Overview

In this chapter results are presented from experiments conducted to test hypotheses generated by modelling the legume-rhizobia symbioses (**chapter 4**). The results are to be published together with the results of **chapter 4**. Results are discussed briefly, but their implications for future research are discussed in **chapter 6**. Supplementary information can be found in Appendix D and contains: (*i*) supplementary results; (*ii*) information on data; and (*iii*) R code used to analyse the data. The data that support the findings of this study are available in the Oxford Research Archive (ORA-Data) repository, DOI: 10.5287/bodleian:BDVypApax.

5.2 Abstract

Plants are known to balance carbon and nitrogen uptake by adjusting resource allocation to the shoot and root in response to local nitrogen availability. In the case of legumes, which form symbioses with nitrogen-fixing rhizobia, this balancing act also includes the nodules in which rhizobia reside. Rhizobial strains differ in (1) how much nitrogen they fix (effectiveness), and (2) the carbon required to do so (efficiency). As both carbon and

nitrogen are valuable resources to the plant, this variation among strains may affect plant growth. We used a plant growth model to generate predictions about optimal resource allocation to different rhizobial strains which we test here.

We used derivatives of rhizobial strain *Rhizobium leguminosarum* bv. *viciae* Rlv3841 which differed in either the rate of nitrogen fixation, or the carbon cost (per gram nitrogen fixed), but were otherwise similar. We inoculated plants with different combinations of those strains to test whether resource allocation to less effective or efficient strains (that provide less nitrogen or require more carbon) depends on whether or not a better, alternative strain is available. We also assessed whether the effect of soil nitrogen on resource allocation to nodules depends on how much nitrogen strains provide.

We found that plants preferentially allocated resources to the most effective strain. An intermediate-fixing strain received fewer resources when co-inoculated with a more effective strain than when co-inoculated with a non-fixing strain, in line with model predictions. In contrast, plants did not preferentially allocate resources to more efficient strains. Resource allocation to a strain with increased carbon costs per gram nitrogen fixed did not depend on whether or not a more efficient strain was available. Increasing soil nitrogen concentrations lead to reduced investment in nodules. For effective strains this was mostly due to a reduction in nodule size; for less effective strains, this was mostly due to a reduction in nodule number. Our results indicate that pea plants have evolved sophisticated mechanisms to compare nitrogen provision from different nodules, integrate information, and adjust resource allocation accordingly.

5.3 Introduction

Carbon and nitrogen are the main constituents of plant tissue. To maximise growth rates, plants have evolved mechanisms to tightly control uptake of carbon and nitrogen, adjusting resource allocation to the shoot and different parts of the root in response to local nutrient availability[1–5]. Legumes can also obtain nitrogen through a symbiosis with nitrogen-fixing rhizobia, in exchange for carbon[6]. Rhizobial strains vary in how much nitrogen they provide[7–11] (effectiveness) and how much carbon they require per gram nitrogen fixed[12–14] (efficiency), both of which may affect plant growth and therefore fitness and yield. Plants may withold resources from strains that provide little nitrogen, which is called sanctioning if it also reduces rhizobial fitness[15–19]. Persuasive evidence for sanctioning has been provided in the case of fixing and non-fixing strains[19–21]. However, it is less clear: (*i*) how resources would be allocated among multiple rhizobial strains with different effectiveness; (*ii*) whether the efficiency of rhizobial strains also affect plant resource allocation; and (*iii*) how soil nitrogen affects resource allocation to strains of different effectiveness. Here, we conducted experiments to test predictions about resource allocation from a theoretical model (**chapter 4**).

Through mathematical modelling, predictions were made about the nature of optimal resource allocation (**chapter 4**). The model predicts that plants should invest all resources in the single, best available strain, rather than provide resources to multiple strains (for example, in proportion to the nitrogen they provide). The best available strain is more effective or more efficient, and the relative importance of these two characteristics depends on the extent of carbon-limitation in the plant. The model also predicts that plants cease to invest in rhizobial strains above a threshold soil nitrogen concentration, which is higher for more effective strains. However, allocation of resources may deviate from model predictions, for example, if plants lack the required

regulatory mechanisms or if not all aspects are under plant control. Establishing how plants actually allocate resources therefore requires experiments, which have so far not been easy to design so that hypotheses about resource allocation can be properly tested.

Split-root experiments have been instrumental in studying resource allocation to roots in response to local nutrient availability[5], and have also been used with nodulated roots[22–25]. However, it is not technically feasible to create more than two or maybe three different compartments, making it difficult to study allocation to the multiple rhizobial strains that infect a typical legumes individual[26–29]. It is also problematic to discriminate between resource allocation to a part of the root which receives more nitrogen and resource allocation to specific nodules which provide more nitrogen, as nodules containing different strains may sit very close to each other on normal (non-split) roots (see **chapter 3**[21]).

Another difficulty is the natural variation in both how much nitrogen strains provide[7–11] and how much carbon they use[12–14]. Investment to rhizobial strains is predicted to depend on both (**chapter 4**), but to understand the effect of these two different aspects of rhizobial behaviour on resource allocation requires strains that vary in one aspect but not the other. This is near impossible to achieve with naturally occurring strains. To overcome this problem, previous studies have used argon gas to manipulate the amount of atmospheric nitrogen available to rhizobia[19,20,24,25,30], and therefore their rate of nitrogen fixation. However, this requires the use of either split-root experiments[19,20,24,25], or encasing nodules in a chamber[20,30], which can only be done at a late stage of nodule development, thereby missing the earlier stages of resource allocation.

Here, we use a set of mutant strains derived from *Rhizobium leguminosarum* bv. viciae (Rlv) strain 3841[31] to experimentally manipulate either effectiveness or efficiency

in strains that are otherwise similar. Insertion of spectinomycin resistance genes led to non-fixing mutant strains (**chapter 3**[21]), and also a strain fixing nitrogen at intermediate rates[32]. The *gusA* and *celB* marker gene system was used to discriminate between strains post-harvest (**chapter 3**[21]). To manipulate carbon use, we made use of the *lux* reporter gene system which is widely used to study gene expression. The set of genes causes bacteria to emit light, and is normally placed under the control of a specific promoter, so that the amount of light emitted reports on the levels of gene expression[33]. We used the system to increase the amount of carbon used per nitrogen fixed, by putting *lux* genes under control of the *nifH* gene promoter, essential for nitrogen fixation. Each photon of light produced requires 20 molecules of ATP, so a considerable energetic burden is placed on strains expressing the *lux* genes[34,35]. Thus, we increased carbon usage during nitrogen fixation without interfering with other aspects of rhizobial function, for example, in root colonisation success.

The rhizobial strains varying in how much nitrogen they provide and how much carbon they require were used to test model predictions about plant resource allocation. Specifically, we tested whether: (*i*) in nitrogen-limited environments plants will allocate resources to strains that provide less nitrogen or demand more carbon, if such a strain is the only option, but will sanction the same strain if a better alternative is present, and (*ii*) resource allocation to a given strain drops rapidly above a critical level of soil nitrogen, and the concentration at which this happens is lower for strains providing little nitrogen.

5.4 Results

Results are presented based on measuring nodule size (area) of the ten most-developed (largest) nodules, so that variation in size due to developmental stage was accounted for, but similar results were obtained based on randomly selecting ten nodules (Appendix D, supplementary material**)**.

5.4.1 Characterisation of intermediate-fixing strain LMB827

To study resource allocation to strains that differ in nitrogen provision, we made use of non-fixing (OPS0365 and OPS0366 (Fix^{-}), **chapter 3**[21]) and intermediate-fixing (LMB827 (Fix^{Int})[32]) derivatives of strain Rlv3841 (Fix^{+})[31]. The intermediate-fixing strain is not fully characterised genetically, but contains a duplication of *fixC*, the first copy of which is disrupted with the omega spectinomycin cassette (ΩSpc). The strain has the following genetic structure; *fixABCfixCΩSpc*, followed by *fixCX*, *nifA*, *nifB*. As ΩSpc is highly polar, it prevents auto-upregulation of *nifA* from the *nifA* promoter in front of *fixA*, restricting its expression to the basal constitutive promoter in *fixC*[32]. It therefore fixes nitrogen at reduced rates as compared to Rlv3841 (2.5 ± 0.2 versus 3.6 ± 0.2 μmol ethylene $plant^{-1}$ $hour^{-1}$ respectively at 28 days post inoculation)[32]. Nodules of the intermediate-fixing strain are pink, but slightly paler than fixing nodules (**Figure 5-1**). In single inoculations (single root or split-root with the same strain on each half-root), the intermediate-fixing strain formed intermediate-sized nodules. Intermediate-fixing nodules were 2.66 mm^2 (95% CI [2.21, 3.11]), which is 0.80 mm^2 (95% CI [0.17, 1.43]) larger than non-fixing nodules and 0.49 mm^2 (95% CI [0.08, 0.90]) smaller than fixing nodules (**Figure 5-2**).

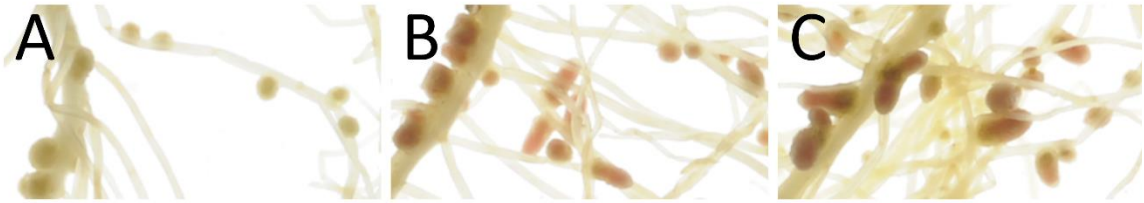


Figure 5-1. Representative nodules of three strains differing in effectiveness. (**a**) Non-fixing strain OPS0365. (**b**) Intermediate-fixing strain LMB827. (**c**) Fixing strain Rlv3841 *gusA*. Each picture represents a 2x1 cm root section of pea plants grown for 28 days without external nitrogen.

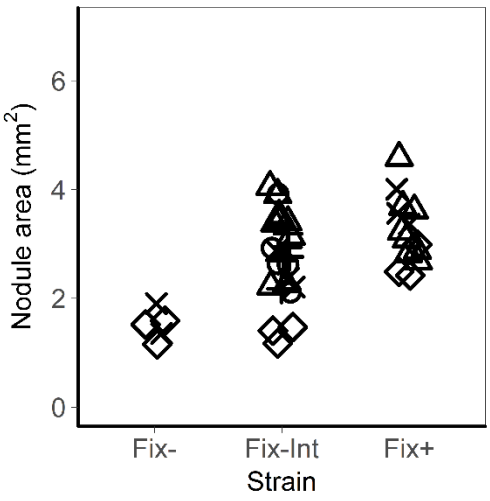


Figure 5-2. Nodule sizes of strains differing in effectiveness in single inoculations (Fix^{-}; OPS0365, Fix^{Int}; LMB827, Fix^{+}; Rlv3841 *gusA*). Points represent independent replicate plants, and are an average based on measuring ten nodules. Symbols indicate data from independent experiments, grown at different times. All plants were grown for 28 days without external nitrogen.

5.4.2 Resource allocation to an intermediate-fixing strain depends on whether or not a more effective strain is available to the plant.

To test whether plants allocate resources to the single – best available – rhizobial strain based on nitrogen fixation rates, we co-inoculated plants with different combinations of a

fixing strain (Rlv3841 *gusA*[36]; Fix$^+$), a non-fixing *nifH* mutant strain (OPS0365 or OPS0366 (**chapter 3**[21]); Fix$^-$) and a strain fixing nitrogen at intermediate rates (LMB827[32]; Fix^{Int}). The fixing and non-fixing strain were *gusA* or *celB* marked so that after a post-harvest staining procedure it was easy to visually determine which nodule contained which strain (**Figure 5-3** and **Figure 5-6**).

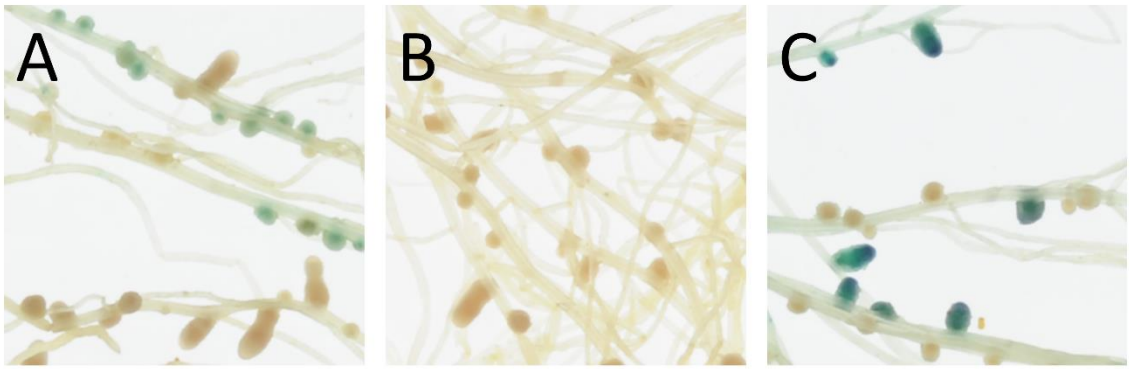


Figure 5-3. Resource allocation to an intermediate-fixing strain depends on which other strain is present on the plant. Pea root sections (2x2 cm) of roots inoculated with different combinations of a non-fixing strain (Fix$^-$, OPS0365), an intermediate-fixing strain (Fix^{Int}, LMB827) and a fixing strain (Fix$^+$, Rlv3841 *gusA*). (**a**) Unmarked Fix^{Int} strain co-inoculated with *gusA* marked (blue) Fix$^-$ strain. (**b**) Unmarked Fix^{Int} strain singly inoculated. (**c**) Unmarked Fix^{Int} strain co-inoculated with *gusA* marked (blue) Fix$^+$ strain. Plants were grown for 28 days without external nitrogen.

In two independent experiments, the intermediate-fixing strain was co-inoculated with the non-fixing strain (OPS0365), the fixing strain (Rlv3841 *gusA*), or itself (single inoculation) (**Figure 5-3**). Nodule size of the intermediate-fixing strain depended on which other strain was present on the plant ($F_{2, 26}$ = 53.2537, p < 0.0001) (**Figure 5-4**). Nodules formed by the intermediate-fixing strain measured 2.80 mm^2 (95% CI [2.50, 3.10]) when singly inoculated, and were of similar size when co-inoculated with the non-

fixing strain (a non-significant difference of 0.06 mm^2, 95% CI [-0.30, 0.42]). However, in the presence of the fixing strain, nodules of the intermediate-fixing strain were less than half this size (a significant difference of 1.54 mm^2, 95% CI [1.18, 1.90]). For comparison, the nodule sizes of the non-fixing and the fixing strains, as measured on the same plants, are shown in **Figure 5-5**. In the presence of the fixing strain, nodules of the intermediate-fixing strain became as small as non-fixing nodules (t_{18} = 0.60612, p = 0.552). Thus, the difference in size between the intermediate-fixing strain and the fixing strain was larger when both strains were co-inoculated than in single inoculations (significant interaction between strain and type of inoculation (singly or co-inoculated), $F_{1,36}$ = 17.36887, p = 0.0002). The intermediate-fixing strain typically formed about 50% of the total nodules on the plant (46 ± 13% (mean ± s.e.)), independent of whether the competing strain was fixing or non-fixing ($F_{1,8}$ = 3.3892, p = 0.1029), indicating that it was equally effective at colonising the roots and forming nodules.

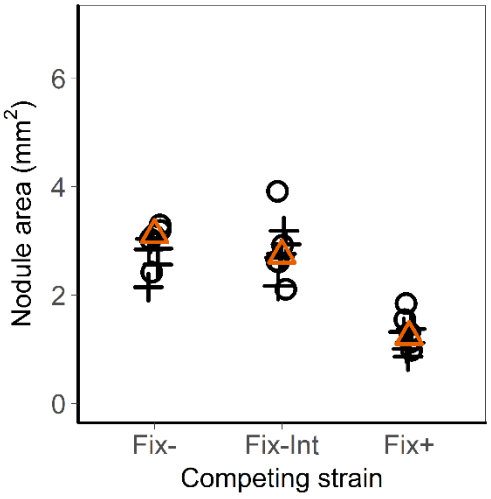


Figure 5-4. Nodule sizes of the intermediate-fixing strain (LMB827; Fix^{Int}) when co-inoculated with a non-fixing strain (OPS0365; Fix^{-}), with itself, or with

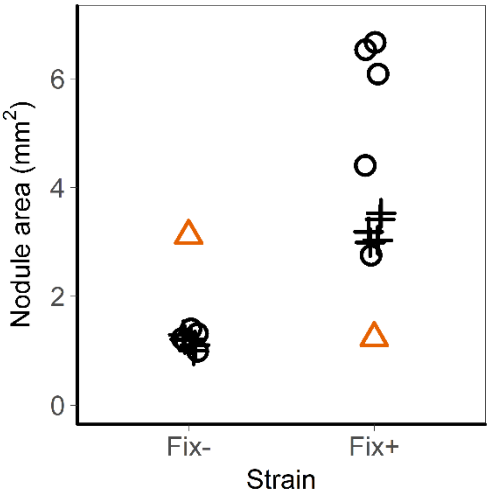


Figure 5-5. Nodule sizes of the non-fixing strain (OPS0365; Fix^{-}) and the fixing strain (Rlv3841 *gusA*; Fix^{+}), both when co-inoculated with the intermediate-fixing

the fixing strain (Rlv3841 *gusA*; Fix$^+$). The overall mean is indicated with an orange triangle. Points represent independent replicate plants, and are an average based on measuring ten nodules. Symbols indicate data from independent experiments. Plants were grown for 28 days without external nitrogen.

strain (LMB827). The mean size of the intermediate-fixing strain, as measured on the same plants, is indicated as an orange triangle for reference. Points represent an average based on measuring ten nodules. Symbols indicate data from independent experiments. Plants were grown for 28 days without external nitrogen.

These results were repeated in an experiment in which all three strains were co-inoculated in equal ratios, the non-fixing strain now being *celB* marked strain OPS0366 (**chapter 3**[21]) (**Figure 5-6**). Again, the non-fixing strain and the intermediate-fixing strain formed equally small nodules (a non-significant difference of 0.19 mm^2, 95% CI [-0.15, 0.53]), while the fixing strain formed significantly larger nodules (a difference of 2.92 mm^2, 95% CI [2.58, 3.26]), as compared to the non-fixing strain) (**Figure 5-7**). These results indicate that plants allocate resources preferentially to the strain that fixes the most nitrogen, and fewer resources to other strains present on the plant. As a result, resource allocation to an intermediate-fixing strain depends on whether or not a better fixing strain is available to the plant.

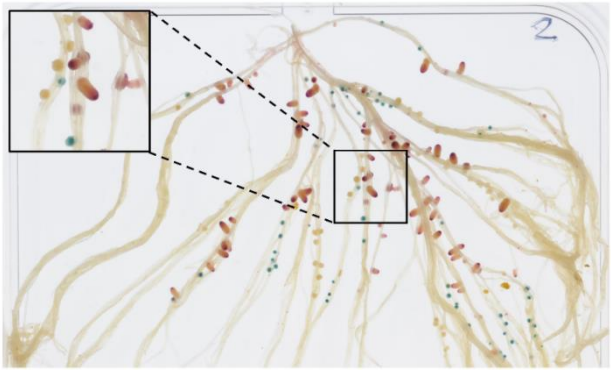


Figure 5-6. Section of pea root (12 x 7.3 cm) inoculated with three strains differing in how much nitrogen they provide. The fixing strain is magenta (Rlv3841 *gusA*), the non-fixing strain is blue (*celB* marked strain OPS0366), and the intermediate-fixing strain (LMB827) is unmarked.

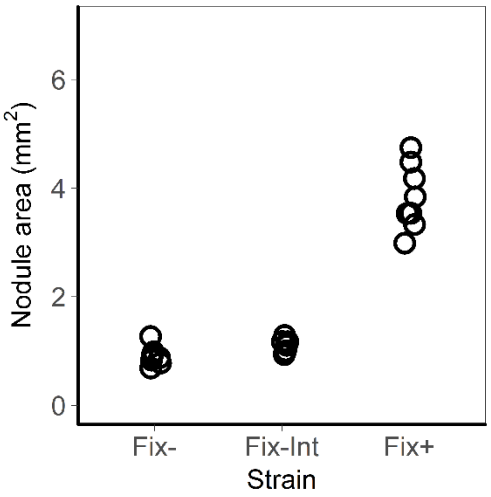


Figure 5-7. Nodule sizes of three co-inoculated strains. Nodules of the non-fixing strain (OPS0366; Fix^{-}) and the intermediate-fixing strain (LMB827; Fix^{Int}) are equally small, while the fixing strain (Rlv3841 *gusA*; Fix^{+}) forms nodules that are more than twice as large. Each point represents the average of ten nodules. Plants were grown for 29 days without external nitrogen.

5.4.3 Reduced resource allocation results in reduced nitrogen fixation rates of the intermediate-fixing strain

To test whether the reduced resource allocation to the intermediate-fixing strain by the plant also resulted in reduced nitrogen fixation, a split-root experiment was performed. Again, plants were co-inoculated with different combinations of a fixing strain (Rlv3841 *gusA*[36]), a non-fixing *nifH* mutant strain (OPS0365, **chapter 3**[21]) and a strain fixing nitrogen at intermediate rates (LMB827[32]). The split-root design allowed us to measure acetylene reduction rates of the focal strain on one half-root, given the presence of another strain on the second half-root. Using the split-root system, we again observed that the intermediate-fixing strain formed smaller nodules when co-inoculated with the fixing strain than when co-inoculated with the non-fixing strain (**Figure 5-8**), although this difference was not quite significant (a difference of 0.71 mm^2, 95% CI [-0.04, 1.46], t_{11} = -2.069, p = 0.0628). The reduction in nodule size was associated with a reduction in the rate of nitrogen fixation. Half-roots with the intermediate-fixing strain reduced

acetylene at a rate of 0.99 μmol hour^{-1}, 95% CI [0.57, 1.40] when co-inoculated with the non-fixing strain, but this rate was reduced by 0.65 μmol hour^{-1}, 95% CI [0.02, 1.27] in the presence of the fixing strain (t_7 = -2.435, p = 0.04512) – a significant reduction of 66%. Nitrogen fixation rates were positively correlated with nodule size (t_{10} = 2.714, p = 0.0218). For every mm^2 increase in average nodule size of the most developed nodules, half-roots reduced 0.91 μmol hour^{-1}, 95% CI [0.16, 1.67] more acetylene. (**Figure 5-9**). This suggests that nodule size is a good proxy for nitrogen fixation rates.

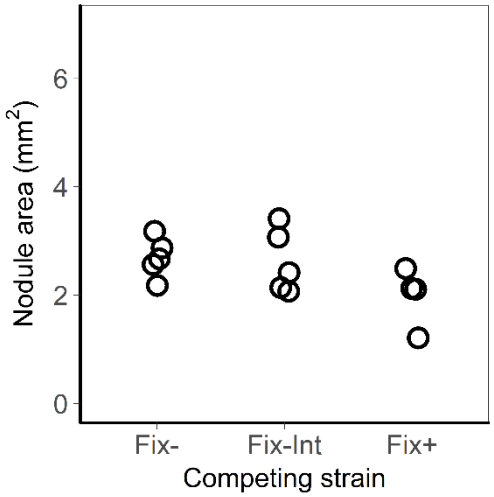


Figure 5-8. Nodule sizes of the intermediate-fixing strain (LMB827; Fix$_{Int}$) when co-inoculated with a non-fixing strain (OPS0365; Fix$^-$), itself, or the fixing strain (Rlv3841 *gusA*, Fix$^+$) in a split-root design. Points represent an average of ten nodules. Plants were grown for 28 days without external nitrogen.

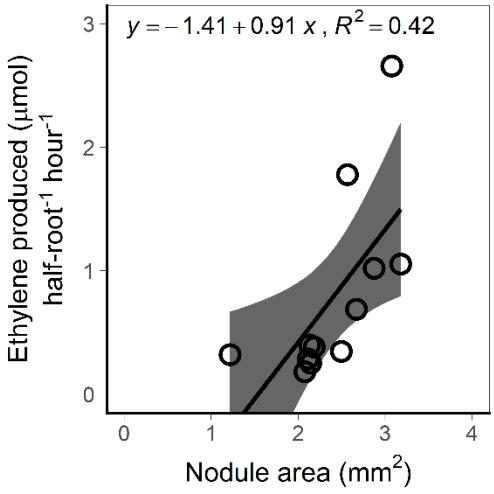


Figure 5-9. Acetylene reduction rates per half-root positively correlate with the average nodule area of intermediate-fixing strain LMB827. Regression with 95% CI shown. Acetylene reductions were performed on half-roots (with attached shoots) of plants grown for 28 days without external nitrogen.

5.4.4 Soil nitrate availability affects resource allocation to nodules in a strain-dependent manner

To test whether the effect of soil nitrate on nodule size and number depends on how much nitrogen each strain provides, plants were inoculated with a fixing strain (Rlv3841 *gusA*[36]), an intermediate-fixing strain (LMB827[32]) or a non-fixing strain (OPS0365, **chapter 3**[21]) at a range of external nitrate concentrations. Nodules of all strains decreased in size as soil nitrate availability increased ($F_{1,79}$ = 106.20611, $p < 0.001$) (**Figure 5-10**). The interaction between soil nitrate concentration and rhizobial strain was not quite significant ($F_{2,79}$ = 2.59777, p = 0.0808), but the effect was most pronounced for the fixing strain (a steeper slope as compared to the non-fixing strain, a difference of -2.25 mm^2 / g KNO_3, 95% CI [-4.95, 0.46]). Nodules of the fixing strain were largest at zero soil nitrogen (1.43 mm^2, 95% CI [1.05, 1.81] larger than non-fixing nodules), but similar in size to the less effective strains at higher soil nitrogen concentrations (**Figure 5-10**).

The number of nodules also decreased with increasing soil nitrate concentrations ($F_{1,79}$ = 129.7125, $p < 0.001$) (**Figure 5-11**). This effect depended on rhizobial strain (strain x nitrate interaction, $F_{2,79}$ = 7.4144, p = 0.0011) and was least pronounced for the fixing strain (a less steep slope as compared to the non-fixing strain, a difference of 378 nodules / g KNO_3, 95% CI [145, 611]). The fixing strain formed the smallest number of nodules at zero soil nitrogen (96 fewer nodules, 95% CI [63, 129]), but all strains formed similar numbers of nodules at higher soil nitrogen concentrations (**Figure 5-11**).

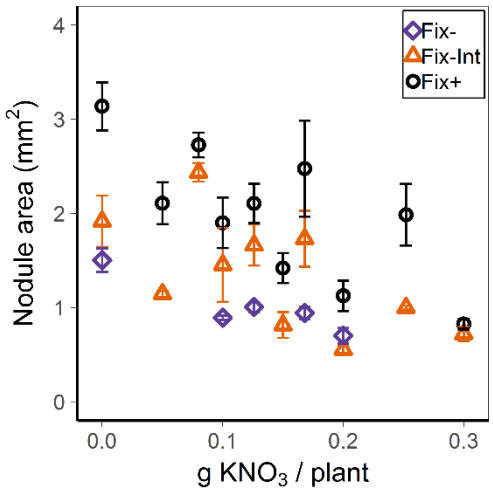


Figure 5-10. Nodule size decreases as a function of soil nitrate availability for all strains. The effect is most pronounced for the fixing strain (black circles), which at zero nitrogen forms the largest nodules, while at high nitrate concentrations all nodules are similar in size. Nodule size for each plant was based on the average of ten nodules. Points represent mean ± s.e. of 3-6 replicate plants grown for 27-29 days. KNO_3 was added at planting in the stated amounts. Fix$^-$; OPS0365, Fix^{Int}; LMB827, Fix$^+$; Rlv3841 *gusA*.

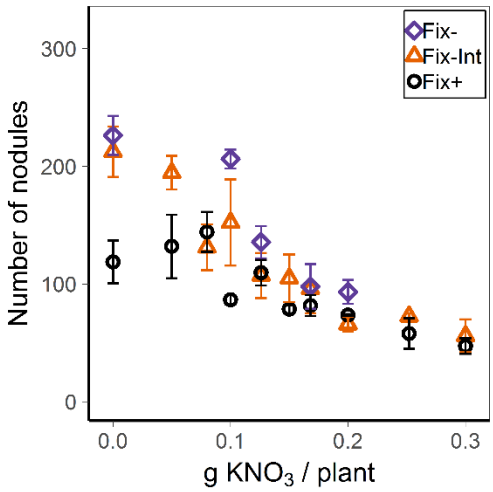


Figure 5-11. The number of nodules decreases as a function of soil nitrate availability. The effect is least pronounced for the fixing strain (black circles), which at zero nitrogen forms the smallest number of nodules, so that at high nitrate concentrations all strains form similar numbers of nodules. Points represent mean ± s.e. of 3-6 replicate plants grown for 27-29 days. KNO_3 was added at planting in the stated amounts. Fix$^-$; OPS0365, Fix^{Int}; LMB827, Fix$^+$; Rlv3841 *gusA*.

5.4.5 Resource allocation to a strain with increased carbon costs does not depend on whether or not a more efficient strain is available to the plant.

To test whether plants sanction less efficient strains, we compared resource allocation to a strain with increased carbon demands (LMB553; Lux$^+$, which emits light when fixing nitrogen due to *lux* genes under control of *nifH* promoter) and the reference strain

(LMB542; Lux^-, with *lux* plasmid but no promoter) in both single and co-inoculations. We found that the Lux^+ strain formed slightly smaller nodules than the Lux^- strain in both single inoculations (a non-signifianct difference of 0.69 mm^2, 95% CI [-0.12, 1.50], t_{12} = 1.8584, p = 0.08781) and in co-inoculations (a significant difference of 0.80 mm^2, 95% CI [0.32, 1.29], t_{10} = 3.6922, p = 0.004162) (**Figure 5-12**). Both strains were slightly smaller in co-inoculation than in single inoculations (a difference of 0.687 mm^2, 95% CI [0.06, 1.32]), which is possibly an artefact of measuring the twenty largest nodules in co-inoculation and only the ten largest nodules in single inoculation. However, the difference in size between the Lux^+ and Lux^- strains was similar whether the strains were present on their own or co-inoculated (no significant interaction between strain and inoculation type (singly or co-inoculated), $F_{1,32}$ = 0.0678, p = 0.796216). Thus, although nodules respiring more carbon are smaller, there is no indication that plants preferentially allocated resources to strains requiring less carbon, when given the choice. In fact, Lux^+ nodules were found to be very well developed in co-inoculated plants (**Figure 5-13**).

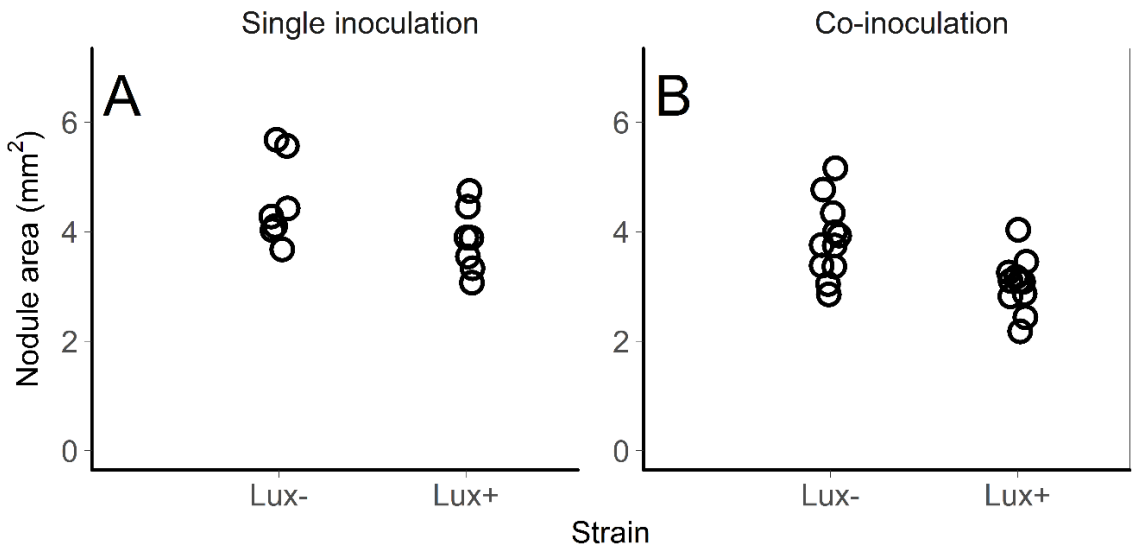


Figure 5-12. Nodule sizes of strains with increased carbon costs per gram nitrogen fixed (Lux^+; LMB553) as compared to the control strain (Lux^-; LMB542). The carbon-costly Lux^+ strain formed slightly smaller nodules than the Lux^- strain, but the difference was independent of whether strains were singly or co-inoculated. Therefore, there is no indication that plants preferentially allocated resources to a more efficient strain, given the choice. Plants were grown for 29 days without external nitrogen.

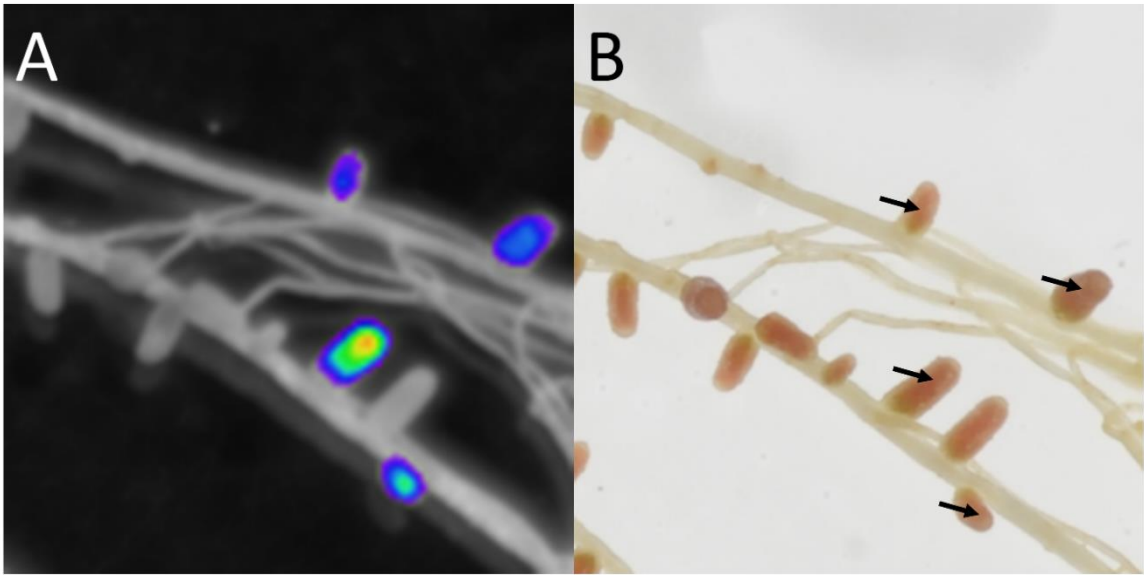


Figure 5-13. Root sections (2x2 cm) of pea roots co-inoculated with strains carrying the *lux* plasmid under the *nifH* promoter (Lux^+; LMB553), which is emitting light when fixing nitrogen, and a strain carrying the *lux* plasmid without a promoter (Lux^-; LMB542), which does not emit light. (**A**) The NightOwl image (Berthold Technologies) shows which nodules are emitting light (1000-10000 cps). (**B**) Standard photography image in which nodules found to emit light are indicated with arrows. Nodules emitting light appeared to be as well developeded as nodules not emitting light. Therefore, there is no indication that plants sanctioned nodules which used more carbon per gram nitrogen fixed.

5.5 Discussion

We used otherwise similar strains differing in both their effectiveness and their efficiency to investigate: (*i*) how plants allocate resources among multiple strains; (*ii*) whether allocation to rhizobia is based only on nitrogen provided, or also on carbon required; and (*iii*) how allocation to different strains is affected by external soil nitrogen. Mathematical modelling predicted that optimal resource allocation entails investing in the single – best available – strain, and that both effectiveness and efficiency may determine which strain is best. The model also predicted a sharp decrease in resource allocation to nodules above a threshold soil nitrogen concentration, which is lower for less effective strains (**chapter 4**). Model predictions were partially upheld. We found that plants preferentially allocated resources to more effective strains, so that a strain fixing nitrogen at reduced rates formed smaller nodules when co-inoculated with a fixing strain than when co-inoculated with itself or with a non-fixing strain. In contrast, nodule size of a strain with increased carbon demands was not affected by the presence of a more efficient strain. Increasing soil nitrogen concentrations reduced mostly nodule size of the effectively fixing strain, and mostly nodule numbers of less effective strains. At high external nitrogen concentrations, all strains formed similar numbers of nodules of similar sizes.

Our results indicate that plants preferentially invest in the most effective strain, at the expense of a less effective strain, when both are present. This result is in line with model predictions, but is likely to be contingent on numbers of nodules formed. The reduction in nodule size of the intermediate-fixing strain was associated with a reduction in nitrogen-fixation rates of those nodules, so that nitrogen fixation potential was lost. Investing in a better strain makes sense if such a strain is present, but if only very few nodules are occupied by the more effective strain, it is possible that more resources will

be allocated to the intermediate-fixing strain. The reduction in nodule size may not be so dramatic when nodule numbers are unequal, as was the case in our experiment. Further modelling taking into account numbers of nodules formed, and experiments with a range of ratios of the different strains, can help to elucidate how important nodule numbers are. An early study[37] indicates that biases in nodule numbers may have smaller effects than expected. Effectively fixing nodules increased in nodule size as they decreased in nodule number, so that nitrogen fixation potential did not linearly decrease with the number of effectively fixing nodules formed[37].

Plants did not discriminate among strains differing in efficiency, contrary to model predictions. Nodules of the less efficient strain were slightly smaller than the reference strain, but the difference in size was independent of whether the more efficient strain was co-inoculated or not. There is therefore no indication that plants discriminated between the two strains and supplied fewer resources to the less efficient strain. We cannot exclude the possibility that the strain fixed less nitrogen, so that total carbon demands did not increase much. However, carbon demands *per gram nitrogen fixed* were increased, and the plant did not discriminate on this basis. One possibility is that the difference in efficiency was too small to be relevant. Light emission is energy-intensive[34,35], but so is nitrogen fixation[38], so the relative increase may have been small. However, the difference in nodule size indicates that there was indeed a difference between strains, and the model predicts any difference to be relevant. Nonetheless, it could be that only more strongly carbon-limited plants discriminate among strains based on efficiency. Alternatively, pea plants lack the required sensory mechanisms to detect how much carbon different nodules use and discriminate between strains based on nitrogen provision, but supply carbon as needed for respiration. To see whether discrimination based on efficiency is possible, it would be worthwhile repeating the

experiment using even less efficient strains (possibly using another system than the *lux* genes), or in more carbon-limited plants, for example by providing nitrogen and shading leaves.

Strains differing in how effectively they provide nitrogen were affected by soil nitrogen in different ways, but we did not observe the predicted sharp decrease in resource allocation beyond a strain-dependent threshold nitrogen concentration. Differences in responses between strains may be better observed in a more sophisticated flow-through system in which external nitrogen concentrations are kept constant over the growing season. But if differences are small they may also be of little relevance in the field. The differential effects on nodule number and size may be more relevant, if unsurprising. It is well established that soil nitrogen affects nodule number and size[39–41], that plants with larger nodules tend to have fewer nodules[40], and that the number of nodules formed is subject to a process called autoregulation, which is known to overlap with nitrate signalling[42]. The relevance lies in the fact that the stronger effect on nodule size of the more effective strain meant that at high soil nitrogen concentrations all nodules were of similar size. Therefore, growing plants under conditions of high external nitrogen may lead to less effective discrimination between strains. Indeed, others have found that breeding plants in highly fertile environments has led to newer soybean varieties being less effective at discriminating between strains[43].

In contrast to earlier studies on sanctioning and resource allocation to nodules, our results cannot be explained by resource allocation proportional to nitrogen received. Earlier studies on sanctioning compared a more and a less effective strain, and the observed smaller nodule sizes of the less effective strain could be explained by local regulation based on nitrogen fixation rates. One other study included multiple rates of nitrogen fixation, and found that intermediate-fixing nodules received intermediate

amounts of resources[30]. However, only a small number of nodules could be exposed to treatments for a short period of time, given the limitations of using argon gas, making interpretation difficult. The regulation of nodule number and size in response to soil nitrogen requires systemic regulation based on plant nitrogen status, for which pathways are being elucidated[42,44,45], but resource allocation to individual nodules may still be proportional to nitrogen received and locally regulated. In contrast, our finding that pea plants preferentially allocated resource to the single – best available – strain, means that plants must regulate resource allocation in a more complex way. Co-inoculated plants must have been able to compare nitrogen fixation rates of multiple nodules, and adjust resource allocation accordingly, indicating the ability to integrate information in a more sophisticated way than previously established. This finding raises questions about how differential resource allocation might be regulated mechanistically, how it affects rhizobial populations in the soil, and how widespread these sophisticated mechanisms may be in the legume family.

5.6 Methods

5.6.1 Bacterial strains and culture conditions

We made use of derivatives of *R. leguminosarum* bv. viciae (Rlv) strain 3841[31], which differ in how much nitrogen they provide or how much carbon they use per gram nitrogen fixed. For strains differing in how much nitrogen they provide, we used a *gusA* marked fixing strain (Rlv3841 *gusA*[36]), and *gusA* or *celB* marked non-fixing strains (OPS0365 and OPS0366 respectively), which are completely ineffective at fixing nitrogen due to insertion of a spectinomycin resistance cassette in the *nifH* gene which is essential for nitrogen fixation (**chapter 3**[21]). As an intermediate-fixing strain, we used LMB827 which fixed nitrogen at reduced rates due to a complex insertion of the spectinomycin resistance gene[32]. For strains differing in how much carbon they cost per gram nitrogen fixed, we used strains carrying a plasmid with *lux* genes either without a promoter (LMB542[46], not emitting light), or under the *nifH* promoter (LMB553 (Poole lab,[47], emitting light when fixing nitrogen). Strains were stocked at -80°C in 15-20% glycerol, and otherwise maintained on TY agar[48] medium at 28°C, supplemented with 500 μg ml^{-1} streptomycin (all strains) and 100 μg ml^{-1} spectinomycin (OPS0365, OPS0366 and LMB827 only) or 5 μg ml^{-1} tetracycline (LMB542 and LMB553 only). Further details on all strains can be found in Table 2-1 (**chapter 2**).

5.6.2 Plant growth

Pea (*Pisum sativum* cv Avola) were surface sterilised (immersed first in 95% ethanol for 30 s and then in 2% sodium hypochlorite for 5-6 min, rinsed with distilled water) and left to germinate on 0.8% distilled water agar plates in the dark at room temperature. For single root experiments, peas were planted after five days by transferring them to sterilised 500 ml Azlon beakers filled with a 1:1 (v/v) mixture of sand and fine

vermiculite, to which 75 ml sterilised nutrient solution (as in Poole *et al*., 1994[49], but 2.67 times more concentrated, nitrogen-free unless otherwise stated) and 1 ml of the appropriate rhizobial inoculum (or water as control) was added at planting. For the split-root experiment, the last few millimetres of the emerging radicle were cut off after three days to induce lateral root formation[50], and seedlings were left to grow on agar for a further ten days. At planting, two Azlon beakers were prepared as described for the single roots and taped together, after which the root system was equally divided over both beakers, with the pea seed being kept in place with cling-film on the rim of the two beakers. In all experiments, beakers were covered with cling-film, which was slit a few days after planting to allow the seedling to grow. Plants were grown in the growth room (21°C, 16 hour photoperiod) for four weeks after planting, and watered with distilled water as necessary from two weeks onwards.

5.6.3 Experimental designs

To test how plants allocate resources among strains differing in how much nitrogen they provide, intermediate-fixing strain LMB827 was competed against non-fixing strain OPS0365, itself, or fixing strain Rlv3841 *gusA* by inoculating plants with 1:1 mixtures of the strains at a total density of 1 x 10^2 cells pot^{-1}. Two independent experiments were conducted, with each experiment consisting of five replicate plants per treatment plus two water controls grown in a fully randomised design. Resource allocation among three strains was tested by inoculating plants with a 1:1:1 mixture of the fixing strain (Rlv3841 *gusA*), the intermediate-fixing strain (LMB827), and non-fixing strain OPS0366, at a total density of 1 x 10^2 cells pot^{-1}. Eight plants and two water controls were grown in a fully randomised design.

To study how resource allocation among strains affects acetylene reduction rates, both intermediate-fixing strain LMB827 and fixing strain Rlv3841 *gusA* were competed against non-fixing strain OPS0365, intermediate-fixing strain LMB827 and fixing strain Rlv3841 *gusA* in a split-root design. Control plants had the non-fixing strain on one half-root and water control on the second half-root. Rhizobial inocula were always 1×10^2 cells half-root^{-1}. Five plants per treatment were grown in a fully randomised design.

To study the effect of soil nitrogen availability on strains differing in nitrogen fixation, plants were grown with each of the strains (non-fixing OPS0365, intermediate-fixing LMB827 and fixing Rlv3841 *gusA*) separately (density of 1×10^2 cells pot^{-1}), at a range of nitrogen concentrations. Potassium nitrate (KNO_3) was added to the nitrogen-free nutrient solution, resulting in 0, 0.05, 0.1, 0.15, 0.2 and 0.3 g KNO_3 added plant^{-1}. Three plants per treatment plus water controls were grown in a fully randomised design. The experiment was repeated independently, but with only 84% of the original nutrient solution. The difference in nitrogen concentration is accounted for in the analysis, while other nutrients are so much in abundance that results are unlikely to be affected.

To test whether resource allocation to nodules depends on carbon costs per gram nitrogen fixed, plants were grown with a strain with increased carbon costs per gram nitrogen fixed due to *lux* genes under control of the *nifH* promoter (LMB553), an otherwise isogenic strain lacking the promoter (thus never emitting light, LMB542), or both strains co-inoculated (density of 1×10^2 cells pot^{-1} in all cases). In a completely randomised design, we grew seven replicates for each of the single inoculations, twelve replicates for the co-inoculation treatment, and two water control plants.

5.6.4 Harvest

At harvest, plants were gently removed from the pots and shaken to remove any clinging substrate. For acetylene reductions, the half-root with the competing strain was removed and the half-root with the focal strain (with shoot attached) was transferred to 250 ml Duran bottles with airtight neoprene lids. Bottles contained a bit of wet tissue paper to prevent drying of the samples. Acetylene was added with a syringe through the lid to a final concentration of 2%. After one hour incubation at room temperature, three samples were taken from the jar and analysed for acetylene and ethylene by gas chromatography (Clarus 480®, Perkin-Elmer®).

For GusA and CelB staining[36,51], roots were gently washed with water, and submerged in 40 ml phosphate buffer (7 g L^{-1} NaH_2PO_4, 7.2 g L^{-1} Na_2HPO_4, 1 mM EDTA (pH 8), 1% Sarkosyl, 1 ml L^{-1} Triton) in Falcon tubes covered with aluminium foil. For GusA staining, Magenta-glc (5-bromo-6-chloro-3-indolyl-β-D-glucuronide) was added to a final concentration of 0.2 mg ml^{-1} and roots were incubated overnight at 28°C. If additional CelB staining was necessary, roots were then transferred to fresh preheated phosphate buffer and incubated for ± 2 hours at 70°C to destroy endogenous β-galactosidase activity, after which X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) was added to a final concentration of 0.25 mg ml^{-1} and roots were incubated overnight at 37°C.

To manipulate the amount of carbon used for nitrogen fixation, we used a strain with *lux* genes under the control of the *nifH* promoter (LMB553[47]), and a reference strain with the *lux* plasmid but no promoter (LMB542, Poole lab,[46]). The *lux* genes encode the luciferase enzyme which converts energy-rich compounds into light[52,53], which can be detected using sensitive charge coupled device cameras[54]. To detect whether nodules were emitting light, roots were carefully laid out on 120 x 120 mm^2 plates and imaged

with a NightOwl II camera (Berthold Technologies) at an exposure time of 120 seconds using inbuilt IndiGO software v2.0.5.0.

After any treatment, roots were gently washed, carefully laid out on 120 x 120 mm^2 plates so that all nodules were visible, and photographed for further data collection.

5.6.5 Data collection

From the photographs taken of the root systems, the number of nodules formed by each strain was counted. Nodule size per plant and per strain was estimated by measuring the area of the ten most developed (largest) nodules, so that variation in size due to developmental stage was accounted for. Using the software ImageJ[55] v1.49v, areas of irregular shapes could be accurately measured. Strains were discriminated (if necessary) based on GusA or CelB staining, or based on light emission, in which case the NightOwl images were used to indicate which nodules were emitting light. Nodule sizes were measured from the standard photography images in all cases for consistency.

5.6.6 Statistical analyses

For all experiments, a single data point per plant and per strain was obtained by averaging the nodule area measurements of the ten most developed (largest) nodules. Nodule sizes of strains when only one strain was present on the plant were analysed in a mixed effect model with experiment as random effect. Nodule sizes of the intermediate-fixing strain were analysed as a function of the competing strain in a mixed effect model including experiment as a random effect. To compare the size of non-fixing nodules with nodules of the intermediate-fixing strain in the presence of the fixing strain, a two-sided t-test assuming equal variances was performed. To confirm that the difference in nodule size between the intermediate-fixing strain and the fixing strain depended on whether

they were co-inoculated or not, a mixed-effect model was fitted with nodule size as function of strain (Fix^{Int} or Fix^{+}), inoculation (single or co-inoculation) and the interaction, with experiment as random effect. The percentage of nodules formed by the intermediate-fixing strain was analysed as a function of competing strain (Fix^{-} or Fix^{+}), with experiment as random effect. The triple inoculation experiment was analysed as a mixed-effects model, including plant as random effect, to account for the fact that on each plant nodule sizes of all three strains were measured.

The results of the split-root experiment were analysed in linear models, modelling either nodule area or ethylene produced half-$root^{-1}$ $hour^{-1}$ (average of three technical replicates) of the intermediate-fixing strain as function of the competing strain. The acetylene reduction rates per half-root of the intermediate-fixing strain were regressed against average nodule size in a linear model.

The effects of soil nitrate availability on nodule sizes and numbers of different strains were analysed in a mixed-effect models, modelling nodule area or number of nodules as a function of both strain and nitrate concentration (ANCOVA), including the interaction, and accounting for experiment as a random effect.

To test for the differences in resource allocation to strains with increased carbon costs per gram nitrogen fixed, we used a linear model in which we analysed nodule area as a function of strain (Lux^{-} and Lux^{+}), including the interaction with treatment applied (single or co-inoculation). To obtain 95% confidence intervals of the differences in size between the two strains, we did a two-sided t-test (single inoculations) and a paired t-test (co-inoculation).

5.7 References

1. Hsieh, M. H., Lam, H. M., van de Loo, F. J. & Coruzzi, G. A PII-like protein in *Arabidopsis*: putative role in nitrogen sensing. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 13965–13970 (1998).
2. Coruzzi, G. M. & Zhou, L. Carbon and nitrogen sensing and signaling in plants: emerging 'matrix effects'. *Curr. Opin. Plant Biol.* **4,** 247–253 (2001).
3. Schachtman, D. P. & Shin, R. Nutrient sensing and signaling: NPKS. *Annu. Rev. Plant Biol.* **58,** 47–69 (2007).
4. Nunes-Nesi, A., Fernie, A. R. & Stitt, M. Metabolic and signaling aspects underpinning the regulation of plant carbon nitrogen interactions. *Mol. Plant* **3,** 973–996 (2010).
5. Bisseling, T. & Scheres, B. Nutrient computation for root architecture. *Science* **346,** 300–301 (2014).
6. Vance, C. P. & Heichel, G. H. Carbon in N_2 fixation: limitation or exquisite adaptation. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **42,** 373–390 (1991).
7. Bergersen, F. J. Some Australian studies relating to the long-term effects of the inoculation of legume seeds. *Plant Soil* **32,** 727–736 (1970).
8. Burdon, J. & Gibson, A. Variation in the effectiveness of symbiotic associations between native rhizobia and temperate Australian *Acacia*: within-species interactions. *J. Appl. Ecol.* **36,** 398–408 (1999).
9. Fening, J. O. & Danso, S. K. A. Variation in symbiotic effectiveness of cowpea bradyrhizobia indigenous to Ghanaian soils. *Appl. Soil Ecol.* **21,** 23–29 (2002).
10. Sprent, J. Mutual sanctions. *Nature* **422,** 672–674 (2003).
11. Thrall, P. H., Laine, A.-L., Broadhurst, L. M., Bagnall, D. J. & Brockwell, J. Symbiotic effectiveness of rhizobial mutualists varies in interactions with native Australian legume genera. *PLoS One* **6,** e23545 (2011).
12. Voisin, A. S., Salon, C., Jeudy, C. & Warembourg, F. R. Symbiotic N_2 fixation activity in relation to C economy of *Pisum sativum* L. as a function of plant phenology. *J. Exp. Bot.*

54, 2733–44 (2003).

13. Kaschuk, G., Kuyper, T. W., Leffelaar, P. A., Hungria, M. & Giller, K. E. Are the rates of photosynthesis stimulated by the carbon sink strength of rhizobial and arbuscular mycorrhizal symbioses? *Soil Biol. Biochem.* **41,** 1233–1244 (2009).
14. Kaschuk, G., Hungria, M., Leffelaar, P. a, Giller, K. E. & Kuyper, T. W. Differences in photosynthetic behaviour and leaf senescence of soybean (*Glycine max* [L.] Merrill) dependent on N_2 fixation or nitrate supply. *Plant Biol.* **12,** 60–69 (2010).
15. Denison, R. F. Legume sanctions and the evolution of symbiotic cooperation by rhizobia. *Am. Nat.* **156,** 567–576 (2000).
16. West, S. A., Kiers, E. T., Simms, E. L. & Denison, R. F. Sanctions and mutualism stability: why do rhizobia fix nitrogen? *Proc. R. Soc. B.* **269,** 685–94 (2002).
17. Kiers, E. T. & Denison, R. F. Sanctions, cooperation, and the stability of plant-rhizosphere mutualisms. *Annu. Rev. Ecol. Evol. Syst.* **39,** 215–236 (2008).
18. Heath, K. & Tiffin, P. Stabilizing mechanisms in a legume–rhizobium mutualism. *Evolution (N. Y).* **63,** 652–662 (2009).
19. Oono, R., Anderson, C. G. & Denison, R. F. Failure to fix nitrogen by non-reproductive symbiotic rhizobia triggers host sanctions that reduce fitness of their reproductive clonemates. *Proc. R. Soc. B.* **278,** 2698–703 (2011).
20. Kiers, E. T., Rousseau, R. A., West, S. A. & Denison, R. F. Host sanctions and the legume-rhizobium mutualism. *Nature* **425,** 78–81 (2003).
21. Westhoek, A. *et al.* Policing the legume-Rhizobium symbiosis: a critical test of partner choice. *Sci. Rep.* 1–10 (2017).
22. Ruffel, S. *et al.* Systemic signaling of the plant nitrogen status triggers specific transcriptome responses depending on the nitrogen source in *Medicago truncatula*. *Plant Physiol.* **146,** 2020–35 (2008).
23. Gubry-Rangin, C., Garcia, M. & Béna, G. Partner choice in *Medicago truncatula-Sinorhizobium* symbiosis. *Proc. Biol. Sci.* **277,** 1947–51 (2010).
24. Jeudy, C. *et al.* Adaptation of *Medicago truncatula* to nitrogen limitation is modulated via

local and systemic nodule developmental responses. *New Phytol.* **185,** 817–28 (2010).

25. Laguerre, G. *et al.* Local and systemic N signaling are involved in *Medicago truncatula* preference for the most efficient *Sinorhizobium* symbiotic partners. *New Phytol.* **195,** 437–449 (2012).

26. Dowling, D. N. & Broughton, W. J. Competition for nodulation of legumes. *Annu. Rev. Microbiol.* **40,** 131–157 (1986).

27. Souza, V. *et al.* Ethnomicrobiology: do agricultural practices modify the population structure of the nitrogen fixing bacteria *Rhizobium etli* biovar *phaseoli*? *J. Ethnobiol.* **17,** 249–266 (1997).

28. Hagen, M. J. & Hamrick, J. L. A hierarchical analysis of population genetic structure in *Rhizobium leguminosarum* bv. *trifolii*. *Mol. Ecol.* **5,** 177–186 (1996).

29. Silva, C., Eguiarte, L. E. & Souza, V. Reticulated and epidemic population genetic structure of *Rhizobium etli* biovar *phaseoli* in a traditionally managed locality in Mexico. *Mol. Ecol.* **8,** 277–287 (1999).

30. Kiers, T. E., Rousseau, R. A. & Denison, R. F. Measured sanctions: legume hosts detect quantitative variation in rhizobium cooperation and punish accordingly. *Evol. Ecol. Res.* **8,** 1077–1086 (2006).

31. Johnston, A. W. & Beringer, J. E. Identification of the rhizobium strains in pea root nodules using genetic markers. *J. Gen. Microbiol.* **87,** 343–350 (1975).

32. Webb, I. U. C. Characterisation of the *fixABCX* operon in symbiotic nitrogen fixation. (University of East Anglia, 2016).

33. Close, D. *et al.* The evolution of the bacterial luciferase gene cassette (*lux*) as a real-time bioreporter. *Sensors* **12,** 732–752 (2012).

34. Van Der Meer, J. R., Tropel, D. & Jaspers, M. Illuminating the detection chain of bacterial bioreporters. *Environ. Microbiol.* **6,** 1005–1020 (2004).

35. Herron, P. M., Gage, D. J., Arango Pinedo, C., Haider, Z. K. & Cardon, Z. G. Better to light a candle than curse the darkness: illuminating spatial localization and temporal dynamics of rapid microbial growth in the rhizosphere. *Front. Plant Sci.* **4,** 1–14 (2013).

36. Sánchez-Cañizares, C. & Palacios, J. Construction of a marker system for the evaluation of competitiveness for legume nodulation in *Rhizobium* strains. *J. Microbiol. Methods* **92,** 246–249 (2013).

37. Singleton, W. & Stockinger, K. R. Compensation against ineffective nodulation in soybean. *Crop Sci.* **23,** 69–72 (1983).

38. Lodwig, E. & Poole, P. Metabolism of *Rhizobium* bacteroids. *CRC. Crit. Rev. Plant Sci.* **22,** 37–41 (2003).

39. Streeter, J. & Wong, P. P. Inhibition of legume nodule formation and N_2 fixation by nitrate. *CRC. Crit. Rev. Plant Sci.* **7,** 1–23 (1988).

40. Parsons, R., Stanforth, A., Raven, J. a. & Sprent, J. I. Nodule growth and activity may be regulated by a feedback mechanism involving phloem nitrogen. *Plant, Cell Environ.* **16,** 125–136 (1993).

41. Caballero Mellado, J. & Martinez Romero, E. Soil fertilization limits the genetic diversity of *Rhizobium* in bean nodules. *Svmbiosis* **26,** 111–121 (1999).

42. Okamoto, S. *et al.* Nod factor/nitrate-induced *CLE* genes that drive HAR1-mediated systemic regulation of nodulation. *Plant Cell Physiol.* **50,** 67–77 (2009).

43. Kiers, E. T., Hutton, M. G. & Denison, R. F. Human selection and the relaxation of legume defences against ineffective rhizobia. *Proc. R. Soc. B.* **274,** 3119–26 (2007).

44. Ferguson, B. J. *et al.* Molecular analysis of legume nodule development and autoregulation. *J. Integr. Plant Biol.* **52,** 61–76 (2010).

45. Takahara, M. *et al.* TOO MUCH LOVE, a novel kelch repeat-containing F-box protein, functions in the long-distance regulation of the legume-*rhizobium* symbiosis. *Plant Cell Physiol.* **54,** 433–447 (2013).

46. Pini, F. *et al.* Lux bacterial biosensors for in vivo spatiotemporal mapping of root secretion. *Plant Physiol.* 1302 (2017). doi:10.1104/pp.16.01302

47. Frederix, M. *et al.* Mutation of *praR* in *Rhizobium leguminosarum* enhances root biofilms, improving nodulation competitiveness by increased expression of attachment proteins. *Mol. Microbiol.* **93,** 464–478 (2014).

48. Beringer, J. E. R factor transfer in *Rhizobium leguminosarum*. *J. Gen. Microbiol.* **84,** 188–198 (1974).

49. Poole, P. S., Blyth, A., Reid, C. J. & Walters, K. myo-Inositol catabolism and catabolite regulation in *Rhizobium leguminosarum* bv. *viciae*. *Microbiology* **140,** 2787–2795 (1994).

50. Dener, E., Kacelnik, A. & Shemesh, H. Pea plants show risk sensitivity. *Curr. Biol.* **26,** 1763–1767 (2016).

51. Sessitsch, A., Wilson, K. J., Akkermans, A. D. L. & De Vos, W. M. Simultaneous detection of different *Rhizobium* strains marked with either the *Escherichia coli gusA* gene or the *Pyrococcus furiosus celB* gene. *Appl. Environ. Microbiol.* **62,** 4191–4194 (1996).

52. Stewart, S. A. B. & Williams, P. *lux* genes and the applications of bacterial bioluminescence. *J. Gen. Microbiol.* **138,** 1289–1300 (1992).

53. G.M. Gahan, C. The bacterial *lux* reporter system: applications in bacterial localisation studies. *Curr. Gene Ther.* **12,** 12–19 (2012).

54. Contag, C. H. & Bachmann, M. H. Advances in in vivo bioluminescence imaging of gene expression. *Annu. Rev. Biomed. Eng.* **4,** 235–260 (2002).

55. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9,** 671–675 (2012).

6 General discussion

'*The* [legume-rhizobia] *symbiosis is a rich resource for evolutionary fact and speculation, but its complexity and diversity should warn us not to expect easy answers*'.

J.P.W. Young and A.W.B. Johnston, 1989[1]

'There are no miracles in agricultural production.'

Norman Borlaug, Nobel Lecture, 1970

6.1 Overview

This chapter gives an overview of the results (6.2) and discusses them in a more general context than was done in each chapter. Implications of the findings and avenues for future research are highlighted (6.3). The chapter ends with general conclusions (6.4).

6.2 Summary of results

My aim in the work described in this thesis was to establish how legume plants respond to rhizobial strains which differ in how much nitrogen they provide and how much carbon they use, and therefore in how effective they are at enhancing plant growth. First, in **chapter 3**[2] it was established that pea plants are not able to exert partner choice and discriminate against non-fixing strains before establishing the symbiosis: fixing and non-fixing but otherwise isogenic strains have equal chances of forming a nodule on the plant root. It was discussed that partner choice is unlikely to evolve as a mechanism to protect against ineffective strains as it requires accurate detection of a trait which does not become apparent until after the symbiosis has been established. Such accurate detection is unlikely given the potential for dishonest signalling by the rhizobia[3]. Furthermore, even if plants could evolve a mechanism to detect mutations in nitrogen fixation, the shorter generation time of rhizobia[4] as compared to plants would always put plants one step behind. It was argued that sanctioning is a more robust mechanism against ineffective strains – requiring no specific recognition signals and relying just on differential resource allocation based on benefits received. I corroborated earlier findings[2,5,6] that nodules providing little nitrogen are smaller, indicating reduced plant resource allocation.

In **chapter 4**, I further explored how plants might differentially allocate resources to rhizobial strains in a model of plant growth and resource allocation. I considered strains varying in how much nitrogen they provide and how much carbon they use, under varying soil nitrogen conditions, and asked how plants should allocate resources to achieve maximum growth rates. The model predicted that optimal resource allocation involves allocating all resources to the single – best available – strain. Therefore, resource allocation to a given strain depends on other strains present on the same plant.

Both effectiveness and efficiency determine which strain would be the best strain to invest in, and the relative importance depends on how nitrogen limited the plant is. The model also predicted that plants should cease to invest in nodules above a threshold soil nitrogen concentration – which is higher for more effective and for more efficient strains.

In **chapter 5**, I tested the model predictions in growth room experiments. I found that a strain fixing nitrogen at intermediate rates formed intermediate-sized nodules when singly inoculated or co-inoculated with a non-fixing strain, but formed smaller nodules when co-inoculated with a fixing strain. When plants were inoculated with three strains (non-fixing, intermediate-fixing and fixing), the fixing strain formed large nodules while the two other strains formed equally small nodules. So, in line with model predictions, resource allocation to a rhizobial strain depended on other strains present on the plant and resources were preferentially allocated to the single – best available – strain. In contrast to model predictions, resource allocation to a strain with increased carbon requirements did not depend on whether or not a more efficient strain was present. I could also not detect a threshold concentration of soil nitrogen which was strain-dependent, but I did find that at high soil nitrogen concentrations all strains formed similar numbers of nodules of similar sizes.

The results of this thesis have implications for three lines of further research: *i*) understanding how sanctioning is regulated in the plant; *ii*) understanding the evolutionary dynamics of the symbiosis; and *iii*) how the knowledge gained can be used to select for effective (inoculant) strains and to enhance yields.

6.3 Implications and future work

6.3.1 Plant regulation of sanctioning

The results of **chapter 5** indicate that plants regulate nodule growth systemically and integrate information across the root system. Previous studies showing that ineffective nodules were smaller than effective nodules[2,5,6] could be explained by a model in which resource allocation was proportional to nitrogen received. This type of regulation can be carried out locally, i.e. at the level of individual nodules or even bacteroids. In contrast, the findings presented in this thesis indicate that pea plants are able to compare nitrogen received from different nodules and adjust resource allocation – and therefore nodule growth (or vice versa) – accordingly. This integration of information requires systemic regulation. Systemic regulation of nodule initiation, growth, and activity was considered before sanctioning was even suggested, because of the observations that nitrate supply affects nodule initiation and growth, and that plants tend to have few large nodules, or many small nodules[7,8]. It was proposed that concentrations of nitrogenous compounds in the phloem entering the nodule, obviously dependent on plant nitrogen status, affect nodule growth and activity, possibly through regulation of oxygen supply[8]. However, this model does not explain how nitrogen fixation rates of different nodules are compared. Here, I propose an alternative mechanism based on cytokinin concentrations.

Regulation of nodule development based on nitrogen in the phloem seems plausible at first sight, but open questions remain, especially given the results described in this thesis. Undeniably, nitrogenous compounds circulating through the plant affect nodules. Nitrogen does not only come out of nodules, but also goes into the nodules[9–11]. Bacteroids rely not only on plant-derived carbon, but also on amino acid supply from the plant[9,11,12]. This supply can be used to regulate nodule activity and bacteroid development, as has been shown[11,13]. However, it remains unclear how plants would

sense how much nitrogen comes out of nodules, and compare it to nitrogen levels in the plant or to other nodules, which is required to explain the observed results. Forming an idea of how plants might sense and compare how much nitrogen nodules provide is extremely difficult, since levels of the different nitrogenous compounds in the phloem fluctuate during the day[14], as do rates of nitrogen fixation[15–17]. In addition, both the amino acids in the phloem[18] and the type of nitrogenous compounds exported from nodules[19,20] are species-dependent. The regulation of nodule growth and its dependency on nitrogen fixation rates has received little research attention, but insight may be gained from research on systemic regulation of nodule initiation, and from research on nodule development.

Nodule initiation is systemically regulated in a process called autoregulation of nodulation (AON)[21], in which cytokinin is an important regulator. Cytokinin is induced in the early stages of infection and is needed to initiate the cortical cell divisions that will lead to nodule formation[22–28]. Localised high concentrations can induce formation of pseudonodules even in the absence of nodulating rhizobia[29,30]. Nodule formation subsequently inhibits initiation of other nodules, again through regulation by cytokinin. Cytokinin induces the formation of CLE peptides (CLAVATA3/ESR-RELATED)[31–34], which are perceived in the shoot by LR-RKK (leucine-rich repeat receptor kinase) receptors, in turn inducing a shoot-derived systemic signal inhibiting further nodule formation[35–37]. In split-root experiments, addition of cytokinin to one half-root decreased nodule formation in the other half-root[38]. Autoregulation of nodulation overlaps with nitrate signalling pathways. Cytokinin concentrations correlate with nitrogen supply in a range of species[39,40], nitrate also induces formation of CLE peptides (sometimes but not always the same as the rhizobia-induced CLE peptides[31,36,37]) and reduces nodulation via the same LRR-RK receptors as AON[36,41]. So, it seems that although locally high

cytokinin concentrations are needed to induce cortical cell divisions at the start of nodule formation, systemically increased cytokinin levels – through formation of nodules or uptake of soil nitrogen – reduces nodulation.

Nodule growth is regulated by the nodule meristem, which has been likened to lateral root meristems based on similar developmental processes[42–44]. Cytokinin is a key regulator of meristem activity[27,28,44–47]. In lateral roots, cytokinin stimulates root elongation[30,48] and in nodule meristems, cytokinin receptors and response genes are also active[24,27,49–51]. Cytokinin levels are especially high in nodule meristems of young, developing nodules, but decline with maturity, and cytokinin is involved in preventing early senescence[44,52,53]. Given the known roles of cytokinin in systemic regulation of nodule numbers and in regulation of nodule meristems, it is tempting to speculate that cytokinin is also involved in systemically regulating nodule growth and size.

Systemic regulation of nodule size based on cytokinin concentrations could be envisioned as follows (**Figure 6-1**): nodules grow faster and/or senesce later when the cytokinin concentration in the nodule is high compared to the concentration in the incoming phloem. The concentration in the nodule is determined by the rates of nitrogen fixation, and the concentration in the rest of the root system is determined by plant nitrogen status. Such regulation could explain the observed results. In single inoculations, under low external nitrogen conditions nodules of the fixing strain grow as their cytokinin concentration is higher than the root. By fixing nitrogen, they increase systemic levels of cytokinin thus inhibiting further nodule formation (AON). Thus, few, large nodules are formed and increasing external nitrogen affects mostly nodule size (**chapter 5**). In contrast, nodules of ineffective strains grow less since their cytokinin levels are not much higher than levels in the root, but systemic levels of cytokinin are also not increased so that new nodules continue to be formed. Thus, many, small nodules

are formed and increasing external nitrogen affects mostly nodule number (**chapter 5**). Nodules of an intermediate-fixing strain would grow in the absence of a better strain, since cytokinin concentrations in the nodules would be relatively high compared to the root system. However, the presence of a better fixing strain on the same plant increases systemic cytokinin levels, so that the cytokinin concentration of the intermediate-fixing nodules does not exceed the root concentration and nodules fail to grow (**chapter 5**). Under this hypothesis, it would be possible that resources are allocated to more than one strain if the differences between strains are small. Regulation based on cytokinin concentrations, which depend on rates of nitrogen fixation, would also explain why I failed to see discrimination based on differences among strains in carbon requirements.

Some evidence is available to support the proposed hypothesis of systemic regulation of nodule size via cytokinin concentrations. First, there is an indication that cytokinin levels may correlate with nitrogen fixation rates. Nitrogen fixed by rhizobia is incorporated into glutamine in the plant cytosol as the first step in amino acid formation[54–58]. Recently, it was shown in rice that cytokinin levels are increased not only in response to nitrate, but also in response to glutamine, through upregulation of *ipt* genes which encode for the first and rate-limiting step in cytokinin production[40]. It has also been shown that several *ipt* genes are expressed in *Lotus japonicus* nodules. Different *ipt* genes are expressed at different stages, and one gene (LjIPT3) was expressed specifically in vascular bundles and required for full development of mature nodules by preventing senescence[59]. Furthermore, in *Medicago truncatula*, the cytokinin response gene MtRR4 has been proposed to negatively regulate cytokinin signalling in mature nodules[50]. Intriguingly, mutant nodules with reduced expression of MtRR4 continued to grow despite not fixing nitrogen when the wild-type nodules were already senescing[60].

Testing the proposed cytokinin-based regulation still requires a lot of work, and the devil is likely to be in the details. There are several different cytokines which can also be interconverted, and their levels are tightly regulated both spatially and temporally by continuous production and degradation[61]. Cytokinin interacts with other plant hormones such as auxin[62,63] and ethylene[32,64] in regulating processes, and microRNAs have been shown to alter sensitivity to cytokinin, so that cytokines can have different effects at different stages of development[65,66]. A first step would be to obtain data on *ipt* gene expression or cytokinin concentrations in individual nodules under differing circumstances, including different combinations of strains on the same plant and under different nitrogen concentrations. It would also be worthwhile to manipulate cytokinin concentrations in the nodule and see what the effects on nodule growth are. Understanding plant genetic aspects of regulation would be easier in for example *Medicago truncatula*, for which plant mutant libraries exist[67]. Another approach would be to develop a simple mathematical model in which nodule growth is related to the difference in cytokinin concentrations between root and nodule. Such a model could be used to see whether the observed results can indeed be obtained based on this mechanism, and what the difference in cytokinin concentrations is likely to be, which can subsequently be compared to data.

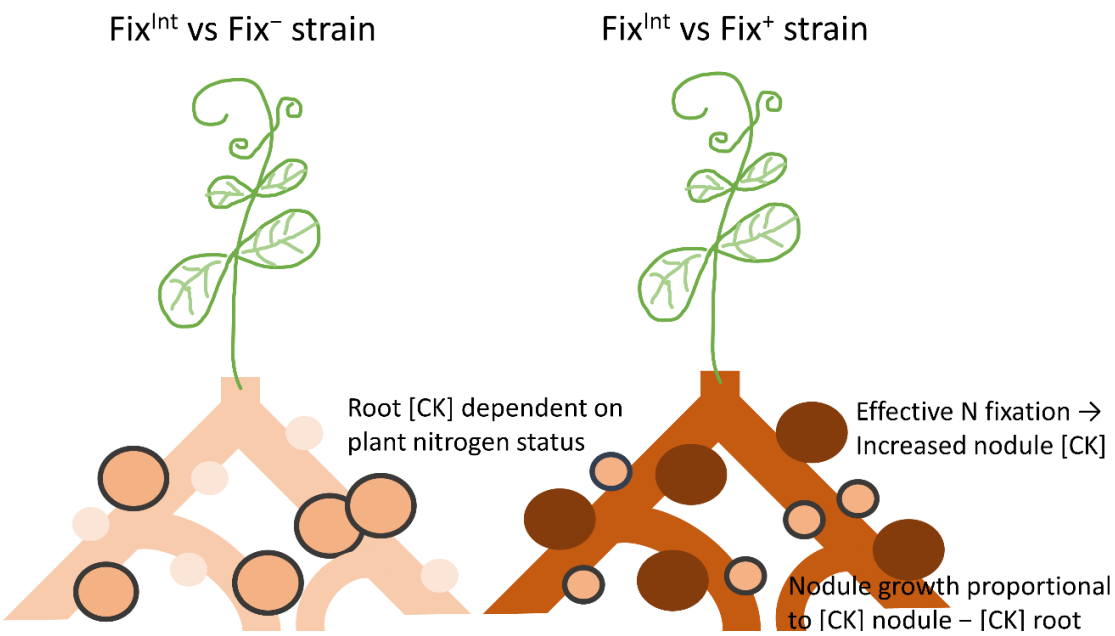


Figure 6-1. Hypothesis of cytokinin-based regulation of nodule initiation and development. The hypothesis explains integration of information needed to compare different nodules and adjust resource allocation accordingly. Cytokinin concentrations in the root system depend on plant nitrogen status, and cytokinin concentrations in the nodule depend on rates of nitrogen fixation. Only nodules with higher cytokinin concentrations than the root develop. Therefore, only nodules of the most effectively fixing strain(s) (darker colours) will develop and a strain fixing nitrogen at intermediate rates (outlined nodules) will form small nodules when a more effective strain is present.

6.3.2 The legume-rhizobia symbiosis from an evolutionary biology perspective

The legume-rhizobia symbiosis has been studied by evolutionary biologists interested in understanding the stability of mutualisms[68]. They predicted ineffective strains are likely to arise for two reasons[3,69,70]: (*i*) legume hosts tend to be infected with multiple strains[71–74], so that in the absence of sanctions, each strain can potentially divert resources from the energy-intensive process of nitrogen fixation into rhizobial fitness, leaving nitrogen fixation to other strains, and (*ii*) rhizobia are horizontally transmitted and therefore their fitness is uncoupled from the plant's fitness[3,69,70]. Partner choice and sanctioning were

proposed as possible mechanisms preventing the spread of ineffective strains[3,69,70,75,76]. In this thesis, it was shown that pea plants are not able to exert partner choice, since fixing and non-fixing strains which were otherwise similar formed equal numbers of nodules on the plant root (**chapter 3**[2]). With regards to sanctioning, I: (*i*) supported previous results[5,6] – which were based on experiments using argon gas – with experiments using mutant strains, showing that effectively fixing nodules are larger than non-fixing nodules, indicating they received more resources (**chapter 3**[2]); (*ii*) showed that plants preferentially allocated resources to the single – best available – strain, so that resource allocation to an intermediate-fixing strain depended on whether or not a better alternative was available (**chapter 5**), in line with predictions about optimal resource allocation (**chapter 4**); (*iii*) found no indication that pea plants sanctioned based on carbon demands of strains (**chapter 5**), which contrasted with model predictions (**chapter 4**); and (*iv*) showed that at high soil nitrogen concentrations resource allocation to all strains became similar, indicating that sanctioning may be less effective in fertile environments (**chapter 5**). The results described in this thesis have implications for further research on the legume-rhizobia symbiosis from an evolutionary biology perspective.

The effects of sanctioning on rhizobial fitness have not been directly addressed in this thesis, but nonetheless the findings may contribute to understanding how sanctioning affects rhizobial fitness. There is considerable evidence that reduced resource allocation leading to smaller nodules does indeed decrease fitness of the strains in those nodules, as fewer viable bacteria are present[5,6,77]. Nonetheless, ineffective strains in the field are common[78–80]. This is sometimes portrayed as a paradox[81], but in fact the two are not mutually exclusive for several reasons. Firstly, rhizobial fitness does not depend solely on resources received from the plant, but also on survival in the soil[82]. Secondly, mixed infection nodules could be a way for ineffective strains to avoid

sanctioning if nodules are sanctioned on the level of whole nodules, rather than individual bacteroids[3,69,83] (as discussed in **chapter 3**[2]). If nodule size is regulated by the nodule meristem, as proposed here (6.3.1), sanctioning would indeed take place on the whole nodule level. Third, strains of different effectiveness may coexist even in the presence of sanctioning[81,84]. Modelling predicted that the best strain to invest in may depend on external conditions (**chapter 4**) and resource allocation to a strain fixing nitrogen at reduced rates was substantial when the plant had no better alternative (**chapter 5**). In addition, sanctioning may not always result in near-complete reduction of resource allocation to the less effective strain. For example, the less effective strain may still receive a relatively large proportion of resources when few nodules of a more effective strain are present (discussed in **chapter 5**). Under the hypothesis of regulation based on cytokinin concentrations in the nodules (6.3.1), if strains differ only slightly, differences in resource allocation may also be small. Modelling plant sanctions with different combinations and proportions of strains combined with (multigenerational) experiments could shed more light on how sanctioning affects rhizobial fitness.

In the literature about sanctioning, it has generally been assumed that the diversion of plant-derived carbon away from nitrogen-fixation into rhizobial fitness leads to reduced rates of nitrogen fixation. I argued that it is equally possible that nitrogen-fixation is not reduced, but the plant simply allocates more carbon. I therefore tested whether pea plants discriminate among strains based on how much carbon they require for nitrogen fixation, and found no indication that plants discriminated based on efficiency (**chapter 5**). This result indicates that rhizobial strains may increase their fitness by demanding more carbon without reducing their rate of nitrogen-fixation. In fact, the model predicts that the higher the rate of nitrogen-fixation, the more carbon a plant should be willing to invest in a given rhizobial strain (**chapter 4**). Interestingly,

rhizobial strains have been described which synthesise compounds called rhizopines during effective nitrogen-fixation, and (later) use them as a carbon source[85–87]. However, the role of these rhizopines remains to be elucidated and it is unlikely that they act only as a growth substrate[86].

From a plant perspective, it would be interesting to see how the evolution of sanctioning relates to other aspects of legume evolution and how widespread sanctioning can be expected to be. The experiments described here were conducted on pea plants, which are part of the Inverted Repeat Lacking Clade (IRLC)[88]. Legumes in this group evolved relatively recently (39 million years ago[89]). It is therefore possible that the sophisticated mechanisms that underlie the observed resource allocation to the single – best available – strain are limited to this clade. On the other hand, in the symbiosis with arbuscular mycorrhizal (AM) fungi –which is older than the rhizobial symbiosis[90] – preferential resource allocation to strains providing more phosphorus has been shown[91]. Furthermore, regulation may be widespread if it is based on mechanisms evolved for resource allocation to roots in heterogeneous environments[92]. IRLC legumes form indeterminate nodules[88] and if nodule size is regulated by the meristem (6.3.1), this raises questions about how effectively regulation can take place in determinate nodules. But discrimination based on effectiveness has been shown, and was in fact first shown, among effective and less effective determinate (soybean) nodules[5]. So regulation of resource allocation in response to effectiveness seems to be widespread, but it remains to be seen whether resource allocation to the single – best available – strain, as observed here, is also widespread. IRLC legumes are known for forming symbioses with a narrow range of strains[88] (in contrast to promiscuous nodulation, thought to be ancestral[93]). Possibly, allocating resources to the single – best available – strain works best when few strains are present (in this thesis up to three strains were tested), but less well in

promiscuous species infected with many strains, each potentially forming only a small number of nodules. Repeating the experiments in a wider range of legume species, and even with AM fungi, may help in understanding how widespread the observed mechanisms of sanctioning are.

6.3.3 The legume-rhizobia symbiosis from an agricultural perspective

From an agricultural perspective, the legume-rhizobia symbiosis has been studied with an emphasis on enhancing yields. A major approach has been, and continues to be, finding and selecting rhizobial inoculant strains competitive in colonising roots and effective at enhancing yields[94–100]. By addressing the question of partner choice, I studied one aspect of rhizobial competitiveness and the finding that an ineffective strain can be equally competitive as an effective strain (**chapter 3**[2]) helps explain the difficulties in using inocula to enhance yields[94,95,100,101]. I studied sanctioning both theoretically (**chapter 4)** and experimentally (**chapter 5**), proposing that understanding how plants allocate resources to different rhizobial strains may also help in understanding how to select for elite strains, and how to effectively use rhizobia to enhance yields. The findings presented in this thesis are limited in that they are based on a single plant species using a single set of similar rhizobial strains. Further research as outlined above (6.3.1 and 6.3.2) can be imagined to be of future help in understanding how to enhance yields. For example, understanding how sanctioning is regulated genetically may guide plant breeding efforts. Understanding whether sanctioning is as effective in promiscuous species can guide which varieties need to be used under different circumstances. Soybean varieties for example differ in how promiscuous they are, and release of varieties has been tailored to specific circumstances before[102]. The current results also

provide some suggestions for further work in the lab and under field conditions, especially if the results can be confirmed for a wider range of species.

In the lab, screening for effective rhizobial strains is often done by singly inoculating plants with each strain to be tested[94,97,98,103]. This process is labour intensive and time consuming. The results from this thesis indicate that screening might work just as well – or even better – when plants are co-inoculated. In single inoculations, the intermediate-fixing strain may have scored well in the traditional approach, forming relatively large nodules, albeit slightly smaller than the more effective strain. However, by co-inoculating the two strains the difference in size became larger, thus more clearly indicating which strain was more effective (**chapter 5**). Obviously, being able to screen strains by co-inoculating plants has the potential to save a lot of time and effort compared to single inoculations. Ideally, this approach would need to be coupled with fast and easy identification methods of the strains, so that differences in competitiveness can be simultaneously assessed. It seems that co-inoculation screening would be best performed under low soil nitrogen conditions, as is common practice[94,97,104], as differences between strains became smaller with increasing soil nitrogen (**chapter 5**).

In screening for effective rhizobial strains, often used indicators are nodule number, nodule weight, plant yield and plant nitrogen content[79,97,98,100,105]. In this thesis, nodule size – as measured from photographs – was successfully used as an indicator of rhizobial effectiveness. The strain known to have the highest rate of nitrogen fixation formed the largest nodules, which was in line with previous observations based on nodule weight[5,6] (**chapter 3**[2]). Although it may be worthwhile to establish the relationship between nodule size and weight for different species, nodule size may be a very good alternative indicator of rhizobial effectiveness. Using nodule size has the advantage of being non-destructive, so that nodule sizes could be followed over time.

Measuring nodule size is also less time consuming and labour intensive than picking and drying nodules to obtain nodule weights. However, development of good software to count and measure size of nodules is strongly recommended. Given the technological advances in image processing[106] this seems feasible, and automating the process would not only save a lot of time and effort, but potentially yield more accurate and consistent results.

Under field conditions, understanding the dynamics of rhizobial strains and how they affect crops becomes even more difficult than under lab conditions. Lab-selected strains need to be tested under field conditions[100], and effects of inoculation are varied[95,100,101]. There is no shortage of studies, but there is a need to understand the varied effects of rhizobial inoculants. Meta-analyses such as recently performed for soybean[100] may prove crucial in obtaining this understanding, but the authors also indicated that simple and cheap methods for measuring nitrogen fixation need to be developed. Although an indirect measure of nitrogen fixation rates, using nodule size may be a possibility to explore. If adequate software is developed, obtaining complete distributions of nodule sizes on plants in the field could become as simple as taking pictures of excavated root systems. Data on nodule sizes may be used to guide what is limiting nitrogen fixation. For example, if most of the sampled plants have small nodules, but a few plants have a few very large nodules, this could indicate the presence of an effective yet not very competitive strain. If, on the other hand, most plants form a decent number of large nodules but yields remain low, this could indicate that the best available strain is not very effective at providing nitrogen under the particular conditions, or that something other than rhizobial nitrogen fixation limits yields. The method could also be used to screen populations of indigenous rhizobia, which is another aspect in need of attention[100]. It remains to be seen how informative data on nodule sizes can be,

but it seems possible that information obtained can be helpful in understanding how (inoculant) rhizobial strains may enhance yields.

6.4 General conclusions

In this thesis, the legume-rhizobia symbiosis was studied from two perspectives – evolutionary biology and agriculture – by integrating modelling and experimental approaches. The combination of two perspectives helped in addressing questions relevant for both. The methods developed because of the agricultural interest in competitive rhizobial strains could be used to address the question of partner choice posed by evolutionary biologists. The idea of sanctioning as proposed by evolutionary biologists was approached in a less theoretical manner as a question of optimal resource allocation, similar to the way agricultural researchers have long studied root growth in heterogeneous environments. This led to a more complete understanding of sanctioning, which in turn has implications for our understanding of using rhizobia to enhance yields. Modelling helped to formulate well-informed and testable hypotheses. Experiments supported some, but not all, model predictions, illustrating the importance of combining both approaches. Testing the model predictions relied on independent manipulation of several rhizobial characteristics. This could be achieved using mutant strains, overcoming previous experimental difficulties, and highlighting that evolutionary and applied questions in the field are unlikely to be answered without the solid foundation of lab-based groundwork. Looking forward, combining modelling and experimental approaches can be expected to further advance our understanding of legume-rhizobia symbioses. Undoubtedly, a deeper understanding will further enhance our appreciation for both legumes and rhizobia, and for how they together underpin our food production.

6.5 References

1. Young, J. P. & Johnston, A. W. The evolution of specificity in the legume-rhizobium symbiosis. *Trends Ecol. Evol.* **4,** 341–349 (1989).
2. Westhoek, A. *et al.* Policing the legume-Rhizobium symbiosis: a critical test of partner choice. *Sci. Rep.* 1–10 (2017).
3. Kiers, E. T. & Denison, R. F. Sanctions, cooperation, and the stability of plant-rhizosphere mutualisms. *Annu. Rev. Ecol. Evol. Syst.* **39,** 215–236 (2008).
4. Gordon, B. R. *et al.* Decoupled genomic elements and the evolution of partner quality in nitrogen-fixing rhizobia. *Ecol. Evol.* **6(5),** 1317–1327 (2016).
5. Kiers, E. T., Rousseau, R. A., West, S. A. & Denison, R. F. Host sanctions and the legume-rhizobium mutualism. *Nature* **425,** 78–81 (2003).
6. Oono, R., Anderson, C. G. & Denison, R. F. Failure to fix nitrogen by non-reproductive symbiotic rhizobia triggers host sanctions that reduce fitness of their reproductive clonemates. *Proc. R. Soc. B.* **278,** 2698–703 (2011).
7. Singleton, W. & Stockinger, K. R. Compensation against ineffective nodulation in soybean. *Crop Sci.* **23,** 69–72 (1983).
8. Parsons, R., Stanforth, A., Raven, J. a. & Sprent, J. I. Nodule growth and activity may be regulated by a feedback mechanism involving phloem nitrogen. *Plant, Cell Environ.* **16,** 125–136 (1993).
9. Lodwig, E., Hosie, A. & Bourdès, A. Amino-acid cycling drives nitrogen fixation in the legume–*Rhizobium* symbiosis. *Nature* **422,** (2003).
10. Fischinger, S. a., Drevon, J. J., Claassen, N. & Schulze, J. Nitrogen from senescing lower leaves of common bean is re-translocated to nodules and might be involved in a N-feedback regulation of nitrogen fixation. *J. Plant Physiol.* **163,** 987–995 (2006).
11. Prell, J. *et al.* Legumes regulate *Rhizobium* bacteroid development and persistence by the supply of branched-chain amino acids. *Proc. Natl. Acad. Sci.* **106,** 12477–12482 (2009).
12. Prell, J. *et al.* Role of symbiotic auxotrophy in the *Rhizobium*-legume symbioses. *PLoS*

One **5,** (2010).

13. Sulieman, S., Fischinger, S. A., Gresshoff, P. M. & Schulze, J. Asparagine as a major factor in the N-feedback regulation of N_2 fixation in *Medicago truncatula*. *Physiol. Plant.* **140,** 21–31 (2010).
14. Pate, J. S., Atkins, C. A., Hamel, K., Mcneil, D. L. & Layzell, David, B. Transport of organic solutes in phloem and xylem of a nodulated legume. *Plant Physiol.* **63,** 1082–1088 (1979).
15. Hardy, R. W. F., Holsten, R. D., Jackson, E. K. & Burns, R. C. The acetylene - ethylene assay for N_2 fixation: laboratory and field evaluation. *Plant Physiol.* **43,** 1185–1207 (1968).
16. Vance, C. P. & Heichel, G. H. Carbon in N_2 fixation: limitation or exquisite adaptation. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **42,** 373–390 (1991).
17. Cabeza, R. A. *et al.* Long-term non-invasive and continuous measurements of legume nodule activity. *Plant J.* **81,** 637–648 (2015).
18. Tegeder, M. & Rentsch, D. Uptake and partitioning of amino acids and peptides. *Mol. Plant* **3,** 997–1011 (2010).
19. Sprent, J. I. & James, E. K. Legume evolution: where do nodules and mycorrhizas fit in? *Plant Physiol.* **144,** 575–81 (2007).
20. Tegeder, M. Transporters involved in source to sink partitioning of amino acids and ureides: opportunities for crop improvement. *J. Exp. Bot.* **65,** 1865–1878 (2014).
21. Mortier, V., Holsters, M. & Goormachtig, S. Never too many? How legumes control nodule numbers. *Plant. Cell Environ.* **35,** 245–58 (2012).
22. Bauer, P., Ratet, P., Crespi, M. D., Schultze, M. & Kondorosi, A. Nod factors and cytokinins induce similar cortical cell division, amyloplast deposition and *MsEnod12A* expression patterns in alfalfa roots. *Plant J.* **10,** 91–105 (1996).
23. Gonzalez-Rizzo, S., Crespi, M. & Frugier, F. The *Medicago truncatula* CRE1 cytokinin receptor regulates lateral root development and early symbiotic interaction with *Sinorhizobium melilot*. *Plant Cell Online* **18,** 2680–2693 (2006).

24. Lohar, D. P. *et al.* Transcript analysis of early nodulation events in *Medicago truncatula*. *Plant Physiol.* **140,** 221–234 (2006).

25. Murray, J. D. *et al.* A cytokinin perception mutant colonized by *Rhizobium* in the absence of nodule organogenesis. *Science* **315,** 101–4 (2007).

26. Tirichine, L. *et al.* A gain-of-function mutation in a cytokinin receptor triggers spontaneous root nodule organogenesis. *Science* **2680,** 104–107 (2007).

27. Plet, J. *et al.* MtCRE1-dependent cytokinin signaling integrates bacterial and plant cues to coordinate symbiotic nodule organogenesis in *Medicago truncatula*. *Plant J.* **65,** 622–633 (2011).

28. Reid, D. E., Heckmann, A. B., Novak, O., Kelly, S. & Stougaard, J. CYTOKININ OXIDASE/DEHYDROGENASE3 maintains cytokinin homeostasis during root and nodule development in *Lotus japonicus*. *Plant Physiol.* **170,** pp.15.00650- (2015).

29. Cooper, J. B. & Long, S. R. Morphogenetic rescue of *Rhizobium meliloti* nodulation mutants by trans-zeatin secretion. *Plant Cell Online* **6,** 215–225 (1994).

30. Jung, J. K. & McCouch, S. Getting to the roots of it: genetic and hormonal control of root architecture. *Front Plant Sci* **4,** 186 (2013).

31. Mortier, V. *et al.* Search for nodulation-related *CLE* genes in the genome of *Glycine max*. *J. Exp. Bot.* **62,** 2571–2583 (2011).

32. Mortier, V., De Wever, E., Vuylsteke, M., Holsters, M. & Goormachtig, S. Nodule numbers are governed by interaction between CLE peptides and cytokinin signaling. *Plant J.* **70,** 367–376 (2012).

33. Soyano, T., Hirakawa, H., Sato, S., Hayashi, M. & Kawaguchi, M. NODULE INCEPTION creates a long-distance negative feedback loop involved in homeostatic regulation of nodule organ production. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 14607–14612 (2014).

34. Nishida, H., Handa, Y., Tanaka, S., Suzaki, T. & Kawaguchi, M. Expression of the *CLE-RS3* gene suppresses root nodulation in *Lotus japonicus*. *J. Plant Res.* **129,** 909–919 (2016).

35. Ferguson, B. J. *et al.* Molecular analysis of legume nodule development and autoregulation. *J. Integr. Plant Biol.* **52,** 61–76 (2010).

36. Reid, D. E., Ferguson, B. J. & Gresshoff, P. M. Inoculation- and nitrate-induced CLE peptides of soybean control NARK-dependent nodule formation. *Mol. Plant-Microbe Interact.* **24,** 606–618 (2011).

37. Okamoto, S., Shinohara, H., Mori, T., Matsubayashi, Y. & Kawaguchi, M. Root-derived CLE glycopeptides control nodulation by direct binding to HAR1 receptor kinase. *Nat. Commun.* **4,** 2191 (2013).

38. van Brussel, A. A. N., Tak, T., Boot, K. J. M. & Kijne, J. W. Autoregulation of root nodule formation: signals of both symbiotic partners studied in a split-root system of *Vicia sativa* subsp. *nigra*. *Mol. Plant. Microbe. Interact.* **15,** 341–349 (2002).

39. Kiba, T., Kudo, T., Kojima, M. & Sakakibara, H. Hormonal control of nitrogen acquisition: roles of auxin, abscisic acid, and cytokinin. *J. Exp. Bot.* **62,** 1399–1409 (2011).

40. Kamada-Nobusada, T., Makita, N., Kojima, M. & Sakakibara, H. Nitrogen-dependent regulation of de novo cytokinin biosynthesis in rice: the role of glutamine metabolism as an additional signal. *Plant Cell Physiol.* **54,** 1881–1893 (2013).

41. Okamoto, S. *et al.* Nod factor/nitrate-induced *CLE* genes that drive HAR1-mediated systemic regulation of nodulation. *Plant Cell Physiol.* **50,** 67–77 (2009).

42. Beveridge, C. A., Mathesius, U., Rose, R. J. & Gresshoff, P. M. Common regulatory themes in meristem development and whole-plant homeostasis. *Curr. Opin. Plant Biol.* **10,** 44–51 (2007).

43. Camp, R. H. M. O. Den *et al.* A phylogenetic strategy based on a legume-specific whole genome duplication yields symbiotic cytokinin type-A response regulators. *Plant Physiol.* **157,** 2013–2022 (2011).

44. Franssen, H. J. *et al.* Root developmental programs shape the *Medicago truncatula* nodule meristem. *Development* **142,** 2941–2950 (2015).

45. Dello Ioio, R. *et al.* A genetic framework for genetic control of cell division and

differentiation in the root meristem. *Science* **1380,** 1543–1547 (2008).

46. Werner, T. & Schmülling, T. Cytokinin action in plant development. *Curr. Opin. Plant Biol.* **12,** 527–538 (2009).

47. Mortier, V. *et al.* Role of *LONELY GUY* genes in indeterminate nodulation on *Medicago truncatula*. *New Phytol.* **202,** 582–593 (2014).

48. Debi, B. R., Taketa, S. & Ichii, M. Cytokinin inhibits lateral root initiation but stimulates lateral root elongation in rice (*Oryza sativa*). *J. Plant Physiol.* **162,** 507–515 (2005).

49. Coba de la Peña, T. *et al.* A salt stress-responsive cytokinin receptor homologue isolated from *Medicago sativa* nodules. *Planta* **227,** 769–779 (2008).

50. Vernié, T. *et al.* EFD Is an ERF transcription factor involved in the control of nodule number and differentiation in *Medicago truncatula*. *Plant Cell* **20,** 2696–2713 (2008).

51. Roux, B. *et al.* An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. *Plant J.* **77,** 817–837 (2014).

52. Badenoch-Jones, J., Rolfe, B. G. & Letham, D. S. Phytohormones, *Rhizobium* mutants, and nodulation in legumes. *Plant Physiol.* **74,** 239–246 (1984).

53. Badenoch-Jones, J., Parker, C. W. & Letham, D. S. Phytohormones, *Rhizobium* mutants, and nodulation in legumes. VII. Identification and quantification of cytokinins in effective and ineffective pea root nodules using radioimmunoassay. *J. Plant Growth Regul.* **6,** 97–111 (1987).

54. Temple, S. J., Vance, C. P. & Gantt, S. J. Glutamate synthase and nitrogen assimilation. *Trends Plant Sci.* **3,** 6 (1998).

55. Miao, G. H., Hirel, B., Marsolier, M. C., Ridge, R. W. & Verma, D. P. S. Ammonia-regulated expression of a soybean gene encoding cytosolic glutamine synthetase in transgenic *Lotus corniculatus*. *Plant Cell* **3,** 11–22 (1991).

56. Schoenbeck, M. A. *et al.* Decreased NADH glutamate synthase activity in nodules and flowers of alfalfa (*Medicago sativa* L.) transformed with an antisense glutamate synthase transgene. *J Exp Bot* **51,** 29–39 (2000).

57. Lea, P. J. & Miflin, B. J. Glutamate synthase and the synthesis of glutamate in plants. *Plant Physiol. Biochem.* **41,** 555–564 (2003).

58. Andrews, M., Raven, J. A. & Lea, P. J. Do plants need nitrate? The mechanisms by which nitrogen form affects plants. *Ann. Appl. Biol.* **163,** 174–199 (2013).

59. Chen, Y. *et al.* Knockdown of *LjIPT3* influences nodule development in *Lotus japonicus*. *Plant Cell Physiol.* **55,** 183–193 (2014).

60. Guan, D. *et al.* Rhizobial infection is associated with the development of peripheral vasculature in nodules of *Medicago truncatula*. *Plant Physiol.* **162,** 107–115 (2013).

61. Mok, D. W. S. & Mok, M. C. Cytokinin metabolism and action. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **52,** 89–118 (2001).

62. Mathesius, U., Charon, C., Rolfe, B. G., Kondorosi, A. & Crespi, M. Temporal and spatial order of events during the induction of cortical cell divisions in white clover by *Rhizobium leguminosarum* bv. *trifolii* inoculation or localized cytokinin addition. *Mol. Plant. Microbe. Interact.* **13,** 617–628 (2000).

63. Suzaki, T., Ito, M. & Kawaguchi, M. Genetic basis of cytokinin and auxin functions during root nodule development. *Plant Cell* **4,** 42 (2013).

64. Ferguson, B. J., Wiebe, E. M. K., Emery, R. J. N. & Guinel, F. C. Cytokinin accumulation and an altered ethylene response mediate the pleiotropic phenotype of the pea nodulation mutant R50 (*sym16*). *Can. J. Bot.* **83,** 989–1000 (2005).

65. Turner, M. *et al.* Ectopic expression of miR160 results in auxin hypersensitivity, cytokinin hyposensitivity, and inhibition of symbiotic nodule development in soybean. *Plant Physiol.* **162,** 2042–55 (2013).

66. Nizampatnam, N. R., Schreier, S. J., Damodaran, S., Adhikari, S. & Subramanian, S. microRNA160 dictates stage-specific auxin and cytokinin sensitivities and directs soybean nodule development. *Plant J.* **84,** 140–153 (2015).

67. Calderini, O. *et al.* Collection of mutants for functional genomics in the legume *Medicago truncatula*. *Plant Genet. Resour.* **9,** 174–176 (2011).

68. Behm, J. E., Geurts, R. & Kiers, E. T. *Parasponia*: a novel system for studying mutualism

stability. *Trends Plant Sci.* **19,** 757–763 (2014).

69. Denison, R. F. Legume sanctions and the evolution of symbiotic cooperation by rhizobia. *Am. Nat.* **156,** 567–576 (2000).

70. Heath, K. & Tiffin, P. Stabilizing mechanisms in a legume–rhizobium mutualism. *Evolution (N. Y).* **63,** 652–662 (2009).

71. Dowling, D. N. & Broughton, W. J. Competition for nodulation of legumes. *Annu. Rev. Microbiol.* **40,** 131–157 (1986).

72. Souza, V. *et al.* Ethnomicrobiology: do agricultural practices modify the population structure of the nitrogen fixing bacteria *Rhizobium etli* biovar *phaseoli*? *J. Ethnobiol.* **17,** 249–266 (1997).

73. Hagen, M. J. & Hamrick, J. L. A hierarchical analysis of population genetic structure in *Rhizobium leguminosarum* bv. *trifolii*. *Mol. Ecol.* **5,** 177–186 (1996).

74. Silva, C., Eguiarte, L. E. & Souza, V. Reticulated and epidemic population genetic structure of *Rhizobium etli* biovar *phaseoli* in a traditionally managed locality in Mexico. *Mol. Ecol.* **8,** 277–287 (1999).

75. Gubry-Rangin, C., Garcia, M. & Béna, G. Partner choice in *Medicago truncatula-Sinorhizobium* symbiosis. *Proc. Biol. Sci.* **277,** 1947–51 (2010).

76. Regus, J., Gano, K. A., Hollowell, A. C. & Sachs, J. L. Efficiency of partner choice and sanctions in *Lotus* is not altered by nitrogen fertilization. *Proc. R. Soc. B.* **281,** 20132587 (2014).

77. Kiers, T. E., Rousseau, R. A. & Denison, R. F. Measured sanctions: legume hosts detect quantitative variation in rhizobium cooperation and punish accordingly. *Evol. Ecol. Res.* **8,** 1077–1086 (2006).

78. Bergersen, F. J. Some Australian studies relating to the long-term effects of the inoculation of legume seeds. *Plant Soil* **32,** 727–736 (1970).

79. Brockwell, J., Bottomley, P. J. & Thies, J. E. Manipulation of rhizobia microflora for improving legume productivity and soil fertility: a critical assessment. *Plant Soil* **174,** 143–180 (1995).

80. Sprent, J. I. & James, E. K. Legume-rhizobial symbiosis: an anorexic model? *New Phytol.* **179,** 3–5 (2008).

81. Moyano, G., Marco, D., Knopoff, D., Torres, G. & Turner, C. Explaining coexistence of nitrogen fixing and non-fixing rhizobia in legume-rhizobia mutualism using mathematical modeling. *Math. Biosci.* **292,** 30–35 (2017).

82. Denison, R. F. & Kiers, E. T. Life histories of symbiotic rhizobia and mycorrhizal fungi. *Curr. Biol.* **21,** R775-85 (2011).

83. Steidinger, B. S. & Bever, J. D. Host discrimination in modular mutualisms: a theoretical framework for meta-populations of mutualists and exploiters. *Proc. R. Soc. B.* **283,** 20152428 (2016).

84. Bever, J. D. Preferential allocation, physio-evolutionary feedbacks, and the stability and environmental patterns of mutualism between plants and their root symbionts. *New Phytol.* **205,** 1503–1514 (2015).

85. Murphy, P. J. *et al.* Genes for the catabolism and synthesis of an opine-like compound in *Rhizobium meliloti* are closely linked and on the Sym plasmid. *Proc. Natl. Acad. Sci. U. S. A.* **84,** 493–7 (1987).

86. Gordon, D. M. *et al.* An experimental test of the rhizopine concept in *Rhizobium meliloti*. **62,** 3991–3996 (1996).

87. Krief, A. *et al.* First synthesis of 3-*O*-methyl-scyllo-inosamine, a natural product which favors the Rhizobium-Leguminosae symbiosis. *Tetrahedron Lett.* **45,** 1461–1463 (2004).

88. Sprent, J. I., Ardley, J. & James, E. K. Biogeography of nodulated legumes and their nitrogen-fixing symbionts. *New Phytol.* (2017).

89. Lavin, M., Herendeen, P. & Wojciechowski, M. Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the Tertiary. *Syst. Biol.* **54,** 575–594 (2005).

90. Osborne, B. Holy alliances? *New Phytol.* **175,** 602–605 (2007).

91. Kiers, E. T. *et al.* Reciprocal rewards stabilize cooperation in the mycorrhizal symbiosis. *Science* **333,** 880–882 (2011).

92. Frederickson, M. E. Rethinking mutualism stability: cheaters and the evolution of sanctions. *Q. Rev. Biol.* **87,** 3–18 (2014).

93. Perret, X., Staehelin, C. & Broughton, W. J. Molecular basis of symbiotic promiscuity. *Microbiol. Mol. Biol. Rev.* **64,** 180–201 (2000).

94. Slattery, J. & Pearce, D. Development of elite inoculant *Rhizobium* strains in Southeastern Australia. *ACIAR Proc. 109e* (2002).

95. Yates, R. J., Howieson, J. G., Reeve, W. G. & O'Hara, G. W. A re-appraisal of the biology and terminology describing rhizobial strain success in nodule occupancy of legumes in agriculture. *Plant Soil* **348,** 255–267 (2011).

96. Tena, W., Wolde-meskel, E. & Walley, F. Symbiotic efficiency of native and exotic *Rhizobium* strains nodulating lentil (*Lens culinaris* Medik .) in soils of Southern Ethiopia. *Agronomy* **6,** 1–10 (2016).

97. Chibeba, A. M., Kyei-Boahen, S., Guimaraes, M. de F., Nogueira, M. A. & Hungria, M. Isolation, characterization and selection of indigenous *Bradyrhizobium* strains with outstanding symbiotic performance to increase soybean yields in Mozambique. *Agric. Ecosyst. Environ.* **246,** 291–305 (2017).

98. Marinho, R. de C. N. *et al.* Symbiotic and agronomic efficiency of new cowpea rhizobia from Brazilian Semi-Arid. *Bragantia* **76,** 273–281 (2017).

99. Kumar, V., Rawat, A. K. & Rao, D. L. N. Population ecology of soybean-rhizobia in diverse crop rotations in Central India. *Agric. Ecosyst. Environ.* **240,** 261–268 (2017).

100. Thilakarathna, M. S. & Raizada, M. N. A meta-analysis of the effectiveness of diverse rhizobia inoculants on soybean traits under field conditions. *Soil Biol. Biochem.* **105,** 177–196 (2017).

101. Slattery, J. F., Pearce, D. J. & Slattery, W. J. Effects of resident rhizobial communities and soil type on the effective nodulation of pulse legumes. *Soil Biol. Biochem.* **36,** 1339–1346 (2004).

102. Mpepereki, S., Javaheri, F., Davis, P. & Giller, K. E. Soyabeans and sustainable agriculture. *F. Crop. Res.* **65,** 137–149 (2000).

103. Kremer, R. J. & Peterson, H. L. Isolation, selection and evaluation of rhizobium under controlled conditions. *Commun. Soil Sci. Plant Anal.* **13,** (1982).

104. Yates, R. J. *et al.* Host-strain mediated selection for an effective nitrogen-fixing symbiosis between *Trifolium* spp. and *Rhizobium leguminosarum* biovar *trifolii*. *Soil Biol. Biochem.* **40,** 822–833 (2008).

105. Brockwell, J. & Bottomley, P. J. Recent advances in inoculant technology and prospects for the future. *Soil Biol. Biochem.* **27,** 683–697 (1995).

106. Russakovsky, O. *et al.* ImageNet large scale visual recognition challenge. *Int. J. Comput. Vis.* **115,** 211–252 (2015).

Appendix A. Published manuscript Chapter 3

Overview

This appendix contains the published manuscript of chapter 3, which was published as:

Policing the legume-Rhizobium symbiosis: a critical test of partner choice.

Annet Westhoek, Elsa Field, Finn Rehling, Geraldine Mulley, Isabel Webb, Philip S. Poole & Lindsay A. Turnbull.

Scientific Reports 7, Article number: 1419 (2017). doi:10.1038/s41598-017-01634-2

SCIENTIFIC REPORTS

OPEN

Policing the legume-Rhizobium symbiosis: a critical test of partner choice

Received: 30 November 2016
Accepted: 3 April 2017
Published online: 03 May 2017

Annet Westhoek[1,2], Elsa Field[1], Finn Rehling[1,3], Geraldine Mulley[4], Isabel Webb[1], Philip S. Poole[1] & Lindsay A. Turnbull[1]

In legume-*Rhizobium* symbioses, specialised soil bacteria fix atmospheric nitrogen in return for carbon. However, ineffective strains can arise, making discrimination essential. Discrimination can occur via partner choice, where legumes prevent ineffective strains from entering, or via sanctioning, where plants provide fewer resources. Several studies have inferred that legumes exercise partner choice, but the rhizobia compared were not otherwise isogenic. To test when and how plants discriminate ineffective strains we developed sets of fixing and non-fixing strains that differed only in the expression of *nifH* – essential for nitrogen fixation – and could be visualised using marker genes. We show that the plant is unable to select against the non-fixing strain at the point of entry, but that non-fixing nodules are sanctioned. We also used the technique to characterise mixed nodules (containing both a fixing and a non-fixing strain), whose frequency could be predicted using a simple diffusion model. We discuss that sanctioning is likely to evolve in preference to partner choice in any symbiosis where partner quality cannot be adequately assessed until goods or services are actively exchanged.

Across the globe primary productivity is nitrogen limited[1]. This limitation has been overcome for plants in the family Fabaceae (commonly known as legumes) through a mutualistic association with nitrogen-fixing bacteria collectively called rhizobia[2]. The nitrogen provided through this symbiosis makes legumes rich in protein and important crops in human diets[3]. But, as ineffective strains will inevitably arise through mutation, there is the potential for the relationship to break down. Ineffective strains are known to be common, at least in some situations, which for agricultural legumes means poor yields and reduced nutritional quality[4,5].

Theory predicts that ineffective strains could be successful within legume – *Rhizobium* symbioses for two reasons[6,7]. First, rhizobia are not transmitted directly from parent plant to offspring. Instead, plants acquire rhizobia from the soil through an intricate signalling process in which bacteria enter specialized root nodules, where they fix nitrogen in return for plant-derived carbon[8]. This horizontal transmission means that rhizobial fitness is not perfectly aligned with the fitness of the host plant[7]. Second, although each nodule is usually occupied by the clonal descendants of a single *Rhizobium*[9,10], a plant is usually infected by multiple rhizobial strains[11,12]. Thus, a non-fixing strain can potentially thrive by taking plant resources while leaving the costly process of nitrogen fixation to others[7,13]. To prevent losing resources to ineffective rhizobial strains that provide little or no nitrogen, legumes have two options: partner choice or sanctions[7].

Partner choice is usually defined as any mechanism that allows detection of suitable partners *before* a mutualistic relationship is established[7,13–15], while sanctioning is a mechanism to discriminate against low-quality partners once the relationship is underway[6,7,13–16] (although confusingly 'partner choice' has also been used to describe a broader concept which includes sanctioning)[13,17,18]. Partner choice might seem to be the more attractive option as resources are not wasted setting up a relationship that is doomed to fail. But crucially, effective partner choice requires accurate assessment of the quality of partners in advance[13]. This is likely to be problematical for any symbiosis in which key traits are not manifested prior to the relationship being established. For example, in the legume-*Rhizobium* symbiosis, nitrogen fixation does not begin until the bacteria have entered the roots and nodule formation is sufficiently advanced for rhizobia to have differentiated into nitrogen-fixing bacteroids[8].

[1]Department of Plant Sciences, University of Oxford, Oxford, OX1 3RB, UK. [2]Systems Biology Doctoral Training Centre, University of Oxford, Oxford, OX1 3RQ, UK. [3]Department of Ecology, Philipps-University Marburg, Marburg, D-35043, Germany. [4]School of Biological Sciences, University of Reading, Reading, RG6 6AJ, UK. Correspondence and requests for materials should be addressed to P.S.P. (email: philip.poole@plants.ox.ac.uk) or L.A.T. (email: lindsay.turnbull@plants.ox.ac.uk)

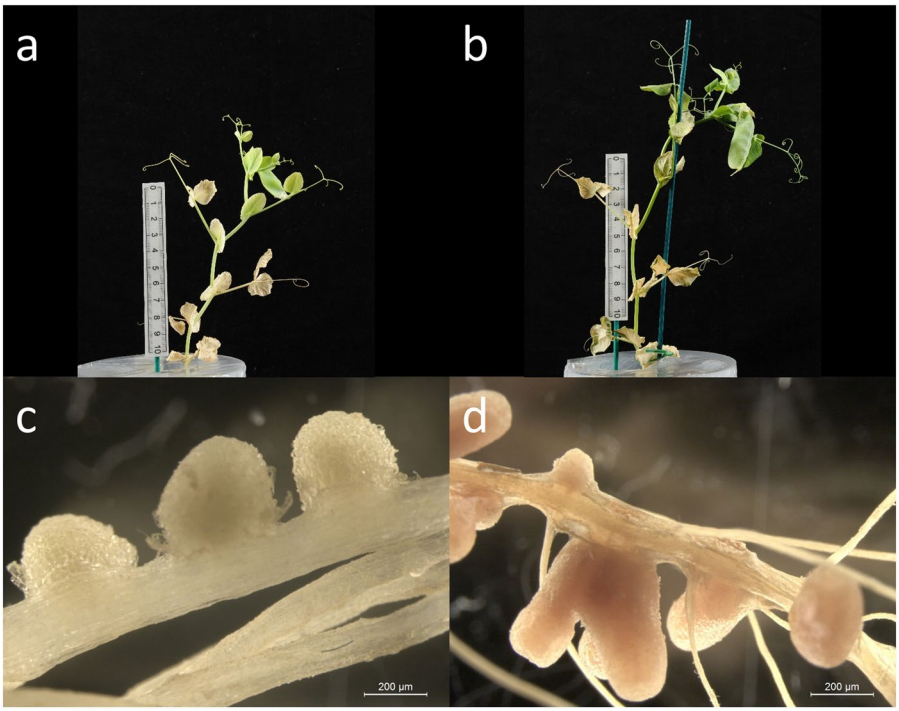


Figure 1. Comparison of the non-fixing strain RU3940, (**a**,**c**) and the otherwise isogenic fixing strain Rlv3841 (**b**,**d**). After five weeks of growth, plants inoculated with the non-fixing strain (**a**) had reduced biomass and produced fewer peas than those inoculated with the fixing strain (**b**). The non-fixing strain formed white (**c**), rather than pink (**d**), nodules, indicating a lack of leghaemoglobin. Scale bars indicate 200 µm.

Once nodules are established they can be sanctioned, if they prove to be ineffective, by cutting off their supply of carbon, oxygen or other nutrients and this has been demonstrated empirically using argon gas to force nodules to fix less nitrogen[19–21].

Despite the empirical support for sanctioning and its apparent advantages, there are nevertheless several studies that claim evidence for partner choice[7,14,15]. However, the interpretation of these studies is problematic because the tested strains are rarely isogenic – meaning that strains differ in several traits, and not just in how much nitrogen they provide. Most importantly, strains are likely to differ in their competitiveness in colonizing plant roots and forming nodules. A range of traits affects competitiveness: examples include motility[22], production of antibiotics[23] and the secretion of proteins and polysaccharides involved in biofilm formation and root attachment[24]. Such differences in competitiveness explain why poorly-fixing strains can also end up occupying a higher proportion of nodules – a problem that is often encountered when developing effective strains for use in agricultural settings[4,25,26]. Thus comparing the nodulation success of naturally occurring strains is difficult to interpret as a test of partner choice.

To test whether plants can directly assess the effectiveness of potential rhizobia prior to nodulation we created a non-fixing mutant from a fixing strain and compared their success in colonising pea plant nodules. There are several key genes involved in nitrogen fixation[8], any of which could undergo a mutation that would render the gene non-functional and hence transform the fixing into a non-fixing strain. We chose the *nifH* gene in *Rhizobium leguminosarum* bv. viciae (*Rlv*) 3841, and created a non-fixing mutant that was otherwise identical to its fixing parent strain. We then assessed when and how the plant discriminated between the two strains. To identify strains, they were marked with *gusA* or *celB* marker genes, rendering strains magenta or blue (respectively) following the application of a simple post-harvest staining protocol. Insertion of marker genes solves a secondary problem as it is usually extremely time-consuming to identify different strains using antibiotic markers and this limits the number of nodules that can be assessed.

One possible complication is that a non-fixing strain can potentially thrive via mixed nodules (where two different bacteria have entered and colonised). If a non-fixing strain can take advantage of mixed nodules to increase its fitness at the expense of the fixing strain, then this would provide a route by which non-fixing strains could increase in frequency. Currently little is known about the frequency at which mixed nodules occur, and the relative fitness of strains within mixed nodules. The staining protocol rendered mixed nodules easily visible, so we assessed the frequency of mixed nodules under different inoculation densities. We discuss mixed nodules, partner choice and sanctioning in the context of the evolutionary stability of the legume-*Rhizobium* symbiosis.

Results

In single inoculations, the non-fixing mutant strain reduced shoot mass and formed small white nodules, while the fixing parent strain formed larger pink nodules (indicating the presence of leghaemoglobin) (Fig. 1).

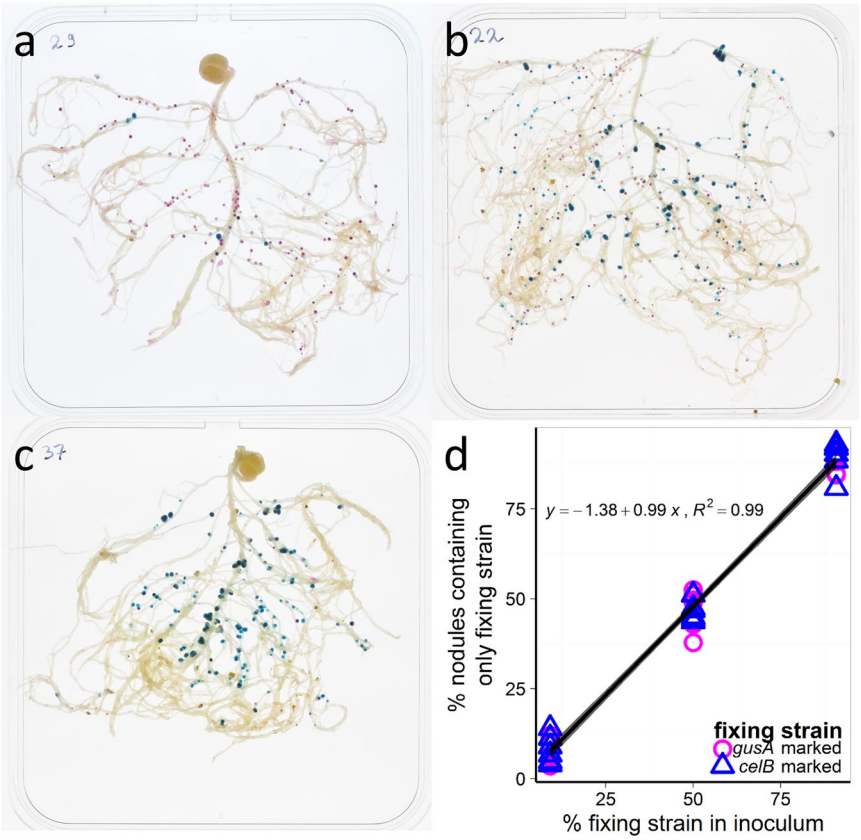


Figure 2. We found no evidence of partner choice. (**a**–**c**) show pea root systems with stained nodules following inoculation with different ratios of a fixing *celB* (blue) marked strain (3841 *celB*) and a non-fixing *gusA* (magenta) marked mutant strain (OPS0365): (**a**) 1:10 (fixing:non-fixing); (**b**) 1:1 (fixing:non-fixing); (**c**) 10:1 (fixing:non-fixing). (**d**) The percentage of nodules containing only the fixing strain exactly reflected the percentage of the fixing strain in the inoculum. The slope of the regression was 0.99 ± 0.02, 95% CI [0.95, 1.02] and this is not significantly different from one ($t_{39} = -0.835$, $p = 0.41$).

Partner choice. To test for partner choice, we varied the ratio of the fixing: non-fixing strains in the inoculum (1:10, 1:1 and 10:1) and counted the number of fixing and non-fixing nodules after three weeks. In the absence of partner choice, we expect the percentage of nodules containing the fixing strain to reflect the percentage in the inoculum. In other words, we expect a 1:1 relationship between the percentage of fixing nodules and the percentage of the fixing strain in the inoculum (a slope of 1.0 through the origin). We found that the percentage of nodules containing only the fixing strain exactly reflected the percentage of the fixing strain in the inoculum (Fig. 2). The slope of the regression line was 0.99 ± 0.02 (95% CI [0.95, 1.02]) and this is not significantly different from 1.0 ($t_{39} = -0.835$, $p = 0.41$). The intercept was not significantly different from zero ($t_{39} = -1.338$, $p = 0.189$). Whether the fixing strain was marked with *gusA* or *celB* had no impact on the proportion of fixing nodules formed ($t_{37} = -0.177$, $p = 0.861$) or on the interaction with the inoculum ratio ($t_{37} = -0.189$, $p = 0.851$). Thus, pea plants do not discriminate between fixing and non-fixing strains prior to nodule formation and are therefore unable to exercise partner choice against the mutant strain.

Nodule size. In the absence of partner choice, sanctions remain the only option for discriminating against non-fixing strains. If pea plants are able to sanction and differentially allocate resources on the basis of the nitrogen provided, we expect nodules containing a fixing strain to be larger than nodules containing a non-fixing strain. We therefore measured the sizes of a sample of nodules colonised by the different strains. Nodules containing the nitrogen-fixing strain were significantly larger than nodules containing the non-fixing strain (paired t-test, $t_{13} = 7.7176$, $p = 3.307 \times 10^{-6}$), indicating that pea plants preferentially allocated resources to fixing nodules, supporting sanctioning. Fixing nodules had an average area of $1.20 \pm 0.07\,mm^2$ (mean ± s.e.), and non-fixing nodules were $0.68\,mm^2$ (95% CI [0.49; 0.87]) smaller – less than half of the area of fixing nodules (Fig. 3).

Total number of nodules. Each pea plant formed 183 ± 12 nodules (mean ± s.e.) on average. The total number of nodules (Fig. 4) decreased as the percentage of nitrogen-fixing nodules increased (slope $= -0.88 \pm 0.34$, $t_{39} = -2.565$, $p = 0.0143$), probably because the nodulation process is inhibited once fixing nodules have been successfully established. A ten-fold increase in the percentage of nitrogen-fixing nodules resulted in 72 fewer nodules in total (95% CI [9, 135]) – a decrease of about 30%.

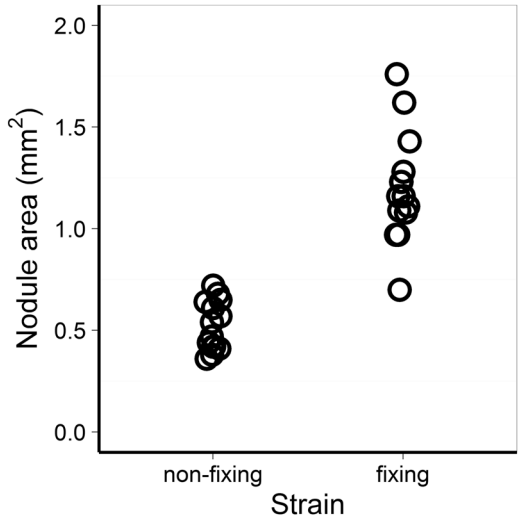


Figure 3. Legumes sanction non-fixing nodules. Nodules containing the fixing strain were significantly larger than nodules containing the fixing strain (paired t-test, $t_{13} = 7.7176$, $p = 3.307 \times 10^{-6}$), indicating that pea plants preferentially allocated resources to fixing nodules.

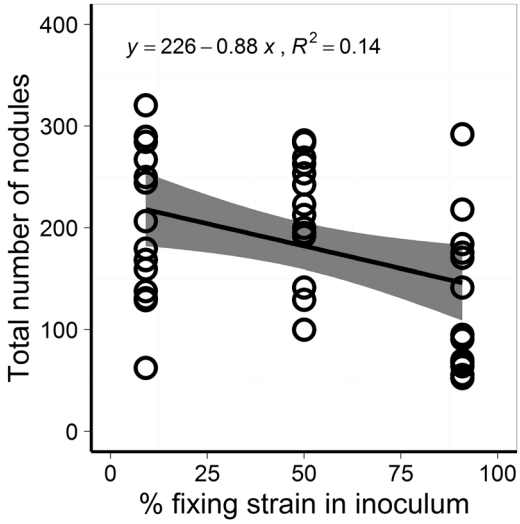


Figure 4. Total nodule number. The total number of nodules decreased as the percentage of fixing nodules increased (slope $= -0.88 \pm 0.34$, $t_{39} = -2.565$, $p = 0.0143$). Regression with 95% confidence interval shown.

Shoot mass. After three weeks of growth we found no differences in shoot dry mass among treatments, including water controls, ($F_{3,44} = 2.1921$, $p = 0.1024$), presumably because the pea seeds contain large reserves of nitrogen. However, after five weeks of growth, independently grown plants inoculated with non-fixing RU3940 weighed 0.32 ± 0.04 grams (mean $\pm$ s.e.) while plants inoculated with the fixing strain Rlv3841 weighed 0.68 ± 0.04 grams, a significant difference of 0.37 grams (95% CI [0.23, 0.51]).

Frequency of mixed nodules. The staining technique allowed clear visualization of mixed nodules, which occasionally appeared to result from multiple independent infections (Fig. 5A), but normally consisted of two – a fixing and a non-fixing strain (Fig. 5B,C). Mixed nodules occurred at an average frequency of $2.02\% \pm 0.35\%$ (mean $\pm$ s.e.), and the frequency depended on the inoculum ratio (Fig. 5D). According to a simple diffusion model (in which we assume that the two rhizobial strains in the soil are well mixed) the predicted frequency of mixed nodules (F) is simply given by $F = pq\varepsilon$, where p is the frequency of the fixing strain, q is the frequency of the non-fixing strain and ε is the unknown probability with which two bacteria simultaneously enter the same nodule; hence: $F \propto pq$. We regressed F against pq ('encounter rate', Fig. 5D) and found that the frequency of mixed nodules indeed increased in proportion to the increase in encounter rate: a $\pm$ three-fold increase in encounter rate increased the odds of a mixed infection 3.5 times (95% CI [2.2, 5.9], $n = 41$). In an additional experiment, we varied the total inoculation density, as the diffusion model also predicts that mixed nodules occur more often at

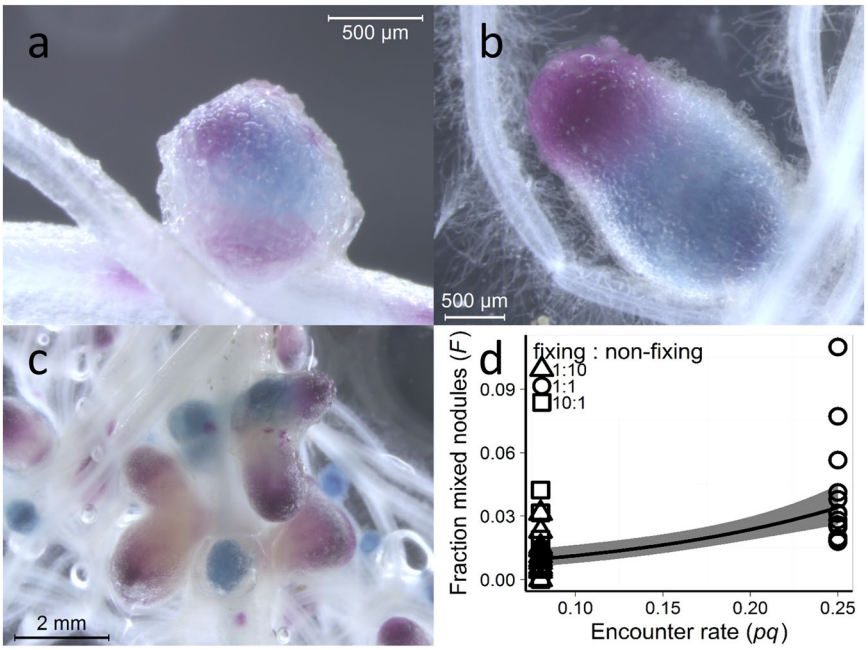


Figure 5. Mixed nodules. (**a**) Microscopic image of a mixed nodule on a pea root, which seems to have been infected by more than two rhizobia. The nitrogen-fixing strain is *celB* (blue) marked (3841 *celB*), and the non-fixing mutant strain is *gusA* (magenta) marked (OPS0365). (**b**) Microscopic image of a mixed nodule on a pea root infected by two rhizobia. The *gusA* marked nitrogen-fixing strain (3841 *gusA*) is magenta, and the *celB* marked non-fixing mutant strain (OPS0366) is blue. (**c**) Nodules, including mixed nodules, on a pea root. The *gusA* marked nitrogen-fixing strain (3841 *gusA*) is magenta, and the *celB* marked non-fixing mutant strain (OPS0366) is blue. (**d**) Mixed nodules occur more often when the probability that different strains encounter each other during nodule formation is higher ($t_{39} = 4.913$, $p = 1.65 \times 10^{-5}$). Generalized linear model with quasibinomial distribution and logit link function shown with 95% confidence interval.

higher densities of rhizobia. As predicted, we found that the percentage of mixed nodules increased with inoculation density (slope $= 1.87 \pm 0.41$, $t_{18} = 4.545$, $p = 0.000251$). A 10-fold increase in inoculation density, increased the percentage of mixed nodules by almost 2%.

Fitness of rhizobial strains within mixed nodules. To assess the relative fitness of fixing versus non-fixing strains within mixed nodules, we measured the area occupied by each strain as indicated by the area stained blue versus magenta (or vice versa). We found that the percentage of nodule area occupied by the fixing strain ranged from 1.1 to 99%. Unfortunately, although the staining technique is valuable in visualizing mixed nodules, we conclude that it cannot be reliably used to assess the relative fitness of different strains within mixed nodules. The percentage of the nodule occupied by the fixing strain depended highly on the marker used to identify the fixing strain ($t_{218} = 14.49$, $p < 2 \times 10^{-16}$). If the fixing strain was marked with *gusA*, it appeared to occupy 31% of the nodule (95% CI [27.5; 33.6%]), but if the fixing strain was marked with *celB*, it appeared to occupy 75% of the nodule (95% CI [68.7; 80.0%].

Discussion

We created a non-fixing but otherwise isogenic mutant to mimic a process that might occur in nature, where a mutation arises in a fixing rhizobial strain, rendering it ineffective. We found that pea plants could not discriminate between these fixing and non-fixing strains prior to nodule formation. Pea plants were therefore unable to detect whether the potential partner was effective at fixing nitrogen and could not prevent the formation of non-fixing nodules. Our results are supported by an early study using similar isogenic strains[27], but this study was severely limited in sample size and has therefore been overlooked.

Our results indicate that partner choice is not a robust mechanism against ineffective strains as pea plants were unable to prevent non-fixing strains from entering. It could be argued that legumes may use genes other than *nifH* to assess the nitrogen fixation capacity of fixing strains before nodule formation; however we believe that this is unlikely. A mutation rendering a strain less effective can arise in any gene and effective partner choice would then require a mutation in the plant genome to detect this change. If the new mutation stops the cheat from entering, then it will spread through the plant population; however, given that rhizobial generation times are much faster than host plant generation times[28], it seems that the host plants will be locked in an evolutionary arms race that they are doomed to lose; hence partner choice seems to be an ineffective way to stabilise the mutualism in the long term[21]. Furthermore, partner choice is susceptible to dishonest signals[13].

In contrast, we found that pea plants did discriminate against ineffective strains via sanctioning, which has been previously reported using argon gas to replace atmospheric nitrogen[19–21]. In our experiment, nodules

containing the non-fixing strain were roughly half the size of fixing nodules, indicating reduced plant resources. In contrast to the case of partner choice, sanctioning can stabilise the mutualism in the long-term. If a mutation arises that allows a plant to detect and sanction a partner that is not delivering the goods it would be effective against a wide variety of future ineffective strains. Thus, sanctioning allows an instantaneous response to ineffective strains and does not require specific recognition genes or rely on honest signals. Sanctioning is therefore a more robust[13] mechanism against ineffective strains and can provide long-term stability to legume-*Rhizobium* mutualisms[16, 29].

However, any discussion of sanctioning should take into account both plant and rhizobial fitness. Sanctioning can only be selected for when it saves plant resources and thus increases plant fitness. The reduced nodule size that we and others[20, 21] have seen indicates that plants allocate fewer resources to non-fixing nodules. Whether this reduced resource allocation also reduces rhizobial fitness and thus stabilizes the mutualism on evolutionary time-scales is more difficult to establish. This may depend on whether or not the nitrogen-fixing bacteroids can still reproduce (usually determinate nodules), or are terminally differentiated and unable to reproduce (indeterminate nodules)[30]. In studies using argon gas, reduced rhizobial fitness has been shown in both determinate[19, 20] and indeterminate[21] non-fixing nodules. However, in a study using a non-fixing isogenic strain, the fitness of non-fixing rhizobia was not reduced in determinate soybean nodules up to five weeks old[31]. Because effects on rhizobial viability may emerge later in the sanctioning process, perhaps the best test would be a multigenerational experiment, where in the first generation plants are inoculated with both fixing and non-fixing strains, and new plants are then repeatedly grown in the same soil for several generations to see how quickly the non-fixing strain is eliminated. It would be of special interest to perform such an experiment with both indeterminate and determinate species.

Evidence from other mutualisms suggests that whether partner choice evolves in preference to sanctioning critically depends on how well partners can assess quality prior to establishment of the mutualistic relationship including the potential for dishonest signalling. For example, clients of the cleaner fish *Labroides dimidiatus* have evolved partner choice to counter cheating by individuals that take healthy tissue while removing parasites[32]. Partner choice is highly effective in this mutualism because the quality of service is known from previous experience and there are repeated interactions between individuals. In contrast, sanctioning has evolved in mutualisms between yucca moths and fig wasps and their respective plant hosts. In both cases the insects deposit seed-eating larvae in the flowers of host plants in return for pollination. In these mutualisms, plants cannot prevent eggs being laid nor assess partner quality, hence sanctioning has evolved: flowers containing too many eggs[33], or too little pollen[33, 34] are selectively aborted.

While our study shows that partner choice is not a robust mechanism to exclude ineffective strains, legumes do not form symbioses with all potential rhizobial strains. Instead, an extensive signalling process[35–37] between legumes and their rhizobial partners can impose a high degree of selectivity on the relationship[36, 38, 39], although the degree of selectivity varies greatly among hosts[39]. There are two explanations for this selectivity that are commonly proposed and are not mutually exclusive. First, specificity may arise in order to prevent the entry of pathogenic bacteria which utilise similar signalling pathways to gain access to host roots[38, 40, 41]. Second, by fine-tuning signalling pathways to target rhizobia that are particularly effective for a specific host, legume species might achieve greater nitrogen-fixation efficiency[42, 43]. This is likely to be true if host environments are sufficiently different that specialization by rhizobia is selected for. Support for specialization comes from the observation that a single rhizobial strain can vary greatly in its effectiveness among hosts[42]. This type of co-evolutionary process is separate from the need to avoid non-fixing rhizobia, which can arise by mutation at any time, in any strain, even those that are usually highly effective. That these two processes are indeed separate is supported by the fact that the genes involved in nitrogen fixation (*nif* and *fix* genes) are only expressed once the symbiosis has been established[44] and are different from the signalling genes involved in infection (*nod* genes)[45]. Currently, genomic analyses are shedding more light on the selective pressures affecting both legume and rhizobial genes[46–48].

Sanctioning is a robust mechanism against ineffective strains, but requires hosts to monitor partner quality and provide resources accordingly. Currently, little is known about the exact mechanism behind sanctioning in legume-*Rhizobium* symbioses, and whether it only takes place at the nodule level, or also occurs within nodules[13, 16]. If sanctioning takes place at the nodule level, mixed nodules could be a way for ineffective strains to avoid sanctions[6, 13, 49]. Indeed certain endosymbionts, even those belonging to different genera and lacking any genes for nitrogen fixation, have been shown to co-infect nodules by "piggybacking" on the genuine symbionts as they infect the root hairs[50]. Whether mixed nodules allow ineffective strains to persist depends on the frequency of mixed nodules, and on the relative fitness of fixing and non-fixing strains within mixed nodules. Estimates of the frequency of mixed nodules in the literature range from 2% to 74%[6, 51]. Our findings at least partly explain this variability as the frequencies we found could be adequately represented by a simple diffusion model, which predicts that more mixed nodules are expected: (1) at high rhizobial densities; (2) when the proportions of different strains are similar; and (3) when rhizobia diffuse more easily, which might occur, for example, under wet conditions. Although the staining technique is valuable in identifying strains and characterising mixed nodules, it could not be used reliably to assess fitness of rhizobial strains within mixed nodules. Further work on the mechanism of sanctioning, how it is affected by external conditions such as soil nitrogen, and how it affects rhizobial fitness will help illuminate how the legume-*Rhizobium* mutualism has persisted for much longer than humans have been around to reap its benefits.

Methods

Bacterial strains and culture conditions. We used *Rhizobium leguminosarum* bv. viciae (*Rlv*) 3841 strains labelled at identical positions in the chromosome with either a *gusA* or a *celB* marker gene[52] (Table 1). Non-fixing mutants of these strains were made by replacing the *nifH* gene with a *nifH* gene disrupted by a spectinomycin resistance cassette (*nifH::ΩSpc*). *Rhizobium leguminosarum* only has a single copy of the *nifH* gene[53],

Strain, plasmid, or primer	Genotype or sequence	Reference of source
Strains		
Rlv3841	St^r derivative of *R. leguminosarum* bv. *viciae* strain 300	61
RU3940	Rlv3841 *nifH::Ω* Sp^r	This work
3841 *gusA*	Rlv3841 with *gusA* marker	52
3841 *celB*	Rlv3841 with *celB* marker	52
OPS0365	3841 *gusA nifH::Ω* Sp^r	This work
OPS0366	3841 *celB nifH::Ω* Sp^r	This work
Plasmids		
pRU1907	*nifH* cloned in pCR2.1	This work
pHP45Ω-Sp	Plasmid containing Ω spec cassette, Sp^r	62
pJQ200SK	pACYC derivative, P15A origin of replication, Gm^r	63
pRU1908	*nifH::Ω*, Sp^r in pJQ200SK	This work
Primers		
pOTforward	CGGTTTACAAGCATAAAGC	
p1002	TTCCTCCATGTGCCTGGAGA	
p1118	GGTTCTTCGGAGTTTCTAT	
oxp0460	GCTTGATCATCGCCGGAAAC	
oxp0461	TGTCACCGCCGAAAACGATG	

Table 1. Strains, plasmids and primers.

which is essential for nitrogen fixation[54]. First, an unmarked non-fixing strain was made, then the non-functional *nifH* gene was transduced into the marked fixing strains. To make the unmarked non-fixing strain, the *nifH* gene was amplified by PCR (2,660 bp product) using primers p950 and p951 that contain a SacI and SpeI restriction site and cloned into pCR2.1 in Dam- *E. coli* SCS110 (Stratagene) (pRU1907). A *Ω* spectinomycin cassette from pHP45Ω-Sp was cloned between two NruI sites within *nifH* resulting in a 166 bp deletion, and the *nifH::ΩSpc* fragment was cloned into the suicide vector pJQ200SK (pRU1908). Plasmid pRU1908 was conjugated into Rlv3841, and cells were plated on TY agar with gentamicin (20 μg ml^{-1}) and then on AMS agar supplemented with 10% sucrose, 10 mM NH_4Cl, and spectinomycin (100 μg ml^{-1}) to select for gene replacement. Strain RU3940 was found to contain the correct *nifH::ΩSpc* mutation by PCR mapping using pOT forward with either p1002 or p1118 in separate reactions. To confirm differences in the nitrogen-fixing capacity of the two strains, we performed acetylene reductions on plants inoculated with either RU3940 or Rlv3841. RU3940 does not fix nitrogen (0.05 ± 0.02 (mean ± s.e, $n = 3$) μmol ethylene per plant per hour, which is not different from water control plants ($t_6 = -0.025$, $p = 0.98$)), whereas the parent strain Rlv3841 does fix nitrogen (2.58 ± 0.33 (mean ± s.e, $n = 3$) μmol ethylene per plant per hour).

The *nifH::ΩSpc* cassette from non-fixing strain RU3940 was transduced into the *gusA* or *celB* marked strains using phage RL38[55], yielding non-fixing *gusA* (OPS0365) and *celB* (OPS0366) marked strains which are otherwise isogenic to their fixing *gusA* and *celB* marked parent strains. Correct insertion of the *nifH::ΩSpc* cassette from RU3940 was confirmed by PCR mapping using pOT forward with either oxp460 or opx461 in separate reactions (Supplementary Fig. S1). Enzyme assays on free living cultures confirmed conservation of the *gusA* and *celB* marker genes in the non-fixing mutants (Supplementary Fig. S2 and Supplementary Fig. S3 respectively). Bacterial cultures were maintained on TY agar[56] with 500 μg ml^{-1} streptomycin (all strains) and 100 μg ml^{-1} spectinomycin (RU3940, OPS0365 and OPS0366 only).

Plant growth. Pea (*Pisum sativum* cv Avola) seeds were germinated in the dark for five days on agar plates at room temperature and then transferred to 500 ml pots containing a mixture of silver sand - fine vermiculite (1:1 v/v) substrate, 75 ml nitrogen free nutrient solution (as in Poole *et al.*[57] but 2.67 times more concentrated) and 1 ml rhizobial inoculum. Peas were grown in the growth room at 21 °C with a 16 hour photoperiod.

Experimental designs. To test for partner choice, we applied rhizobial inocula consisting of 1:10, 1:1 and 10:1 ratios of fixing to non-fixing strains. To exclude any effect of the *gusA* or *celB* marker genes, both combinations were tested: fixing *gusA* (3841 *gusA*) with non-fixing *celB* (OPS0366) strains and fixing *celB* (3841 *celB*) with non-fixing *gusA* (OPS0365) strains. Total inoculum density for all treatments was 1×10^4 cells per pot. Rhizobial cultures for the inocula were spread-plated to confirm numbers of rhizobia in the inocula. In a fully randomized design, we grew 49 plants in total: six treatments plus a water control (no rhizobia), all replicated seven times. Plants were grown for 21 days without additional watering.

To assess the effect of inoculation density on the frequency of mixed nodules, plants were inoculated with 1:1 ratios of fixing *gusA* (3841 *gusA*) to non-fixing *celB* (OPS0366) strains at total densities of 1×10^2, 1×10^3, 1×10^6

and 1×10^8 cells per pot. In a fully randomized design, we grew five plants at each inoculum density plus two water controls. Plants were grown for 38 days and watered with sterilized water as needed after 21 days.

To test for differences in markers on the percentage of the nodule area occupied by the fixing strain, six plants were grown in an additional independent experiment, inoculated with 1:1 ratios of fixing *celB* (3841 *celB*) to non-fixing *gusA* strains (OPS0365) at a total density of 1 × 10^8 cells per pot. Plants were grown for 37 days and watered with sterilized water as needs after 21 days.

To assess the phenotype of non-fixing mutant strain RU3940, plants were inoculated with either fixing strain Rlv3841 or non-fixing strain RU3940 and grown for four or five weeks in independent experiments.

Harvest. At harvest, roots were gently washed and then stained for *gusA* and *celB* marker genes[58]. In aluminium covered tubes, roots were incubated overnight at 28 °C submerged in phosphate buffer (7 g L^{-1} NaH_2PO_4, 7.2 g L^{-1} Na_2HPO_4, 1 mM EDTA (pH 8), 1% Sarkosyl, 1 ml L^{-1} Triton) supplemented with 0.2 mg ml^{-1} Magenta-glc (5-bromo-6-chloro-3-indolyl-β-D-glucuronide). In fresh phosphate buffer, roots were then incubated at 70 °C for 1 hour 45 min. After cooling down, X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) was added to a final concentration of 0.25 mg ml^{-1} and roots were incubated overnight at 37 °C. This yielded visibly magenta (*gusA* marked strains), blue (*celB* marked strains) or magenta and blue (mixed) nodules (Fig. 2A–C). Shoots were dried at 70 degrees Celsius for 24–72 hours and then weighed.

Data collection. Roots were carefully laid out so that all nodules were visible and then photographed. Two people independently counted blue and magenta nodules from photographs taken of whole roots, zooming in as necessary, and the average was taken as the final count. Mixed nodules were counted and photographed using a dissecting microscope (Leica M165 FC) with accompanying software (LAS v4.5). To estimate size of nodules containing fixing versus non-fixing strains, we measured the area of nodules from the photographs taken of the whole root, using ImageJ v1.49v[59], which allowed for measuring areas of irregular shapes. Ten fixing and ten non-fixing nodules were randomly selected for each of the 14 plants inoculated with a 1:1 ratio of fixing to non-fixing strains (7 of which had the *gusA* marked fixing strain and 7 had the *celB* marked fixing strain). To estimate the percentage of mixed nodules that was occupied by the fixing strain, we measured the areas of blue and magenta in all mixed nodules for which this was possible (distinct enough areas of blue and magenta), using ImageJ. Additional microscopy was done on whole nodules (five-week-old nodules). Images were taken with a dissecting microscope (Leica M165 FC) with accompanying software (LAS v4.5).

Statistical analyses. We tested for partner choice by regressing the percentage of nodules containing nitrogen-fixing rhizobia against the percentage of nitrogen-fixing rhizobia in the inoculum. If plants do not exert partner choice, we expect a 1:1 relationship (a slope of 1.0). Water controls and mixed nodules were excluded from this analysis. To test whether the presence of *gusA* versus *celB* marker genes affected competitive ability, we carried out an ANCOVA with the marker gene as the categorical variable. Results presented are from the regression model without this interaction, as marker genes did not affect the slope ($t_{37} = -0.189$, $p = 0.851$) or intercept ($t_{37} = -0.177$, $p = 0.861$) of the regression line. The intended inoculation ratios were used as the explanatory variable, as it was confirmed using colony counts that the actual ratios in inocula did not deviate by more than 11% from intended inoculation ratios ($n = 6$, mean ± s.e. 3.78 ± 1.49%, Supplementary Table S1). Accounting for this deviation did not change results, and if anything would lead to a decrease of the estimate of the slope (whereas for partner choice we would expect an increase of the estimate of the slope).

To test for differences in size between fixing and non-fixing nodules, we performed a paired t-test on the average size of ten nodules, pairing fixing and non-fixing nodules of each plant. To test whether the total number of nodules formed depended on the ratio of fixing to non-fixing strains in the inoculum, we regressed the total number of nodules against the percentage of fixing strain in the inoculum. Shoot mass was also regressed against the percentage of the fixing strain in the inoculum. To assess the frequency of mixed nodules, we analysed how the occurrence of mixed nodules depended on the inoculation ratio of fixing to non-fixing strains with a generalized linear model using a quasibinomial distribution and a logit link function (because data were expressed as proportions and there was significant overdispersion). We also regressed the percentage of mixed nodules against the log of the total inoculation density. Whether the fixing strain was marked with *gusA* or *celB* had no impact on the proportion of mixed nodules detected ($t_{37} = 0.494$, $p = 0.62493$) or on the interaction with the encounter rate ($t_{37} = 1.137$, $p = 0.26301$), so we present results from analyses which do not include this interaction. To assess fitness of fixing to non-fixing strains within mixed nodules, we used a linear model testing for the effect of the marker (*gusA* or *celB*) used to mark the fixing strain. Individual nodules were seen as the unit of replication, and this was confirmed to be a valid assumption as the plant from which nodules came accounted for only 9% of the variation in the data. All analyses were performed in R version 3.0.2[60].

References

1. Lebauer, D. S. & Treseder, K. K. Nitrogen limitation of net primary productivity in terrestrial ecosystems is globally distritubed. *Ecology* **89**, 371–379, doi:10.1890/06-2057.1 (2008).
2. Sprent, J. I. Evolving ideas of legume evolution and diversity: a taxonomic perspective on the occurrence of nodulation. *New Phytol* **174**, 11–25, doi:10.1111/j.1469-8137.2007.02015.x (2007).
3. Graham, P. H. & Vance, C. P. Legumes: importance and constraints to greater use. *Plant Physiol* **131**, 872–877, doi:10.1104/pp.017004 (2003).
4. Hafeez, F., Hameed, S., Ahmad, T. & Malik, K. A. Competition between effective and less effective strains of Bradyrhizobium spp. for nodulation on Vigna radiata. *Biol. Fertil. Soils* **33**, 382–386, doi:10.1007/s003740000337 (2001).
5. Kiers, E. T., Hutton, M. G. & Denison, R. F. Human selection and the relaxation of legume defences against ineffective rhizobia. *Proc. R. Soc. B* **274**, 3119–26, doi:10.1098/rspb.2007.1187 (2007).
6. Denison, R. F. Legume sanctions and the evolution of symbiotic cooperation by rhizobia. *Am. Nat.* **156**, 567–576, doi:10.1086/316994 (2000).

7. Heath, K. & Tiffin, P. Stabilizing mechanisms in a legume–rhizobium mutualism. *Evolution (N. Y)* **63**, 652–662 (2009).
8. Udvardi, M. & Poole, P. S. Transport and metabolism in legume-rhizobia symbioses. *Annu. Rev. Plant Biol.* **64**, 781–805, doi:10.1146/annurev-arplant-050312-120235 (2013).
9. Gage, D. J. Analysis of infection thread development using Gfp- and DsRed-expressing Sinorhizobium meliloti. *J. Bacteriol.* **184**, 7042–7046, doi:10.1128/JB.184.24.7042-7046.2002 (2002).
10. Oldroyd, G. E. D. & Downie, J. A. Coordinating nodule morphogenesis with rhizobial infection in legumes. *Annu. Rev. Plant Biol.* **59**, 519–46, doi:10.1146/annurev.arplant.59.032607.092839 (2008).
11. Hagen, M. & Hamrick, J. Population level processes in Rhizobium leguminosarum bv. trifolii: the role of founder effects. *Mol. Ecol.* **5**, 707–714, doi:10.1111/j.1365-294X.1996.tb00367.x (1996).
12. Silva, C., Eguiarte, L. E. & Souza, V. Reticulated and epidemic population genetic structure of Rhizobium etli biovar phaseoli in a traditionally managed locality in Mexico. *Mol. Ecol* **8**, 277–287, doi:10.1046/j.1365-294X.1999.00564.x (1999).
13. Kiers, E. T. & Denison, R. F. Sanctions, cooperation, and the stability of plant-rhizosphere mutualisms. *Annu. Rev. Ecol. Evol. Syst.* **39**, 215–236, doi:10.1146/annurev.ecolsys.39.110707.173423 (2008).
14. Gubry-Rangin, C., Garcia, M. & Béna, G. Partner choice in Medicago truncatula-Sinorhizobium symbiosis. *Proc. Biol. Sci* **277**, 1947–51, doi:10.1098/rspb.2009.2072 (2010).
15. Regus, J., Gano, K. A., Hollowell, A. C. & Sachs, J. L. Efficiency of partner choice and sanctions in Lotus is not altered by nitrogen fertilization. *Proc. R. Soc. B* **281**, 20132587–20132587, doi:10.1098/rspb.2013.2587 (2014).
16. West, S. A., Kiers, E. T., Pen, I. & Denison, R. F. Sanctions and mutualism stability: when should less beneficial mutualists be tolerated? *J. Evol. Biol* **15**, 830–837, doi:10.1046/j.1420-9101.2002.00441.x (2002).
17. Simms, E. & Taylor, D. Partner choice in nitrogen-fixation mutualisms of legumes and rhizobia. *Integr. Comp. Biol.* **42**, 369–380, doi:10.1093/icb/42.2.369 (2002).
18. Simms, E. L. *et al*. An empirical test of partner choice mechanisms in a wild legume-rhizobium interaction. *Proc. Biol. Sci* **273**, 77–81, doi:10.1098/rspb.2005.3292 (2006).
19. Kiers, E. T., Rousseau, R. A., West, S. A. & Denison, R. F. Host sanctions and the legume-rhizobium mutualism. *Nature* **425**, 78–81, doi:10.1038/nature01931 (2003).
20. Kiers, T. E., Rousseau, R. A. & Denison, R. F. Measured sanctions: legume hosts detect quantitative variation in rhizobium cooperation and punish accordingly. *Evol. Ecol. Res.* **8**, 1077–1086 (2006).
21. Oono, R., Anderson, C. G. & Denison, R. F. Failure to fix nitrogen by non-reproductive symbiotic rhizobia triggers host sanctions that reduce fitness of their reproductive clonemates. *Proc. R. Soc. B* **278**, 2698–703, doi:10.1098/rspb.2010.2193 (2011).
22. Malek, W. The role of motility in the efficiency of nodulation by Rhizobium meliloti. *Arch. Microbiol.* **158**, 26–28, doi:10.1007/BF00249061 (1992).
23. Robleto, E. A., Kmiecik, K., Oplinger, E. S., Nienhuis, J. & Triplett, E. W. Trifolitoxin production increases nodulation competitiveness of Rhizobium etli CE3 under agricultural conditions. *Appl. Environ. Microbiol.* **64**, 2630–2633 (1998).
24. Frederix, M. *et al*. Mutation of praR in Rhizobium leguminosarum enhances root biofilms, improving nodulation competitiveness by increased expression of attachment proteins. *Mol. Microbiol.* **93**, 464–478, doi:10.1111/mmi.12670 (2014).
25. Triplett, E. W. & Sadowsky, M. Genetics of competition for nodulation of legumes. *Annu. Rev. Microbiol.* **46**, 399–428, doi:10.1146/annurev.mi.46.100192.002151 (1992).
26. Friesen, M. L. Widespread fitness alignment in the legume – rhizobium symbiosis. *New Phytol* **194**, 1096–1111, doi:10.1111/nph.2012.194.issue-4 (2012).
27. Hahn, M. & Studer, D. Competitiveness of a nif- Bradyrhizobium japonicum mutant against the wild-type strain. *FEMS Microbiol. Lett.* **33**, 143–148 (1986).
28. Gordon, B. R. *et al*. Decoupled genomic elements and the evolution of partner quality in nitrogen-fixing rhizobia. *Ecol. Evol* **6**(5), 1317–1327, doi:10.1002/ece3.2016.6.issue-5 (2016).
29. West, S. A., Kiers, E. T., Simms, E. L. & Denison, R. F. Sanctions and mutualism stability: why do rhizobia fix nitrogen? *Proc. R. Soc. B* **269**, 685–94, doi:10.1098/rspb.2001.1878 (2002).
30. Oono, R., Denison, R. & Kiers, E. Controlling the reproductive fate of rhizobia: how universal are legume sanctions? *New Phytol* **183**, 967–979, doi:10.1111/nph.2009.183.issue-4 (2009).
31. Marco, D. E. *et al*. An experimental and modelling exploration of the host-sanction hypothesis in legume-rhizobia mutualism. *J. Theor. Biol.* **259**, 423–33, doi:10.1016/j.jtbi.2009.03.033 (2009).
32. Bshary, R. & Schäffer, D. Choosy reef fish select cleaner fish that provide high-quality service. *Anim. Behav.* **63**, 557–564, doi:10.1006/anbe.2001.1923 (2002).
33. Pellmyr, O. & Huth, C. J. Evolutionary stability of mutualism between yuccas and yucca moths. *Nature* **372**, 257–260, doi:10.1038/372257a0 (1994).
34. Tarachai, Y., Compton, S. G. & Trisonthi, C. The benefits of pollination for a fig wasp. *Symbiosis* **45**, 29–32 (2008).
35. Oldroyd, G. E. D. & Downie, J. A. Calcium, kinases and nodulation signalling in legumes. *Nat. Rev. Mol. Cell Biol.* **5**, 566–576, doi:10.1038/nrm1424 (2004).
36. Garg, N. Symbiotic nitrogen fixation in legume nodules: process and signaling. A review. *Agron. Sustain. Dev.* **27**, 59–68, doi:10.1051/agro:2006030 (2007).
37. Murray, J. D. Invasion by invitation: rhizobial infection in legumes. *Mol. Plant. Microbe. Interact.* **24**, 631–639, doi:10.1094/MPMI-08-10-0181 (2011).
38. Yang, S., Tang, F., Gao, M., Krishnan, H. B. & Zhu, H. R gene-controlled host specificity in the legume-rhizobia symbiosis. *Proc. Natl. Acad. Sci. USA.* **107**, 18735–18740, doi:10.1073/pnas.1011957107 (2010).
39. Lira, M. A., Nascimento, L. R. S. & Fracetto, G. G. M. Legume-rhizobia signal exchange: Promiscuity and environmental effects. *Front. Microbiol* **6**, 1–9 (2015).
40. Jones, K. M. *et al*. Differential response of the plant Medicago truncatula to its symbiont Sinorhizobium meliloti or an exopolysaccharide-deficient mutant. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 704–709, doi:10.1073/pnas.0709338105 (2008).
41. Via, V. D., Zanetti, M. E. & Blanco, F. How legumes recognize rhizobia. *Plant Signal. Behav* **11**, e1120396, doi:10.1080/15592324.2015.1120396 (2015).
42. Young, J. P. & Johnston, A. W. The evolution of specificity in the legume-rhizobium symbiosis. *Trends Ecol. Evol.* **4**, 341–349, doi:10.1016/0169-5347(89)90089-X (1989).
43. Barrett, L. G., Zee, P. C., Bever, J. D., Miller, J. T. & Thrall, P. H. Evolutionary history shapes patterns of mutualistic benefit in Acacia-rhizobial interactions. *Evolution (N. Y).* (pp. 1473–1485 (2016).
44. Terpolilli, J. J., Hood, G. A. & Poole, P. S. What determines the efficiency of N2-fixing Rhizobium-legume symbioses? *Adv. Microb. Physiol* **60**, 325–89, doi:10.1016/B978-0-12-398264-3.00005-X (2012).
45. Downie, J. A. The roles of extracellular proteins, polysaccharides and signals in the interactions of rhizobia with legume roots. *FEMS Microbiol. Rev.* **34**, 150–170, doi:10.1111/j.1574-6976.2009.00205.x (2010).
46. Batstone, R. T., Dutton, E. M., Wang, D., Yang, M. & Frederickson, M. E. The evolution of symbiont preference traits in the model legume Medicago truncatula. *New Phytol.* **213**, 1850–1861, doi:10.1111/nph.14308 (2016).
47. Grillo, M. A., Mita, S., De, Burke, P. V., Sol, K. L. S. & Heath, K. D. Intrapopulation genomics in a model mutualist: Population structure and candidate symbiosis genes under selection in Medicago truncatula. *Evolution (N. Y)* **70**, 2704–2717 (2016).

48. Yoder, J. B. Understanding the coevolutionary dynamics of mutualism with population genomics. *Am. J. Bot.* **103**, 1742–1752, doi:10.3732/ajb.1600154 (2016).
49. Steidinger, B. S. & Bever, J. D. Host discrimination in modular mutualisms: a theoretical framework for meta-populations of mutualists and exploiters. *Proc. R. Soc. B* **283**, 20152428, doi:10.1098/rspb.2015.2428 (2016).
50. Zgadzaj, R., James, E. K., Kelly, S., Kawaharada, Y. & De, N. Infection of Nodules by Symbiotic and Endophytic Bacteria. *PLoS Genet.* **11**, 1–21, doi:10.1371/journal.pgen.1005280 (2015).
51. Moawad, H. A., Ellis, W. R. & Schmidt, E. L. Rhizosphere response as a factor in competition among three serogroups of indigenous Rhizobium japonicum for nodulation of field-grown soybeans. *Appl. Environ. Microbiol.* **47**, 607–612 (1984).
52. Sánchez-Cañizares, C. & Palacios, J. Construction of a marker system for the evaluation of competitiveness for legume nodulation in Rhizobium strains. *J. Microbiol. Methods* **92**, 246–249, doi:10.1016/j.mimet.2012.12.022 (2013).
53. Young, J. P. W. *et al*. The genome of Rhizobium leguminosarum has recognizable core and accessory components. *Genome Biol.* **7**, R34, doi:10.1186/gb-2006-7-4-r34 (2006).
54. Fischer, H. M. Genetic regulation of nitrogen fixation in rhizobia. *Microbiol. Rev.* **58**, 352–386 (1994).
55. Buchanan-Wollaston, V. Generalized transduction in Rhizobium meliloti. *J. Gen. Microbiol.* **112**, 135–142, doi:10.1099/00221287-112-1-135 (1979).
56. Beringer, J. E. R factor transfer in Rhizobium leguminosarum. *J. Gen. Microbiol* **84**, 188–198, doi:10.1099/00221287-84-1-188 (1974).
57. Poole, P. S., Blyth, A., Reid, C. J. & Walters, K. Myo-inositol catabolism and catabolite regulation in Rhizobium leguminosarum bv. viciae. *Microbiology* **140**, 2787–2795, doi:10.1099/00221287-140-10-2787 (1994).
58. Sessitsch, A., Wilson, K. J., Akkermans, A. D. L. & De Vos, W. M. Simultaneous detection of different Rhizobium strains marked with either the *Escherichia coli* gusA gene or the Pyrococcus furiosus celB gene. *Appl. Environ. Microbiol.* **62**, 4191–4194 (1996).
59. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675, doi:10.1038/nmeth.2089 (2012).
60. R Core Team. R: A language and environment for statistical computing (2013).
61. Johnston, A. W. & Beringer, J. E. Identification of the rhizobium strains in pea root nodules using genetic markers. *J. Gen. Microbiol.* **87**, 343–350, doi:10.1099/00221287-87-2-343 (1975).
62. Prentki, P. & Krisch, H. M. *In vitro* insertional mutagenesis with a selectable DNA fragment. *Gene* **29**, 303–313, doi:10.1016/0378-1119(84)90059-3 (1984).
63. Quandt, J. & Hynes, M. F. Versatile suicide vectors which allow direct selection for gene replacement in Gram-negative bacteria. *Gene* **127**, 15–21, doi:10.1016/0378-1119(93)90611-6 (1993).

Acknowledgements

We would like to thank Marcela Mendoza-Suárez and Carmen Sánchez-Cañizares for advice during execution of the experiments, and John Baker for taking the photographs. This work was supported by Engineering and Physical Sciences Research Council (EPSRC) [grant number 1364881]; and the Biotechnology and Biological Sciences Research Council [grant numbers BB/K006134/1, BB/J007749/2, BB/J014524/1].

Author Contributions

A.W., L.A.T. and P.S.P. planned the experiments. A.W. carried out the experiments, F.R. assisted during harvest and with counting the nodules, and I.W. with the acetylene reductions. E.F. and A.W. carried out a pilot experiment. G.M. made the unmarked non-fixing strain RU3940. A.W., L.A.T. and P.S.P. analysed the results and prepared the manuscript.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-01634-2

Accession codes: The data that support the findings of this study are available in the Oxford Research Archive (ORA-Data) repository, at doi:10.5287/bodleian:BeXp1Prb B.

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2017

Appendix B. Supplementary material chapter 3: Partner choice in the legume-rhizobia symbiosis

Overview

The supplementary material of chapter 3 consists of:

1. Supplementary material as published with the published manuscript
2. ReadMe document for data
3. R code used to analyse the data.

The data and R code that support the findings of this study are available in the Oxford Research Archive (ORA-Data) repository, DOI: 10.5287/bodleian:BeXp1PrbB.

1. Supplementary material

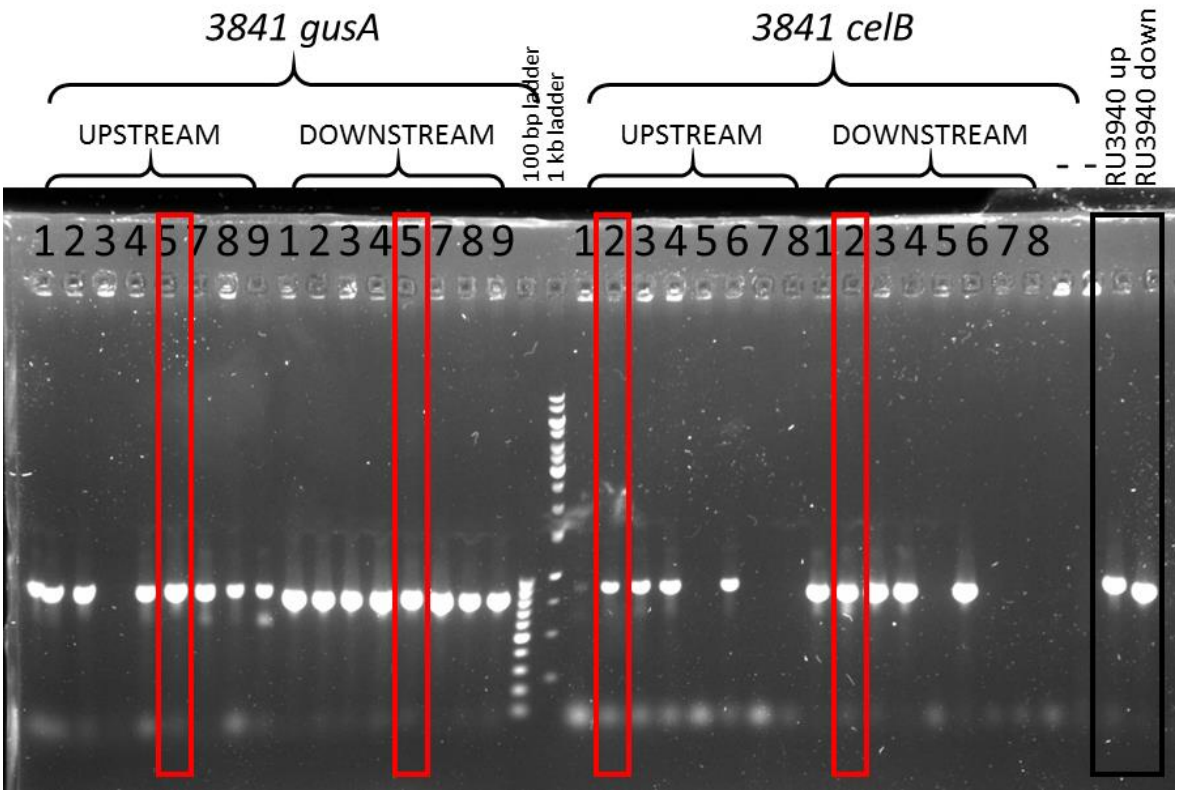
Supplementary information

Policing the legume-Rhizobium symbiosis: a critical test of partner choice.

Annet Westhoek[1, 2], Elsa Field[1], Finn Rehling[1,3], Geraldine Mulley[4], Isabel Webb[1], Philip S. Poole[1*] & Lindsay A. Turnbull[1*]

[1]Department of Plant Sciences, University of Oxford, Oxford OX1 3RB, UK. [2]Systems Biology Doctoral Training Centre, University of Oxford, Oxford OX1 3RQ, UK. [3]Department of Ecology, Philipps-University Marburg, Marburg D-35043, Germany. [4]School of Biological Sciences, University of Reading, RG6 6AJ, UK. Correspondence should be sent to PSP (philip.poole@plants.ox.ac.uk) and/or LAT (lindsay.turnbull@plants.ox.ac.uk).

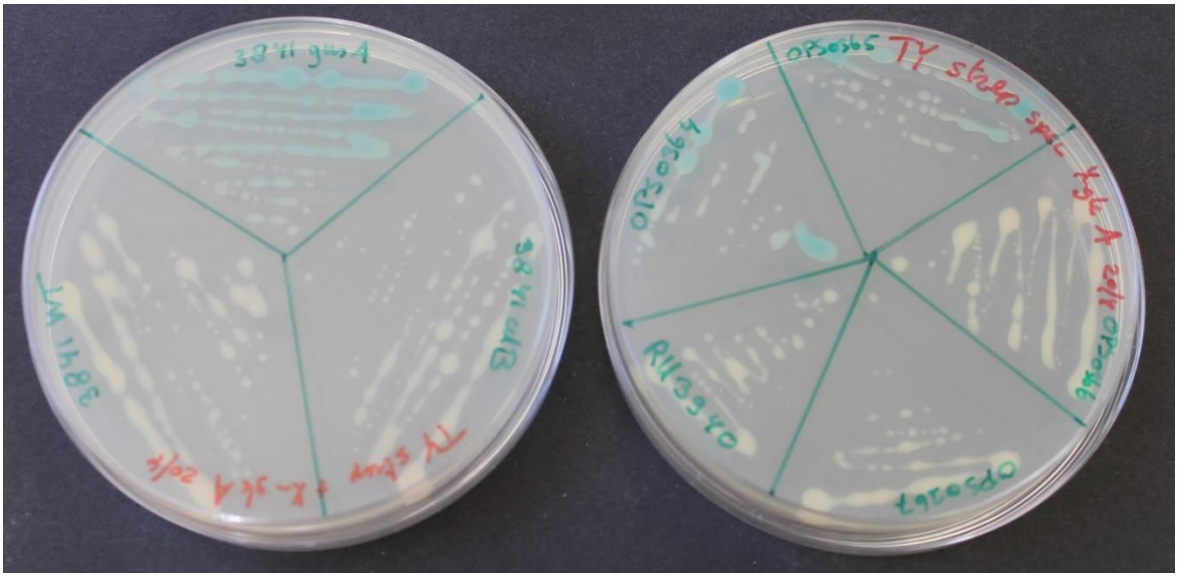
Supplementary figure 1



Supplementary Figure 1. Confirmation of successful insertion of *nifh::Ωspec* mutation into *gusA* and *celB* marked strains. Successful insertion yields *gusA* and *celB* marked *nifh::Ωspec* (non-fixing) mutants. Primers were designed to bind in the *Ωspec* cassette and upstream and downstream of the *nifH* gene, yielding fragments upstream and downstream of the *Ωspec* cassette respectively. If the insertion worked correctly, PCR fragments upstream and downstream of the insertion have the same length in the marked mutants as in the original unmarked *nifh::Ωspec* mutant (RU3940, black box). Multiple strains were made (8 *celB* strains and 9 *gusA* strains (strain 6 not included in PCR because *gusA* marker could not be confirmed)), some of which could not be confirmed. *gusA* marked strain 5 (OPS0365) and *celB* marked strain 2 (OPS0366) were used in this study (red boxes).

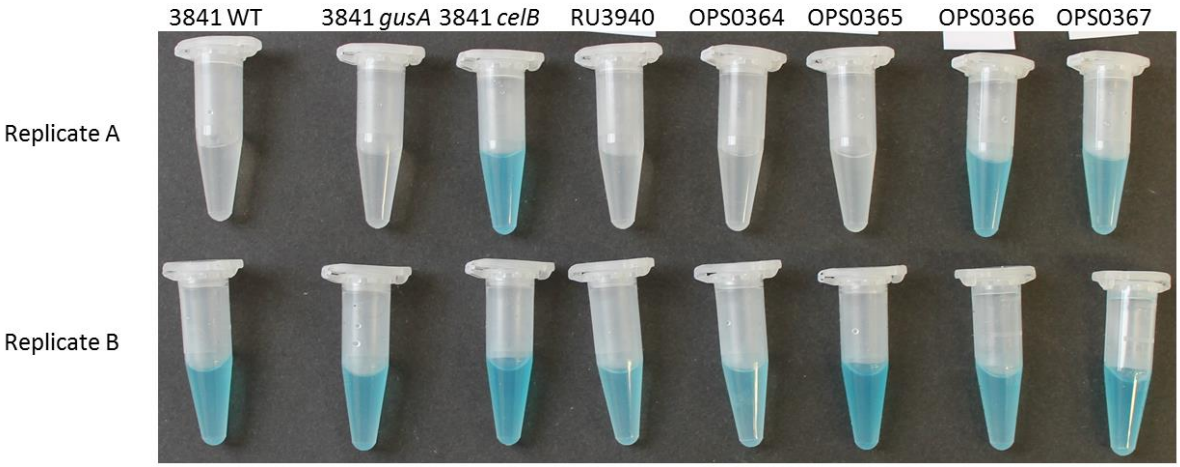
Primers used for upstream fragment: oxp0460 (GCTTGATCATCGCCGGAAAC), binding before the start codon of the *nifH* gene, and pOT forward (CGGTTTACAAGCATAAAGC), binding in the *Ωspec* cassette. Primers used for downstream fragment: pOT forward (binds in two places in *Ωspec* cassette) and oxp0461 (TGTCACCGCCGAAAACGATG), binding downstream of the *nifH* gene.

Supplementary figure 2



Supplementary Figure 2. Confirmation conservation of *gusA* marker gene. It was confirmed that the novel strains with the *nifh::Ωspec* insertion still have the original *gusA* marker. The left hand plate contains the nitrogen-fixing strains on TY medium with streptomycin and X-glcA, the substrate which yields a blue colour if there is *gusA* encoded β-glucuronidase activity. Only the *gusA* marked strain is blue (top section), not the unmarked wild type and *celB* marked strain, as expected. The plate on the right hand side contains *nifh::Ωspec* mutants on TY medium with streptomycin, spectinomycin and X-glcA. Again, only the *gusA* marked strains OPS0364 (not used in this study) and OPS0365 (used in this study) are blue (top two sections).

Supplementary figure 3



Supplementary Figure 3. Confirmation conservation of *celB* marker gene. It was confirmed that the novel strains with the *nifh::Ωspec* insertion still have the original *celB* marker. Free living cultures provided with X-gal yield a blue colour if there is *celB* encoded β-galactosidase activity. Replicate B (bottom) was not heat treated, so the endogenous β-galactosidase activity yields a blue colour for all strains. Replicate A (top) was heat treated, during which endogenous β-galactosidase was destroyed. Only the thermostable *celB* marked strains (*celB* marked nitrogen-fixing 3841, and OPS0366 (this study) and OPS0367) yield a blue colour. Unmarked original wild type 3841 and *nifh::Ωspec* (RU3940) are white, just as *gusA* marked strains OPS0364 and OPS0365.

Supplementary table 1

Ratio fixing:non-fixing strain	Fixing strain	Intended percentage fixing strain	Estimated achieved percentage fixing strain	Deviation (percentage)
1:10	*gusA*	9.09	13.5	4.5
1:10	*celB*	9.09	10.0	1.0
1:1	*gusA*	50	60.7	10.7
1:1	*celB*	50	52.7	2.7
10:1	*gusA*	90.9	93.9	3.0
10:1	*celB*	90.9	91.8	0.9

Table 1. Intended and achieved percentage of fixing strain in the inocula.

2. ReadMe document for data

Data supporting:

Policing the legume-Rhizobium symbiosis: a critical test of partner choice

Annet Westhoek, Elsa Field, Finn Rehling, Geraldine Mulley, Isabel Webb, Philip S. Poole & Lindsay A. Turnbull

Data gathered at University of Oxford, August 2015 – January 2017.
Data comes from five independent experiments:

Exp. 8 (partner choice)	fixing to non-fixing strains in varying ratios at inoculation density of 1 x 10^4 cells / plant. Fixing strain both *gusA* and *celB* marked
Exp. 9 (inoculation density)	fixing to non-fixing strains in 1:1 ratio at varying inoculation densities. Fixing strain *gusA* marked.
Exp. 11 (marker check)	fixing to non-fixing strain in 1:1 ratio at inoculation density of 1x10^8 cells / plant. Fixing strain *celB* marked.
Acetylene reductions	Plants with single inoculations of fixing Rlv3841 strain, non-fixing *nifH* mutant RU3940 and water control.
Shoot mass single inoculation	Plants with single inoculations of fixing Rlv3841 strain, non-fixing *nifH* mutant RU3940 and water control.

Data consists of:

- **Pictures**
  Whole root systems with stained nodules. File names: exp8_1_white_Imgxxxxx. The number indicates plant number.
  Microscopic images of mixed nodules for three experiments. File names: x_x. First number indicates plant number, second number indicates picture number.
- **Nodule number data**. Containing data on the number of *gusA*, *celB* and mixed infection nodules counted in partner choice experiment (exp. 8). Also contains data on shoot weight.

Plant	Plant number
wild_type	Whether fixing strain is *gusA* or *celB*
marked nifh_mutant	Whether non-fixing strain is *gusA* or *celB* marked
ratio	Ratio of fixing to non-fixing strain in inoculum
perc_wt_inoculum	Percentage of fixing (wild-type) strain in inoculum
number_gusA_nodules_person1	Number of *gusA* nodules as counted by person 1
number_celB_nodules_person1	Number of *celB* nodules as counted by person 1
number_gusA_nodules_person2	Number of *gusA* nodules as counted by person 2

number_celB_nodules_person2	Number of *celB* nodules as counted by person 2
number_mixed_nodules	Number of mixed nodules as counted with microscope
perc_wt_inoculum_estimated	Estimated achieved percentage of fixing strain in inoculum, based on cell counts plated out inoculum
shoot_weight	Dry weight of shoot in grams.

- **Nodule size data**. Containing data on the size of fixing versus non-fixing nodules in partner choice experiment (exp. 8).

Nodule	Nodule number. 10 nodules per type (fixing or non-fixing) per plant were measured
Label	Indicating from which plant (picture) nodules were measured
Area	Area measured (as measure of nodule size) in cm^2. Used size of square plate (12x12 cm) to set scale.
colour	Colour of nodule measured (blue for *celB* and pink for *gusA* stained nodules)
type	Whether the nodule was fixing or non-fixing

- **Number mixed nodules data.** Containing data on number of mixed nodules as function of inoculation density (exp. 9). Fixing strain is *gusA* marked.

Plant	Plant number
wild_type	Whether fixing strain is *gusA* or *celB* marked
nifh_mutant	Whether non-fixing strain is *gusA* or *celB* marked
ratio	Ratio of fixing to non-fixing strain in inoculum
Density	Total inoculation density
Nr_celB_nodules_person1	Number of *celB* nodules as counted by person 1
Nr_gusA_nodules_person1	Number of *gusA* nodules as counted by person 1
Nr_mixed_nodules	Number of mixed nodules as counted with microscope
Nr_celB_nodules_person2	Number of *celB* nodules as counted by person 2
Nr_gusA_nodules_person2	Number of *gusA* nodules as counted by person 2

- **Mixed nodules area data.** Containing data on area within mixed nodules occupied by fixing and non-fixing strains. Data from three experiments: exp. 8 (partner choice), exp. 9 (inoculation density) and exp. 11 (marker check).

Label	Indicating from which plant (picture) nodules were measured
Area	Area measured (as measure of nodule size) in mm^2.
Color	Color of nodule measured (blue for *celB* and magenta for *gusA* stained nodules)

Comments	If more than two mixed nodules were present in picture, indicating which nodule measurements are from (left, right, top, bottom).
Marker_Fixing_Strain	Whether fixing strain was *gusA* or *celB* marked.
Experiment	From which experiment this plant and picture are a part (exp. 8, exp. 9 or exp. 11).

- **Acetylene reductions** Containing data on acetylene reductions of plants singly inoculated with fixing 3841 strain, non-fixing RU3940 mutant and water control.

Strain	Which strain plant was inoculated with
Plant	Plant number
Replicate	Technical replicate (three per plant)
Perc_acetylene	Percentage acetylene measured
Perc_ethylene	Percentage ethylene measured

- **Shoot mass single inoculation** Containing data on shoot mass of plants singly inoculated with fixing 3841 strain, non-fixing RU3940 mutant and water control.

Strain	Which strain plant was inoculated with
Plant	Plant number
Weight	Dry weight of shoot in grams

- **R code** used to analyse the data.

3. R code used to analyse the data

Data analysis partner choice

Annet Westhoek

January 25, 2017

```
print(sessionInfo(), locale = FALSE)

## R version 3.3.2 (2016-10-31)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 15063)
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] dplyr_0.5.0    scales_0.4.1   reshape_0.8.6  plyr_1.8.4
##  [5] boot_1.3-18    arm_1.9-3      lme4_1.1-12    Matrix_1.2-7.1
##  [9] MASS_7.3-45    ggplot2_2.2.1
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_0.12.9      knitr_1.15.1     magrittr_1.5     splines_3.3.2
##  [5] munsell_0.4.3    colorspace_1.3-2 lattice_0.20-34  R6_2.2.0
##  [9] minqa_1.2.4      stringr_1.1.0    tools_3.3.2      grid_3.3.2
## [13] nlme_3.1-128     gtable_0.2.0     DBI_0.5-1        coda_0.19-1
## [17] htmltools_0.3.5  abind_1.4-5      yaml_2.1.14      lazyeval_0.2.0
## [21] rprojroot_1.2    digest_0.6.12    assertthat_0.1   tibble_1.2
## [25] nloptr_1.0.4     evaluate_0.10    rmarkdown_1.3    stringi_1.1.2
## [29] backports_1.0.5

# last update January 25, 2017

## Code supporting: 'Policing the legume - Rhizobium symbiosis: a
# critical test of partner choice. Westhoek et al., 2017.

## Note: wild type refers to fixing strain, mutant refers to
# non-fixing nifH mutant strain

## Data analysed:
# nodule number data:     data on the number of nodules and mixed nodules
#                         and shoot weight
#                         in partner choice experiment ('exp. 8')
# nodule size data:       data on the size of nodules in
#                         partner choice experiment ('exp. 8')
# mixed nodule numbers:   data on the number of mixed nodules
#                         as function of inoculation density ('exp. 9')
# mixed nodules area:     data on area of fixing and non-fixing strain
#                         within mixed nodules from
#                         three different experiments.(exp. 8, exp. 9, exp. 11
)
#                         Exp. 11 just had 6 plants at inoculation density of
#                         10^8 CFUs with fixing strain celB marked
#                         (exp. 9 had fixing strain gusA marked, so exp. 11
#                         was just to check for effect of marker gene)
```

```
# acetylene reductions:   data on acetylene reductions of strains 3841
#                         (fixing wild type), RU3940 (non-fixing nifH mutant)
#                         and water control.
# shoot mass single       data on shoot mass of plants singly
#                         inoculated with strains 3841
# inoculations            (fixing wild type), RU3940 (non-fixing nifH mutant)
#                         and water control ('exp. 25')

# read in the results for nodule number data partner choice
# experiment (exp. 8)
results <- read.csv("Westhoek et al., 2017, nodule number.csv")
head(results)
str(results)

# some data processing:
# number of gusA and celB nodules is average of counts
# two different persons
# Person 2 missed one plant,
# so for that one only take counts of Person 1
results$number_gusA_nodules <-
  ifelse(is.na(results$number_gusA_nodules_person2),
                       results$number_gusA_nodules_person1,
                      ((results$number_gusA_nodules_person1 +
                      results$number_gusA_nodules_person2) / 2))
results$number_celB_nodules <-
  ifelse(is.na(results$number_celB_nodules_person2),
                results$number_celB_nodules_person1,
                  ((results$number_celB_nodules_person1 +
              results$number_celB_nodules_person2) / 2))
# calculate the total number of nodules
results$total_number_nodules <- results$number_gusA_nodules +
  results$number_celB_nodules +
  results$number_mixed_nodules
# calculate total number of nodules excluding mixed
# (because want to compare percentage
# wild type vs nifH in single-infection nodules)
results$total_number_nodules_excl_mixed <-
  results$number_gusA_nodules +
  results$number_celB_nodules
# calculate percentage wild type nodules
results$perc_wt_nodules <-
  ifelse(results$wild_type == "gusA", (results$number_gusA_nodules / results$t
otal_number_nodules_excl_mixed)*100,
(results$number_celB_nodules / results$total_number_nodules_excl_mixed)*100)
# calculate fraction of mixed nodules
results$frac_mixed_nodules <-
  results$number_mixed_nodules / results$total_number_nodules

# exclude water controls from further analysis
results_excl_WC <- subset(results, results$wild_type != "WC")

########################################################################
#######            PARTNER CHOICE                        ############
########################################################################

## plot % WT nodules against intended inoculum ratio
plot_partner_choice <- ggplot(data = results_excl_WC,
                              aes(x = perc_wt_inoculum,
                               y = perc_wt_nodules)) +
```

```
  geom_point(aes(colour = wild_type, shape = wild_type),
             stroke = 4, size = 10) +
  geom_smooth(method = lm, colour = "black", fill = "black",
              size = 3, alpha = 0.6) +
   scale_colour_manual(name = "fixing strain",
      labels = c(expression(paste(italic("gusA"), " marked")),
                 expression(paste(italic("celB"), " marked"))),
      values = c("gusA" = "magenta", "celB" = "blue"),
      breaks = c("gusA", "celB")) +
  scale_shape_manual(name = "fixing strain",
      labels = c(expression(paste(italic("gusA"), " marked")),
                 expression(paste(italic("celB"), " marked"))),
      values = c("gusA" = 1, "celB" = 2),
      breaks = c("gusA", "celB")) +
  theme_bw(base_size = 38) +
  theme(panel.grid.major = element_blank(),
        legend.key = element_blank(),
        legend.title = element_text(face = "bold"),
        axis.line.x = element_line(size = 2, color = "black"),
        axis.line.y = element_line(size = 2, color = "black"),
        legend.justification=c(1,0), legend.position=c(1,0)) +
  guides(color=guide_legend(override.aes=list(fill=NA))) +
  xlab("% fixing strain in inoculum") +
  ylab(expression(atop("% nodules containing",
                       paste("only fixing strain")))) +
  annotate('text', x = 39, y = 75,
  label = "italic(y) == -1.38 + 0.99~italic(x)~','~italic(R^2) == 0.99",
  size = 10, parse = TRUE)
plot_partner_choice

# Export graph
tiff("figure 2d.tiff", width = 10, height = 10, units = 'in',
     res = 300, compression = 'none')
plot_partner_choice # Make plot
dev.off()

# regress % fixing nodules against % fixing strain in inoculum
# include possible interaction with marker gene (gusA or celB)
model_partner_choice <- lm(perc_wt_nodules ~ perc_wt_inoculum*wild_type,
                           data = results_excl_WC)
coef(model_partner_choice)
summary(model_partner_choice)
confint(model_partner_choice)
display(model_partner_choice)

# interaction with marker gene non-significant,
# so exclude for further analyses
model_partner_choice <- lm(perc_wt_nodules ~ perc_wt_inoculum,
                           data = results_excl_WC)
coef(model_partner_choice)
summary(model_partner_choice)
confint(model_partner_choice)
display(model_partner_choice)

# 95% CI of slope includes 1 so no reason to reject null hypothesis that
# % fixing strain in nodules reflects % fixing strain in inoculum
# extra test: do t-test on slope (see Whitlock and Schluter p. 475)
# t = (estimate of slope - null hypothesized value) / standard error
# of slope
```

```r
calculated_t <- (0.98544 - 1) / 0.01743
# then use t-distribution with n-2 degrees of freedom (41 - 2 = 39)
# find area (probability) t < found t and
# multiply by two for two-sided
p_value <- 2*pt(calculated_t, 39)
p_value

# check for normality
par(mfrow = c(2,2))
plot(model_partner_choice)
par(mfrow = c(1,1))

#############################################################################
#######                      MIXED NODULES                     ############
#############################################################################

# look at % mixed nodules as function of inoculation ratio
# is % of mixed nodules dependent on inoculation ratio?

# first calculate average frequency of mixed nodules
mean_mixed <- mean(results_excl_WC$frac_mixed_nodules) # 2.02 %
sd_mixed <- sd(results_excl_WC$frac_mixed_nodules) # 2.21 %
se_mixed <- sd_mixed / sqrt(nrow(results_excl_WC)) # 0.35 %

# expect that number of mixed nodules depends on probability
# of two strains encountering
# each other during nodule infection
# add encounter rate to data: if chance of having
# two rhizobia in one spot = 1 (unknown),
# then chance of having mixed nodule is
# 0.909 * 0.0909 = 0.08 (10:1 and 1:10 wt:nifh)
# or 0.5 * 0.5 = 0.25 (1:1 wt:nifh)
results_excl_WC$encounter <-
  round((results_excl_WC$perc_wt_inoculum / 100) *
            ((100 - results_excl_WC$perc_wt_inoculum) / 100),2)

# plot fraction of mixed nodules as function of encounter rate
plot_encounter <- ggplot(data = results_excl_WC,
                         aes(x = encounter,
                             y = frac_mixed_nodules)) +
  geom_point(aes(shape = ratio), size = 10, stroke = 4) +
  geom_smooth(method = "glm",
              method.args = (list(family = quasibinomial)),
              aes(weight = total_number_nodules),
              colour = "black", fill = "black",
              size = 3, alpha = 0.5) +
  scale_shape_manual(name = c("fixing : non-fixing"),
              labels = c("1:10", "1:1", "10:1"),
              values = c("1:01" = 1, "1:10" = 2,"10:01" = 0),
              breaks = c("1:10", "1:01", "10:01")) +
  theme_bw(base_size = 38) +
  theme(panel.grid.major = element_blank(),
        legend.key = element_blank(),
        axis.line.x = element_line(size = 2, color = "black"),
        axis.line.y = element_line(size = 2, color = "black"),
        legend.justification=c(0,1), legend.position=c(0,1)) +
  xlab(expression(paste("Encounter rate (", italic("pq"), ")"))) +
  ylab(expression(paste("Fraction mixed nodules (",italic("F"),")")))
plot_encounter
```

```
# Export graph
tiff("figure 5d.tiff", width = 10, height = 8, units = 'in',
     res = 300, compression = 'none')
plot_encounter # Make plot
dev.off()

# quasibinomial model (need to round because took
# averages person 1 and person 2)
# quasibinomial because of overdispersion
model_encounter <- glm(cbind(number_mixed_nodules,
round(total_number_nodules_excl_mixed)) ~ encounter*wild_type,
                       data = results_excl_WC,
                       family = quasibinomial)
coef(model_encounter)
confint(model_encounter)
summary(model_encounter)
display(model_encounter)

# interaction with marker gene (gusA or celB,
# as specified in column 'wild_type'
# is non-significant, so continue model without this interaction.

# quasibinomial model (need to round because took averages
# person 1 and person 2)
# quasibinomial because of overdispersion
model_encounter <- glm(cbind(number_mixed_nodules,
        round(total_number_nodules_excl_mixed)) ~ encounter,
                       data = results_excl_WC,
                       family = quasibinomial)
coef(model_encounter)
confint(model_encounter)
summary(model_encounter)
display(model_encounter)

# frequency of mixed infection nodules increases with encounter rate.
# do some calculations to get meaningful
# measure of how much (back transform)
# See R in action around p. 321
# the regression coefficient gives the change in log(odds)
# in the respone for a
# unit change in the predictor variable.
# Exponentiate to put results on an odd scale
odds <- exp(coef(model_encounter))[2] # 1823
# with confidence intervals
odds_lower <- exp(confint(model_encounter))[2,1] #98.98
odds_upper <- exp(confint(model_encounter))[2,2] #40852.68

# so for unit increase in encounter rate, the odds of
# finding a mixed nodule
# increases 1823 times (95% CI [99, 40853])
# We are not interested in unit increase in encounter rate,
# but rather in the
# difference between highest and lowest encounter rate:
# increase in encounter rate:
increase <- (0.5*0.5) - ((1/11)*(10/11)) # 0.1673554

# so for increase in encounter rate that we have
# (from 1:10 and 10:1 to 1:1,
```

```
# which is about a 3-fold increase (increase_predictor))
increase_predictor <- (0.5*0.5) / ((10/11)*(1/11)) # 3.025
# the odds of finding a mixed nodule increased with
# factor of also about 3:
odds^increase # 3.51
# lower
odds_lower^increase # 2.1576
# upper
odds_upper^increase # 5.9117

#########################################################################
#######                  TOTAL NUMBER OF NODULES          ############
#########################################################################

# see whether total number of nodules depends on inoculation ratio
# first calculate average number of nodules
mean_total_nr_nodules <- mean(results_excl_WC$total_number_nodules) # 183
sd_total_nr_nodules <- sd(results_excl_WC$total_number_nodules) # 78
se_total_nr_nodules <- sd_total_nr_nodules / sqrt(nrow(results_excl_WC)) # 12

# plot total number of nodules as function of perc wt in inoculum
plot_total_nr <- ggplot(data = results_excl_WC,
                        aes(x = perc_wt_inoculum,
                            y = total_number_nodules)) +
  geom_point(size = 10, stroke = 4, pch = 1) +
  geom_smooth(method = lm, colour = "black", fill = "black",
              size = 3, alpha = 0.5) +
  scale_x_continuous(limits = c(0, 100)) +
  scale_y_continuous(limits = c(0, 400)) +
  theme_bw(base_size = 38) +
  theme(panel.grid.major = element_blank(),
        legend.key = element_blank(),
        axis.line.x = element_line(size = 2, color = "black"),
        axis.line.y = element_line(size = 2, color = "black"),
        legend.justification=c(1,0), legend.position=c(1,1)) +
  guides(color=guide_legend(override.aes=list(fill=NA))) +
  xlab("% fixing strain in inoculum") +
  ylab("Total number of nodules") +
  annotate('text', x = 37, y = 380,
  label = "italic(y) == 226 - 0.88~italic(x)~','~italic(R^2) == 0.14",
  size = 10, parse = TRUE)
plot_total_nr

# Export graph
tiff("figure 4.tiff", width = 10, height = 10, units = 'in',
     res = 300, compression = 'lzw')
plot_total_nr # Make plot
dev.off()

# regress total number of nodules against percentage fixing
# strain in inoculum
# (interaction with wild type gusA or celB marked
# again non-significant)
model_total_nodules <- lm(total_number_nodules ~ perc_wt_inoculum,
                          data = results_excl_WC)
coef(model_total_nodules)
display(model_total_nodules)
confint(model_total_nodules)
summary(model_total_nodules)
```

```r
# check for normality
par(mfrow = c(2,2))
plot(model_total_nodules)
par(mfrow = c(1,1))

# put the slope into perspective; what is the decrease
# in number of nodules?
# estimate of slope: -0.88 (95% CI [-1.572302, -0.185894])
# number of nodules at 1:10 wt: mutant:
(1/11) * 100 * -0.8791 + 226 # 218
# number of nodules at 10:1 wt:mutant:
(10/11) * 100 * -0.8791 + 226 # 146 (95% CI: 83;209 )
(10/11) * 100 * -1.572302 + 226 # (lower estimate slope): 83
(10/11) * 100 * -0.185894 + 226 # (upper estimate slope): 209

# difference:
218-146 #72
# lower CI:
218-209 #9
# upper CI:
218-83 # 135
# percentage decrease
(72/218) * 100 # 33%

# ##############################################################################
# ###### ACHIEVED VERSUS INTENDED INOCULATION RATIO CHECK         #######
# ##############################################################################

# check that if account for achieved (estimated) inoculation ratio,
# results don't change significantly

# calculate mean and standard error of the deviation
results_excl_WC$inoculum_deviation <-
  results_excl_WC$perc_wt_inoculum_estimated -
  results_excl_WC$perc_wt_inoculum
mean_deviation <- mean(unique(results_excl_WC$inoculum_deviation))
# 3.78371238075
sd_deviation <- sd(unique(results_excl_WC$inoculum_deviation))
# 3.65425
se_deviation <- sd_deviation / sqrt(6) # 1.49
max(unique(results_excl_WC$inoculum_deviation))

# plot percentage fixing nodules as function of estimated
# percentage fixing strain in inoculum
plot_check <- ggplot(data = results_excl_WC,
                     aes(x = perc_wt_inoculum_estimated,
                         y = perc_wt_nodules)) +
  geom_point(aes(colour = wild_type, shape = wild_type), size = 30) +
  geom_smooth(method = lm, colour = "black", fill = "black",
              size = 3, alpha = 0.6) +
  scale_colour_manual(name = "fixing strain",
       labels = c(expression(paste(italic("gusA"), " marked")),
                  expression(paste(italic("celB"), " marked"))),
                values = c("gusA" = "magenta", "celB" = "blue"),
                breaks = c("gusA", "celB")) +
  scale_shape_manual(name = "fixing strain",
       labels = c(expression(paste(italic("gusA"), " marked")),
                  expression(paste(italic("celB"), " marked"))),
```

```
                values = c("gusA" = "+", "celB" = "o"),
                breaks = c("gusA", "celB")) +
  theme_bw(base_size = 56) +
  theme(panel.grid.major = element_blank(),
        legend.key = element_blank(),
        axis.line.x = element_line(size = 2, color = "black"),
        axis.line.y = element_line(size = 2, color = "black"),
        legend.justification=c(1,0), legend.position=c(1,0)) +
  guides(color=guide_legend(override.aes=list(fill=NA))) +
  xlab("% nitrogen-fixing strain in inoculum") +
  ylab(expression(atop("% nodules containing",
                       paste("only fixing strain")))) +
  annotate('text', x = 37, y = 75,
  label = "italic(y) == -1.38 + 0.99~italic(x)~','~italic(R^2) == 0.99",
  size = 12, parse = TRUE)
plot_check

# regress % fixing nodules against % fixing strain in inoculum
# include possible interaction with marker gene (gusA or celB)
model_partner_choice <-
  lm(perc_wt_nodules ~ perc_wt_inoculum_estimated*wild_type,
                           data = results_excl_WC)
coef(model_partner_choice)
summary(model_partner_choice)
confint(model_partner_choice)
display(model_partner_choice)

# interaction with marker gene non-significant, so exclude
# for further analyses
model_partner_choice <- lm(perc_wt_nodules ~ perc_wt_inoculum,
                           data = results_excl_WC)
coef(model_partner_choice)
summary(model_partner_choice)
confint(model_partner_choice)
display(model_partner_choice)

##############################################################################
######         SHOOT WEIGHT ANALYSIS                                 ########
##############################################################################

# Analyze whether the percentage of fixing nodules
# (including 0% fixing nodules (water controls)
# affects shoot weight)
head(results)

p_weight <- ggplot(data = results) +
  geom_point(aes(x = perc_wt_inoculum, y = shoot_weight),
             size = 10, stroke = 4, pch = 1) +
  scale_y_continuous(lim = c(0, 0.5)) +
  theme_bw(base_size = 38) +
  theme(panel.grid.major = element_blank(),
        legend.key = element_blank(),
        axis.text.x = element_text(),
        # axis.title.y = element_text(margin = margin(0, 0, 0, 0)),
        axis.line.x = element_line(size = 2, color = "black"),
        axis.line.y = element_line(size = 2, color = "black")) +
  xlab("% fixing strain in inoculum") +
  ylab(expression(paste("Shoot weight (g)")))
p_weight
```

```r
# Export graph
tiff("weight.tiff", width = 10, height = 10, units = 'in',
     res = 300, compression = 'none')
p_weight # Make plot
dev.off()

m1 <- lm(shoot_weight ~ ratio, data = results)
anova(m1)
display(m1)
confint(m1)
summary(m1)

## Shoot weights from separate experiment (exp. 25) after five weeks

# clear workspace
rm(list=ls())

results <-
  read.csv("Westhoek et al., 2017, shoot mass single inoculation.csv")
results
str(results)

m1 <- lm(Weight ~ Strain, data = results)
anova(m1)
display(m1)
confint(m1)
summary(m1)

results$Strain <- relevel(results$Strain, ref = "RU3940")
m1 <- lm(Weight ~ Strain, data = results)
anova(m1)
display(m1)
confint(m1)
summary(m1)

#########################################################################
#######         NODULE SIZE                                  ##########
#########################################################################

# code to analyze nodule size data of
# partner choice experiment (exp. 8)

# clear workspace
rm(list=ls())

# read in the results for nodule size data
# partner choice experiment (exp. 8)
results_size <- read.csv("Westhoek et al., 2017, nodule size data.csv")
head(results_size)
str(results_size)
# relevel to show non-fixing first on the graph
results_size$type <- relevel(results_size$type, ref = "non-fixing")
str(results_size)

# add plant number from label
results_size$plant <- strtoi(sub("_.*$", "", results_size$Label ))
# convert Area in cm2 to area in mm2
results_size$area_mm2 <- results_size$Area * 100
```

```r
# calculate average nodule size per type/colour per plant
results_size_means <- ddply(results_size, c("plant", "type", "colour"),
                            summarize,
                            nodule_size = mean(area_mm2))

# plot the results
p_nodule_size <- ggplot(data = results_size_means) +
  geom_point(aes(x = type, y = nodule_size),
             position = position_jitter(w = 0.1),
             size = 10, stroke = 4, pch = 1) +
  scale_y_continuous(limits = c(0, 2)) +
  theme_bw(base_size = 38) +
  theme(panel.grid.major = element_blank(),
        legend.key = element_blank(),
        axis.line.x = element_line(size = 2, color = "black"),
        axis.line.y = element_line(size = 2, color = "black")) +
  guides(color=guide_legend(override.aes=list(fill=NA))) +
  xlab("Strain") +
  ylab(expression(paste("Nodule area (mm"^"2", ")")))
p_nodule_size

# Export graph
tiff("figure 3.tiff", width = 10, height = 10, units = 'in',
     res = 300, compression = 'lzw')
p_nodule_size # Make plot
dev.off()

# do paired t-test to see whether nodule size of fixing strain is
# significantly larger than nodule size of non-fixing strain
# paired t-test, because within plant is not independent
# (so pair fixing and non-fixing of each plant)

# first get data in correct format
# change name of type 'non-fixing' to 'non_fixing'
results_size_means$type <-
  ifelse(results_size_means$type == "non-fixing", "non_fixing", "fixing")
# reshape the data to wide format
m_results_size_means <- melt(results_size_means,
                             id.vars = c("plant", "type", "colour"),
                             measure.vars = "nodule_size")
wide_results_size_means <- cast(m_results_size_means,
                                plant ~ variable + type)

# do paired t-test
t.test(wide_results_size_means$nodule_size_fixing,
       wide_results_size_means$nodule_size_non_fixing,
       paired = TRUE)

# calculate mean and standard error of fixing nodules
mean_fixing <- mean(wide_results_size_means$nodule_size_fixing) #1.199286
sd_fixing <- sd(wide_results_size_means$nodule_size_fixing) # 0.2695713
se_fixing <- sd_fixing / sqrt(nrow(wide_results_size_means)) # 0.072

#########################################################################
#######                NUMBER OF MIXED NODULES           ##############
#########################################################################

# code to analyze number of mixed nodules as function of
```

```r
# inoculation density (exp. 9)
# how does frequency of mixed nodules depend on inoculation density?

# clear workspace
rm(list=ls())
# read in the results for data number of mixed nodules (exp. 9)
results_nr_mixed <-
  read.csv("Westhoek et al., 2017, number mixed nodules.csv")
head(results_nr_mixed)
str(results_nr_mixed)

# some data processing
# calculate the average number of gusA and celB nodules
# (average two people who counted)
results_nr_mixed$Nr_gusA_nodules <-
  (results_nr_mixed$Nr_gusA_nodules_person1 +
  results_nr_mixed$Nr_gusA_nodules_person2) / 2
results_nr_mixed$Nr_celB_nodules <-
  (results_nr_mixed$Nr_celB_nodules_person1 +
       results_nr_mixed$Nr_celB_nodules_person2) / 2
# calculate the percentage of mixed nodules
results_nr_mixed$perc_mixed <- (results_nr_mixed$Nr_mixed_nodules /
    (results_nr_mixed$Nr_gusA_nodules +
       results_nr_mixed$Nr_celB_nodules +
              results_nr_mixed$Nr_mixed_nodules)) * 100

head(results_nr_mixed)

# plot the data
p_perc_mixed <- ggplot(data = results_nr_mixed,
                       aes(x = Density, y = perc_mixed)) +
  geom_point(size = 10, stroke = 4, pch = 1) +
  geom_smooth(method = lm, colour = "black", fill = "black",
              size = 3, alpha = 0.5) +
  scale_x_continuous(trans = 'log10',
                     breaks = trans_breaks("log10", function(x) 10^x),
                     labels = trans_format("log10", math_format(10^.x))) +
  theme_bw(base_size = 38) +
  theme(panel.grid.major = element_blank(),
        legend.key = element_blank(),
        axis.line.x = element_line(size = 2, color = "black"),
        axis.line.y = element_line(size = 2, color = "black")) +
  annotate('text', x = 10000, y = 20,
  label = "italic(y) == 0.03 + 1.87~italic(x)~','~italic(R^2) == 0.53",
  size = 10, parse = TRUE) +
  xlab("Inoculation density (CFUs)") +
  ylab("% mixed nodules")
p_perc_mixed

# Export graph
tiff("perc_mixed.tiff", width = 10, height = 10,
     units = 'in', res = 300, compression = 'none')
p_perc_mixed # Make plot
dev.off()

# fit linear regression to percentage of mixed nodules data
# as function of log of inoculation
# density (without partner choice data). Include quadratic term.
results_nr_mixed$log_Density <- log10(results_nr_mixed$Density)
```

```r
results_nr_mixed$log_Density2 <- results_nr_mixed$log_Density^2

m1 <- lm(perc_mixed ~ log_Density + log_Density2,
         data = results_nr_mixed)
summary(m1)
anova(m1)
display(m1)
confint(m1)

# Quadratic term non-significant so take it out
m2 <- lm(perc_mixed ~ log_Density, data = results_nr_mixed)
summary(m2)
anova(m2)
display(m2)
confint(m2)

#############################################################
##### AREA OCCUPIED BY FIXING AND                    #######
##### NON-FIXING STRAINS WITHIN MIXED NODULES        #######
#############################################################

# code to analyze area occupied by fixing and non-fixing
# strain within mixed nodules
# data on area of fixing and non-fixing strain within mixed
# nodules from
# three different experiments. (exp. 8 (partner choice),
# exp. 9 (effect inoculation density
# on frequency of mixed nodules), exp. 11). Exp. 11
# just had 6 plants at inoculation density of 10^8 CFUs with fixing
# strain celB marked (exp. 9 had fixing strain gusA marked,
# so exp. 11 was just to check for effect of marker gene)

# clear workspace
rm(list=ls())

# read in the results for data area within mixed nodules
results <- read.csv("Westhoek et al., 2017, mixed nodule areas.csv")
head(results)
str(results)

## some data processing

# add column to indicate for which strain the area was measured
# if Marker_Fixing_Strain = gusA and Color = magenta: fixing
# if Marker_Fixing_Strain = celB and Color = blue: fixing
# otherwise non-fixing. If white: "unknown"
results$Strain <-
  as.factor(ifelse(results$Color == "white", "unknown",
                   ifelse((results$Marker_Fixing_Strain == "gusA" &
                      results$Color == "magenta"), "fixing",
                   ifelse((results$Marker_Fixing_Strain == "celB" &
                         results$Color == "blue"), "fixing",
                                      "non-fixing"))))

# add column with 'Picture' which is label without '.tif'
results$Label <- as.character(results$Label)
results$Picture <- sapply(strsplit(results$Label,
                                   split='.',
                    fixed=TRUE), function(x) (x[1]))
```

```r
head(results)

# add column with unique Plant_ID, by combining
# plant number (first number of Picture)
# with Experiment
results$Plant_ID <- paste(sapply(strsplit(results$Picture,
                                          split='_',
                         fixed=TRUE), function(x) (x[1])),
                           results$Experiment, sep = "_")
head(results)

# add column with unique Nodule_ID: combination of
# Experiment, Picture and first letter of Comments
# (because if multiple nodule were measured in same picture,
# they were indicated in the comments section
# as top/bottom/left/right/front)
results$Nodule_ID <- paste(results$Experiment,
          results$Picture, as.character(results$Comments),
                            sep = "_")
head(results)

# calculate for each nodule the total area for each of blue,
# magenta and white
# (some nodules had multiple blobs of a given color)
results_sum <- results %>%
  group_by(Nodule_ID, Color, Experiment, Plant_ID, Strain,
           Marker_Fixing_Strain) %>%
  summarise(Area = sum(Area))  %>%
  as.data.frame()
dim(results)
dim(results_sum)
head(results_sum)

## Data exploration
table(results_sum$Strain)
# 220 mixed infection nodules measured in total,
# all of which obviously have fixing and non-fixing area
# 60 of them also have white area
length(unique(results_sum$Plant_ID))
# Nodules measured of 38 plants
table(results_sum$Experiment)
# Across three indepent experiments
table(results_sum$Marker_Fixing_Strain)
# More than three times more nodules with
# gusA-marked fixing strain than
# with celB-marked fixing strain

## Data analysis

# calculate for each nodule the total area (blue + magenta + white),
# and the total the area excluding white (blue + magenta)
results_sum <- results_sum %>%
  group_by(Nodule_ID) %>%
  mutate(Total_area = sum(Area),
         Total_area_excl_white = sum(Area[Color!="white"])) %>%
  as.data.frame()
head(results_sum)

# drop the rows with Color = white, because of interest is the relative
```

```r
# fractions of blue vs magenta
results_sum_excl_white <- subset(results_sum, results_sum$Color != "white")
dim(results_sum)
dim(results_sum_excl_white)
# indeed 60 rows less

# calculate percentage of each strain as part of
# Total_area_excl_white
results_sum_excl_white$Percentage <-
  (results_sum_excl_white$Area /
        results_sum_excl_white$Total_area_excl_white) * 100
head(results_sum_excl_white)

# check that all percentages add up to 100%
check <- results_sum_excl_white %>%
  group_by(Nodule_ID) %>%
  summarise(Percentage_check = sum(Percentage)) %>%
  as.data.frame()
unique(check$Percentage_check)

# Do analyses on fraction of fixing strain
# (as non-fixing is just the complement)
final_data <- subset(results_sum_excl_white,
                     results_sum_excl_white$Strain == "fixing")
# Make sure variables are correctly identified as factors/numerical etc.
final_data$Experiment <- as.factor(final_data$Experiment)
final_data$Plant_ID <- as.factor(final_data$Plant_ID)
final_data$Marker_Fixing_Strain <- relevel(final_data$Marker_Fixing_Strain,
                                           ref = "gusA")
head(final_data)
str(final_data)

## Plot the results
p_percentage <- ggplot(data = final_data) +
  geom_point(aes(x = Marker_Fixing_Strain, y = Percentage),
             size = 10, stroke = 4, pch = 1) +
  theme_bw(base_size = 38) +
  theme(panel.grid.major = element_blank(),
        legend.key = element_blank(),
        axis.text.x = element_text(face = "italic"),
        # axis.title.y = element_text(margin = margin(0, 0, 0, 0)),
        axis.line.x = element_line(size = 2, color = "black"),
        axis.line.y = element_line(size = 2, color = "black")) +
  xlab("Fixing strain marked with") +
  ylab(expression(paste("% of nodule area with fixing strain")))
p_percentage


# Export the graph
tiff("nodule occupancy.tiff", width = 10, height = 10,
     units = 'in', res = 300, compression = 'none')
p_percentage # Make plot
dev.off()

## Models

# Model the percentage of fixing strain in a mixed nodule as function of
# which marker was used to mark fixing strain, accounting for plant_ID
# Not accounting for Experiment, as overlaps with which marker was used
```

```
# to mark fixing strain (Marker_Fixing_Strain).
# (In exp. 8 markers are both ways round,
# but most datapoints are from Exp. 9 (gusA-marked
# fixing strain) and Exp. 11 (celB-marked fixing strain)).
(m1 <- lmer(Percentage ~ Marker_Fixing_Strain + (1 | Plant_ID),
            data = final_data))
display(m1)
summary(m1)
anova(m1)

(m2 <- lm(Percentage ~ Marker_Fixing_Strain + Plant_ID,
          data = final_data))
anova(m2)

# Plant_ID accounts for small percentage of variation,
# so can take nodule as unit of replication
(m3 <- lm(data = final_data, Percentage ~ Marker_Fixing_Strain))
display(m3)
anova(m3)
confint(m3)
summary(m3)

# get confidence interval of percentage when celB marked
m_celB <- lm(data = subset(final_data,
                           final_data$Marker_Fixing_Strain == "celB"),
             Percentage ~ 1)
display(m_celB)
confint(m_celB)

#########################################################################
######        ACETYLENE REDUCTIONS                              ########
#########################################################################

# clear workspace
rm(list=ls())

# read in the results for nodule number data
# partner choice experiment (exp. 8)
results <- read.csv("Westhoek et al., 2017, acetylene reductions.csv")
head(results)
str(results)
results$Strain <- relevel(results$Strain, ref = "WC")

# do the calculations to convert Perc_ethylene and
# Perc_acetylene data to how much
# acetylene has been reduced.
# First, calculate the initial amount (in umol) of acetylene added:
# 2% acetylene in 0.320 L bottle.
# At 25 degrees Celsius and normal pressure 1 mol
# of acetylene takes up 24.465 L. So:
results$uMol_Acetylene_Added <- (0.02 * 0.320 / 24.465) * 1000000
# Then, calculate the percentage of acetylene converted to ethylene.
# Perc_ethylene and Perc_acetylene give
# the percentage of ethylene and acetylene (in terms of moles)
# measured in sample,
# as derived from peak area by GC.
results$Perc_Converted <- (results$Perc_ethylene /
        (results$Perc_ethylene + results$Perc_acetylene)) * 100
# The amount (in umols) of ethylene produced (per jar per hour,
```

```
# as incubation time
# was one hour) is the percentage of acetylene converted
# multiplied by the initial amount of acetylene added.
results$uMol_Ethylene_Produced <- (results$Perc_Converted / 100) *
  results$uMol_Acetylene_Added
head(results)

# calculate the average of three technical replicates
# to get average value and standard error per plant
means <- results %>%
  group_by(Strain, Plant) %>%
  summarise(
    N = sum(!is.na(uMol_Ethylene_Produced)),
    Mean_Ethylene_Produced = mean(uMol_Ethylene_Produced,
                                  na.rm = TRUE),
    sd = sd(uMol_Ethylene_Produced, na.rm = TRUE),
    se = sd / N)  %>%
  as.data.frame()
means$Strain <- relevel(means$Strain, ref = "RU3940")
(means)

# plot the results
p_acetylene <- ggplot(data = means) +
  geom_point(aes(x = Strain, y = Mean_Ethylene_Produced),
             size = 10, stroke = 4, pch = 1) +
  theme_bw(base_size = 38) +
  theme(panel.grid.major = element_blank(),
        legend.key = element_blank(),
        axis.text.x = element_text(face = "italic",
                                   angle = 90, hjust = 1),
        axis.title.x = element_blank(),
        # axis.title.y = element_text(margin = margin(0, 0, 0, 0)),
        axis.line.x = element_line(size = 2, color = "black"),
        axis.line.y = element_line(size = 2, color = "black")) +
  # xlab("Strain") +
  ylab(expression(paste("Ethylene produced \n per plant per hour")))
p_acetylene

# Export the graph
tiff("acetylene.tiff", width = 10, height = 10,
     units = 'in', res = 300, compression = 'none')
p_acetylene # Make plot
dev.off()

means$Strain <- relevel(means$Strain, ref = "3841")
m1 <- lm(means$Mean_Ethylene_Produced ~ means$Strain)
confint(m1)
summary(m1)

means$Strain <- relevel(means$Strain, ref = "RU3940")
m1 <- lm(means$Mean_Ethylene_Produced ~ means$Strain)
confint(m1)
summary(m1)

RU3940 <- subset(means, means$Strain == "RU3940")
(mean(RU3940$Mean_Ethylene_Produced))
sd <- sd(RU3940$Mean_Ethylene_Produced)
se <- sd / sqrt(3)
se
```

Appendix C. Supplementary material chapter 4: Mathematical modelling of the legume-rhizobia symbiosis

Overview

The supplementary material of Chapter 4 consists of an excerpt of the R code used to implement the model. The excerpt shows the approach followed, but the full version of the R code is deposited in the Oxford Research Archive (ORA-Data) repository, at DOI: 10.5287/bodleian:BDVypApax.

R code

Model analysis resource allocation among strains

Annet Westhoek

August 17, 2017

```
# load libraries
library(deSolve)
library(ggplot2)
library(plyr)
require(gridExtra)

## Loading required package: gridExtra

print(sessionInfo(), locale = FALSE)

## R version 3.3.2 (2016-10-31)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 15063)
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] gridExtra_2.2.1 plyr_1.8.4      ggplot2_2.2.1   deSolve_1.14
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_0.12.9      assertthat_0.1   digest_0.6.12    rprojroot_1.2
##  [5] grid_3.3.2       gtable_0.2.0     backports_1.0.5  magrittr_1.5
##  [9] evaluate_0.10    scales_0.4.1     stringi_1.1.2    lazyeval_0.2.0
```

```
## [13] rmarkdown_1.3    tools_3.3.2      stringr_1.1.0    munsell_0.4.3
## [17] yaml_2.1.14      colorspace_1.3-2 htmltools_0.3.5  knitr_1.15.1
## [21] tibble_1.2

# last update August 17, 2017
# Annet Westhoek, 2017

# This code is used to model plant growth and
# resource allocation to rhizobia, under
# different soil nitrogen conditions.

## PART A: model functions
#     - Single strain model constant soil N
#     - Two strain model constant soil N
#     - Two strain model dynamic soil N
# Constant soil N means that the soil nitrogen concentration is constant over
# the time course of the simulation, but the concentration
# may of course be varied in different simulations.
# Versions presented here use analytically derived
# optimal values for parameters f (allocation to aboveground biomass)
# and a (allocation to nodules).
# Versions in which the user specifies the values for f and a
# are simpler and can be found in the full version of the code.

## PART B: code to run the model
# The version presented here is to run a single simulation
# To run multiple simulations, see the full version of the code.


########################
##### PART A: MODEL FUNCTIONS
########################

#########################################################
# SINGLE STRAIN MODEL, CONSTANT SOIL NITROGEN
#########################################################
single_strain_model <- function(gamma, rho, theta, Nconc, m,
                                dleaf, droot, dnod,
                                Mleaf_init, Mroot_init, Mnod_init,
                                time_step, time_end){

  ###################################
  # MODEL DEFINITION
  ###################################

  model <- function(t, state, params){
    with(as.list(c(state, params)),{
      # C limited growth rate
      G_c <<- gamma * Mleaf - dleaf * Mleaf - droot * Mroot - dnod * Mnod
      # N limited growth rate
      G_n <<- rho * theta * Mroot * Nconc + rho * m * Mnod
      # final growth rate is minimum G
      G <<- min(G_c, G_n)

      # calculate the optimal allocation to the nodules
      a <<- ifelse(dnod >= (dleaf-gamma) +
                     m*((droot+gamma-dleaf)/(theta*Nconc)), 0, 1)

      # calculate the optimal allocation to aboveground biomass f
```

```
      f <<- (droot + rho*theta*Nconc +
               a*(dnod - droot + rho*(m - theta*Nconc))) /
        (gamma - dleaf + droot + rho*theta*Nconc +
           a*(dnod - droot + rho*(m - theta*Nconc)))

      # calculate growth for seperate tissues
      dMleaf <- f * G
      dMroot <- (1-f) * (1-a) * G
      dMnod  <- (1-f) * a * G

      # save the results
      der <<- c(dMleaf, dMroot, dMnod)
      list(der, G_c, G_n, G, gamma, rho, theta,
           Nconc, m, dleaf, droot, dnod, f, a)
    })
  }

  ############################
  # MAIN PROGRAM
  ############################

  ### INITIAL VALUES, PARAMETER SETTINGS AND TIME COURSE
  # initial values: initial values as specified in part B
  state <- c(Mleaf=Mleaf_init, Mroot=Mroot_init, Mnod=Mnod_init)

  # parameter values: parameter values as specified in part B
  params <- c(gamma=gamma, rho=rho, theta=theta, Nconc=Nconc, m=m,
              dleaf=dleaf, droot=droot, dnod=dnod)

  # time for simulation: as specified in part B
  dt    <- seq(0,time_end,time_step)

  ### SIMULATE THE MODEL
  simulation <<- as.data.frame(ode(y = state, times = dt,
                                   func = model, parms = params))

  # give names to the columns that were added to the results
  names(simulation)[5:17] <- c("G_c", "G_n", "G", "gamma", "rho", "theta",
                               "Nconc", "m", "dleaf", "droot", "dnod", "f", "a
")

  # make the simulation available outside function
  simulation <<- simulation
}

#########################################################
# TWO STRAIN MODEL, CONSTANT SOIL NITROGEN
#########################################################
double_strain_model <- function(gamma, rho, theta, Nconc, m1, m2,
                                dleaf, droot, dnod1, dnod2,
                                Mleaf_init, Mroot_init, Mnod1_init, Mnod2_init
,
                                time_step, time_end){

  ####################################
  # MODEL DEFINITION
  ####################################

  model <- function(t, state, params){
```

```
  with(as.list(c(state, params)),{
    # C limited growth rate
    G_c <- gamma * Mleaf - dleaf * Mleaf -
      droot * Mroot - dnod1 * Mnod1 - dnod2 * Mnod2
    # N limited growth rate
    G_n <- rho * theta * Mroot * Nconc +
      rho * m1 * Mnod1 + rho * m2 * Mnod2
    # final growth rate is minimum G
    G <- min(G_c, G_n)

    # calculate the allocation to the nodules
    a1 <- ifelse(dnod1 >= (dleaf-gamma) +
                   m1*((droot+gamma-dleaf)/(theta*Nconc)), 0, 1)
    a2 <- ifelse(dnod2 >= (dleaf-gamma) +
                   m2*((droot+gamma-dleaf)/(theta*Nconc)), 0, 1)

    # if both strains are better than root:
    # if both strains are identical, give 0.5 to each,
    # otherwise choose one over other
    # have to round otherwise equality doesn't work
    if(a1 == 1 & a2 == 1){
      a1 <- ifelse(round(dnod1, 10) == round(dleaf - gamma +
                                   m1 * ((gamma-dleaf+dnod2)/m2), 10), 0.5,
                   ifelse(round(dnod1,10) > round(dleaf - gamma +
                                   m1 * ((gamma-dleaf+dnod2)/m2), 10), 0,1))
      a2 <- 1 - a1
    }

    # calculate the allocation to aboveground biomass f
    f <- (droot + rho*theta*Nconc +
            a1*(dnod1 - droot + rho*(m1 - theta*Nconc)) +
            a2*(dnod2 - droot + rho*(m2 - theta*Nconc)) ) /
      (gamma - dleaf + droot + rho*theta*Nconc +
         a1*(dnod1 - droot + rho*(m1 - theta*Nconc)) +
         a2*(dnod2 - droot + rho*(m2 - theta*Nconc)) )

    # calculate growth for seperate tissues
    dMleaf <- f * G
    dMroot <- (1-f) * (1-a1-a2) * G
    dMnod1  <- (1-f) * a1 * G
    dMnod2 <- (1-f) * a2 * G

    # save the results
    der <- c(dMleaf, dMroot, dMnod1, dMnod2)
    list(der, G_c, G_n, G, gamma, rho, theta, Nconc, m1, m2,
         dleaf, droot, dnod1, dnod2, f, a1, a2)
  })
}

############################
# MAIN PROGRAM
############################

### INITIAL VALUES, PARAMETER SETTINGS AND TIME COURSE
# initial values: initial values
# as specified in part B
state <- c(Mleaf=Mleaf_init, Mroot=Mroot_init,
           Mnod1=Mnod1_init, Mnod2=Mnod2_init)
```

```
  # parameter values: parameter values
  # as specified in part B
  params <- c(gamma=gamma, rho=rho, theta=theta, Nconc=Nconc, m1=m1, m2=m2,
              dleaf = dleaf, droot = droot, dnod1 = dnod1, dnod2=dnod2)

  # time for simulation: as specified in part B
  dt    <- seq(0,time_end,time_step)

  ### SIMULATE THE MODEL
  simulation <<- as.data.frame(ode(y = state, times = dt, func = model,
                                   parms = params))

  # give names to the columns that were added to the results
  names(simulation)[6:21] <- c("G_c", "G_n", "G", "gamma", "rho", "theta",
                               "Nconc", "m1", "m2", "dleaf", "droot",
                               "dnod1", "dnod2", "f", "a1", "a2")

  # make the simulation available outside function
  simulation <<- simulation

}

########################################################
# TWO STRAIN MODEL, DYNAMIC SOIL NITROGEN
########################################################
double_strain_model_dyn_N <- function(gamma, rho, theta, Nconc_init, phi, V,
                                      m1, m2, dleaf, droot, dnod1, dnod2,
                                      Mleaf_init, Mroot_init,
                                      Mnod1_init, Mnod2_init,
                                      time_step, time_end){

    ###################################
    # MODEL DEFINITION
    ###################################

    model <- function(t, state, params){
      with(as.list(c(state, params)),{
        # C limited growth rate
        G_c <<- gamma * Mleaf - dleaf * Mleaf -
          droot * Mroot - dnod1 * Mnod1 - dnod2 * Mnod2
        # N limited growth rate
        G_n <<- rho * theta * Mroot * Nconc +
          rho * m1 * Mnod1 + rho * m2 * Mnod2
        # final growth rate is minimum G
        G <<- min(G_c, G_n)

        # calculate the allocation to the nodules
        a1 <<- ifelse(dnod1 >= (dleaf-gamma) +
                        m1*((droot+gamma-dleaf)/(theta*Nconc)), 0, 1)
        a2 <<- ifelse(dnod2 >= (dleaf-gamma) +
                        m2*((droot+gamma-dleaf)/(theta*Nconc)), 0, 1)

        # if both strains are better than root:
        # if both strains are identical,
        # give 0.5 to each, otherwise choose one over other
        # have to round otherwise equality doesn't work
        if(a1 == 1 & a2 == 1){
          a1 <<-
            ifelse(round(dnod1, 10) == round(dleaf - gamma +
```

```
                                              m1 * ((gamma-dleaf+dnod2)/m2),
10), 0.5,
                  ifelse(round(dnod1,10) > round(dleaf - gamma +
                                              m1 * ((gamma-dleaf+dnod2)/m2),
10), 0,1))
          a2 <<- 1 - a1
        }

        # calculate the allocation to aboveground biomass f
        f <<- (droot + rho*theta*Nconc +
                 a1*(dnod1 - droot + rho*(m1 - theta*Nconc)) +
                 a2*(dnod2 - droot + rho*(m2 - theta*Nconc)) ) /
          (gamma - dleaf + droot + rho*theta*Nconc +
             a1*(dnod1 - droot + rho*(m1 - theta*Nconc)) +
             a2*(dnod2 - droot + rho*(m2 - theta*Nconc)) )

        # calculate growth for seperate tissues
        dMleaf <- f * G
        dMroot <- (1-f) * (1-a1-a2) * G
        dMnod1  <- (1-f) * a1 * G
        dMnod2 <- (1-f) * a2 * G
        dNconc <- phi - ((theta*Mroot*Nconc) / V)

        # save the results
        der <<- c(dMleaf, dMroot, dMnod1, dMnod2, dNconc)
        list(der, G_c, G_n, G, gamma, rho, theta, phi, V, m1, m2,
             dleaf, droot, dnod1, dnod2, f, a1, a2)
      })
    }

    ############################
    # MAIN PROGRAM
    ############################

    ### INITIAL VALUES, PARAMETER SETTINGS AND TIME COURSE
    # initial values: initial values as specified in part B
    state <- c(Mleaf=Mleaf_init, Mroot=Mroot_init, Mnod1=Mnod1_init,
               Mnod2=Mnod2_init, Nconc = Nconc_init)

    # parameter values: parameter values
    # as specified in part B
    params <- c(gamma=gamma, rho=rho, theta=theta, phi=phi, V=V,
                m1=m1, m2=m2, dleaf = dleaf, droot = droot,
                dnod1 = dnod1, dnod2=dnod2)

    # time for simulation: as specified in part B
    dt    <- seq(0,time_end,time_step)

    ### SIMULATE THE MODEL
    simulation <<- as.data.frame(ode(y = state,
                                     times = dt, func = model, parms = params)
)

    # give names to the columns that were added to the results
    names(simulation)[7:23] <- c("G_c", "G_n", "G", "gamma", "rho", "theta",
                                 "phi", "V", "m1", "m2","dleaf", "droot",
                                 "dnod1", "dnod2", "f", "a1", "a2")

    # make the simulation available outside function
```

```
    simulation <<- simulation

  }


#######################
##### PART B: RUNNING THE MODELS
#######################
# The example provided is for the most complex model
# (two strains, dynamic soil nitrogen)

######## RUN A SINGLE SIMULATION AND SAVE RESULTS

# provide parameter values
gamma <- 0.5    # C uptake rate (photosynthesis)
                # unit: per day
rho   <- 7.0    # C:N ratio
                # unit: gC / gN (dimensionless)
theta <- 2.0    # N uptake rate (by root)
                # unit: cm3 per mg per day
Nconc <- 0.05   # (initial) N concentration in soil
                # unit: mg per cm3
phi   <- 0      # fertilisation rate
                # unit: mg per cm3 per day
                # only needed in dynamic soil N model
V     <- 500    # pot volume
                # unit: cm3
                # only needed in dynamic soil N model
m1     <- 0.05  # N fixation rate (by strain 1)
                # unit: mg N per mg nodule per day
m2     <- 0.05  # N fixation rate (by strain 2)
                # unit: mg N per mg nodule per day
                # only needed in two strain models
dleaf <- 0.2    # leaf maintenance
                # unit: mg C per mg leaf per day
droot <- 0.2    # root maintenance
                # unit: mg C per mg root per day
e1     <- 6     # efficiency of N fixation strain 1
                # unit: mg C / mg N
e2     <- 6     # efficiency of N fixation strain 2
                # unit: mg C / mg N
dnod1  <- e1*m1 # nodule maintenance strain 1
                # unit: mg C per mg nodule per day
dnod2  <- e2*m2 # nodule maintenance strain 2     u
                # nit: mg C per mg nodule per day

# provide initial conditions
Mseed <- 200      # start with 200 mg seed
Mleaf_init <-     # inital leaf mass has to satisfy this condition
                  # when assuming optimal allocation function f
  ((rho*theta*Nconc_init + droot) / (gamma - dleaf +
                                       rho*theta*Nconc_init + droot)) * Mseed
Mroot_init <-     # initial root mass has to satisfy this condition
                  # when assuming optimal allocation function f
  ((gamma - dleaf) / (rho*theta*Nconc_init +
                        droot + gamma - dleaf)) * Mseed
Mnod1_init <- 0     # at germination no nodules of strain 1
Mnod2_init <- 0     # at germination no nodules of strain 2
```

```r
# provide time steps and end time of simulation
# (this determines times of output only, not the timestep of integration)
time_step <- 1
time_end  <- 28

# run the model, given the parameter values,
# initial conditions and simulation time provided

double_strain_model_dyn_N(gamma, rho, theta, Nconc_init, phi, V,
                          m1, m2, dleaf, droot, dnod1, dnod2,
                          Mleaf_init, Mroot_init,
                          Mnod1_init, Mnod2_init,
                          time_step, time_end)

# the output is given as table (simulation)
head(simulation)

# check that model is OK and print warnings
if(any(simulation$f < 0 | simulation$f > 1)){
  print("WARNING: f outside limits")
}

# save simulation results to csv file
name_file_csv <- paste(getwd(),
                       "/simulation.csv", sep = "")
write.csv(simulation, file = name_file_csv, row.names = FALSE)

################################################################################
################################################################################
################################################################################
```

Appendix D. Supplementary material chapter 5: Resource allocation in the legume-rhizobia symbiosis

Overview

The supplementary material of chapter 5 consists of:

1. Results based on randomly selected (rather than most developed) nodules.
2. ReadMe document for data
3. R code used to analyse the data.

The data and R code that support the findings of this chapter are deposited in the Oxford Research Archive (ORA-Data) repository, at DOI: 10.5287/bodleian:BDVypApax.

1. Results based on randomly selected nodules

In single inoculations, randomly selected nodules of the intermediate-fixing strain averaged 1.24 mm^2, 95 % CI [1.06, 1.42], which is larger, but not quite significantly so, than Fix^- nodules (a difference of 0.21 mm^2, 95% CI [-0.06, 0.47], and significantly smaller than Fix^+ nodules (a difference of 0.29 mm^2, 95% CI [0.04, 0.54] (**Figure 1**). These results are in line with our findings based on measuring the most developed nodules.

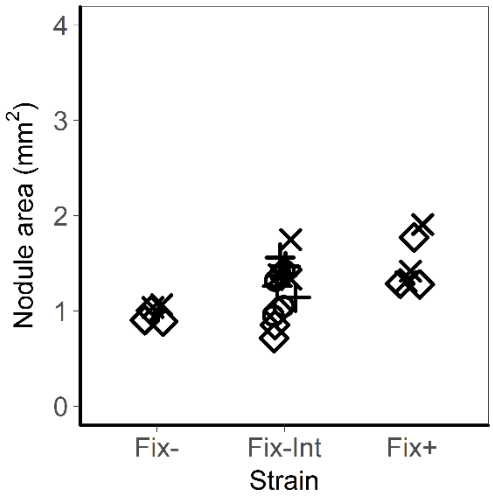


Figure 1. Nodule sizes of strains differing in effectiveness in single inoculations (Fix^-; OPS0365, Fix^{Int}; LMB827, Fix^+; Rlv3841 *gusA*). Points represent an average of ten nodules. Symbols indicate data from independent experiments. Plants were grown for 28 days without external nitrogen.

In line with results based on the most developed nodules, average size of randomly selected intermediate-fixing nodules was decreased when co-inoculated with the fixing strain as compared to when co-inoculated with itself (a difference of 0.48 mm^2, 95% CI [0.29, 0.66]). Co-inoculation with the non-fixing strain led to slightly larger nodules of the intermediate-fixing strain (a difference of 0.29 mm^2, 95% CI [0.09, 0.46]) (**Figure 2**). Such 'compensation' in nodule size when fewer effective nodules are present has been reported before[1], but was not observed based on the ten most developed nodules. For comparison, average nodule sizes of randomly selected nodules of the Fix^- and Fix^+ strain are shown (**Figure 3**).

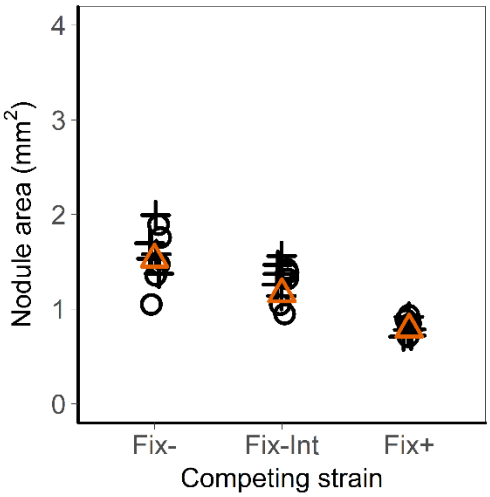


Figure 2. Nodule area of the intermediate-fixing strain (LMB827) when co-inoculated with a non-fixing strain (OPS0365), with itself, or with the fixing strain (Rlv3841 *gusA*). The overall mean is indicated with an orange triangle. Points represent an average of ten nodules. Symbols indicate data from independent experiments. Plants were grown for 28 days without external nitrogen.

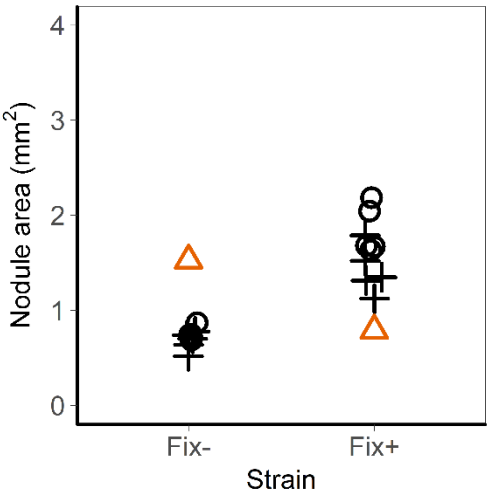


Figure 3. Nodule area of the non-fixing strain (OPS0365) and the fixing strain (Rlv3841 *gusA*), both when co-inoculated with the intermediate-fixing strain (LMB827). The mean size of the intermediate-fixing strain, as measured on the same plants, is indicated as an orange triangle for reference. Points represent an average of ten nodules. Symbols indicate data from independent experiments. Plants were grown for 28 days without external nitrogen.

In line with results based on the most developed nodules, average size of randomly selected nodules decreased with increasing soil nitrogen concentrations ($F_{1,79}$ = 76.31602, $p < 0.001$), and at high soil nitrate concentrations all strains became similar in size (**Figure 4**).

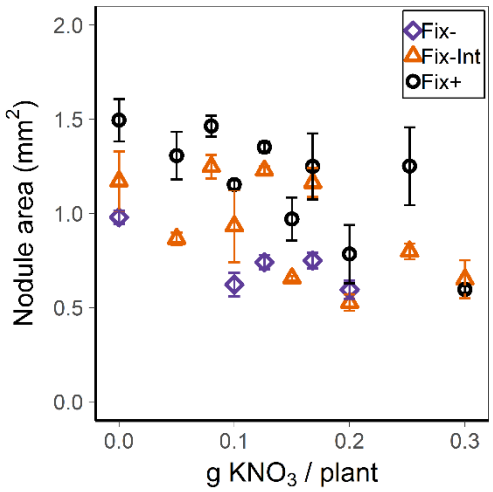


Figure 4. Nodule sizes decrease as a function of soil nitrate availability. The effect was most pronounced for the fixing strain (black circles), which at zero nitrogen forms the largest nodules, so that at high nitrate concentrations all nodules are similar in size. Nodule size for each plant was based on the average of ten nodules. Points represent mean ± s.e. of 3-6 replicate plants grown for 27-29 days. KNO_3 was added at planting in the stated amounts. Fix^-; OPS0365, Fix^{Int}; LMB827, Fix^+; Rlv3841 *gusA*.

References:

1. Singleton, W. & Stockinger, K. R. Compensation against ineffective nodulation in soybean. *Crop Sci.* **23,** 69–72 (1983).

2. ReadMe document for data

Data gathered at University of Oxford, May 2016 – July 2017.

Data comes from seven independent experiments (Table 2).

Table 2. Overview of experimental designs and type of data gathered.

Exp.	Name	Design	Strains	Measurements done
12	Soil nitrate A	Single inoculations of each strain at different nitrogen concentrations	Fix^{-}; OPS0365 Fix^{Int}; LMB827 Fix^{+}; Rlv3841 *gusA*	Nodule number, nodule area based on ten largest and on ten randomly selected nodules
13	Dual inoculations A	1:1 inoculation of LMB827 (Fix^{Int}) against each of the strains	Fix^{-}; OPS0365 Fix^{Int}; LMB827 Fix^{+}; Rlv3841 *gusA*	For each strain: nodule number, nodule area based on ten largest and on ten randomly selected nodules
14	Dual inoculations B	Idem dual inoculations A	Fix^{-}; OPS0365 Fix^{Int}; LMB827	Idem dual inoculations A

			Fix^{+}; Rlv3841 *gusA*	
16	Soil nitrate B	Idem soil nitrate A	Fix^{-}; OPS0365 Fix^{Int}; LMB827 Fix^{+}; Rlv3841 *gusA*	Idem soil nitrate A
18	Triple inoculations	1:1:1 inoculation of all strains	Fix^{-}; OPS0366 Fix^{Int}; LMB827 Fix^{+}; Rlv3841 *gusA*	For each strain: nodule number, nodule area based on ten largest nodules
21	Acetylene reductions	Idem dual inoculations but in split-root design. Also did Fix^{+} against each of the strains	Fix^{-}; OPS0365 Fix^{Int}; LMB827 Fix^{+}; Rlv3841 *gusA*	For each half-root: number of nodules and size of ten largest nodules. Only for focal half-root: acetylene reductions.
28	Different efficiency	Single and 1:1 co-inoculation.	Lux^{-}; LMB542 Lux^{+}; LMB553	For each strain: nodule area based on ten largest nodules

Data files:

- **Pictures**

  (a) of whole or half-root systems with stained nodules if applicable. File names for whole roots: x_white_Imgxxxxx. The first number indicates plant number. File names for half-roots: x_x_L_white or x_x_R_white. The first number indicates plant number, the second number indicates pot number (two pots per plant). L and R indicate an alternative method of indicating the two pots per plant. All pictures are also available with a black background.

  (b) of whole root systems as photographed with NightOwl II camera (Berthold Technologies) at an exposure time of 120 seconds using inbuilt IndiGO software v2.0.5.0. The number indicates plant number.

- **Data in .csv format**

  (a) Nodule sizes.csv. Nodule sizes from all experiments

  (b) Nodule counts.csv Nodule numbers from all experiments

  (c) Acetylene reductions.csv Acetylene reduction data

Specification of data in each of the files:

Nodule sizes.csv

Nodule	Nodule number (1-30 as up to 30 nodules measured on a single plant (10 nodules x three strains).
Label	Label as given by ImageJ indicating from which picture file data comes
Area	Nodule area measured
Plant	Plant number
Experiment	Experiment number
Selection	Randomly selected or largest nodules

Root_Type	Single or split-root
Soil_N	Soil nitrogen addition
Treatment	Treatment applied
Focal_Strain	Focal strain (Rlv3841 for Rlv3841 *gusA*), LMB827, OPS0365, OPS0366, LMB542 or LMB553
Competing_Strain	Focal strain (Rlv3841 for Rlv3841 *gusA*), LMB827, OPS0365, OPS0366, LMB542 or LMB553
Focal_Strain_Name	Fix^+ for Rlv3841 *gusA* Fix^{Int} for LMB827 Fix^- for OPS0365 and OPS0366 Lux^- for LMB542 Lux^+ for LMB553
Competing_Strain_Name	Fix^+ for Rlv3841 *gusA* Fix^{Int} for LMB827 Fix^- for OPS0365 and OPS0366 Lux^- for LMB542 Lux^+ for LMB553
Nodule counts.csv	As for nodule sizes, in addition:
Nr_Nodules_Focal_Strain	Number of nodules of focal strain
Nr_Nodules_Competing_ Strain	Number of nodules of competing strain
Acetylene reductions.csv	As for nodule sizes, in addition:
Perc_Ethylene	Percentage ethylene in sample as measured by gas chromatography (Clarus 480®, Perkin-Elmer®).

Perc_Acetylene	Percentage acetylene in sample as measured by gas chromatography (Clarus 480®, Perkin-Elmer®).
Replicate	1-3 as measured three technical replicates per plant

3. R code used to analyse the data

Data analysis resource allocation among strains

Annet Westhoek

August 15, 2017

```
# last update August 15, 2017

## code to analyse data on resource allocation among strains

## Order of analyses:
# Size of strains in single inoculations
#         (combined from results Exp. 12, 13, 14, 16, 21)
# Resource allocation in dual inoculation (Exp. 13 and Exp. 14)
# Resource allocation in triple inoculation (Exp. 18)
# Acetylene reduction data (Exp. 21)
# Soil nitrate experiments (Exp. 12 and Exp. 16)
# Allocation based on carbon demands (Exp. 28)
# Results based on ten randomly selected nodules (Exp. 12, 13, 14, 16)

print(sessionInfo(), locale = FALSE)

## R version 3.3.2 (2016-10-31)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 15063)
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] nlme_3.1-128    dplyr_0.5.0     lattice_0.20-34 reshape_0.8.6
##  [5] plyr_1.8.4      boot_1.3-18     arm_1.9-3       lme4_1.1-12
##  [9] Matrix_1.2-7.1  MASS_7.3-45     ggplot2_2.2.1
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_0.12.9      knitr_1.15.1     magrittr_1.5     splines_3.3.2
##  [5] munsell_0.4.3    colorspace_1.3-2 R6_2.2.0         minqa_1.2.4
##  [9] stringr_1.1.0    tools_3.3.2      grid_3.3.2       gtable_0.2.0
## [13] DBI_0.5-1        coda_0.19-1      htmltools_0.3.5  abind_1.4-5
## [17] yaml_2.1.14      lazyeval_0.2.0   rprojroot_1.2    digest_0.6.12
## [21] assertthat_0.1   tibble_1.2       nloptr_1.0.4     evaluate_0.10
## [25] rmarkdown_1.3    stringi_1.1.2    scales_0.4.1     backports_1.0.5

# read in the results about nodule sizes and
# check data type (factor, numerical etc.)
sizes <- read.csv("nodule sizes.csv")
head(sizes)
str(sizes)
sizes$Plant <- as.factor(sizes$Plant)
str(sizes)

# read in the results about acetylene reductions
# and check data type
results_ARA <- read.csv("acetylene reductions.csv")
```

```
str(results_ARA)
results_ARA$Plant <- as.factor(results_ARA$Plant)

# read in the results about nodule numbers
counts <- read.csv("nodule counts.csv")

# Unless otherwise stated analyses based on ten largest nodules
# For results based on strains differing in how much nitrogen they provide:
fixation <- subset(sizes, (sizes$Focal_Strain_Name == "Fix-" |
                             sizes$Focal_Strain_Name == "Fix-Int" |
                             sizes$Focal_Strain_Name == "Fix+") &
                     sizes$Selection == "largest")

# For results based on strains differing in how much carbon they cost:
carbon <- subset(sizes, (sizes$Focal_Strain_Name == "Lux-" |
                           sizes$Focal_Strain_Name == "Lux+") &
                   sizes$Selection == "largest")

# calculate the average nodule size per plant and focal strain
means_fixation <- fixation %>%
  group_by(Label, Plant, Experiment, Selection, Root_Type, Soil_N, Treatment,
           Focal_Strain, Competing_Strain,
           Focal_Strain_Name, Competing_Strain_Name) %>%
  summarise(
    N_Area = sum(!is.na(Area)),
    Mean_Area = mean(Area, na.rm = TRUE),
    sd_Area = sd(Area, na.rm = TRUE),
    se_Area = sd_Area / sqrt(N_Area))  %>%
  as.data.frame()
head(means_fixation)
str(means_fixation)

# calculate the average nodule size per plant and focal strain
means_carbon <- carbon %>%
  group_by(Label, Plant, Experiment, Selection, Root_Type, Soil_N, Treatment,
           Focal_Strain, Competing_Strain,
           Focal_Strain_Name, Competing_Strain_Name) %>%
  summarise(
    N_Area = sum(!is.na(Area)),
    Mean_Area = mean(Area, na.rm = TRUE),
    sd_Area = sd(Area, na.rm = TRUE),
    se_Area = sd_Area / sqrt(N_Area))  %>%
  as.data.frame()
head(means_carbon)
str(means_carbon)

################################################################################
####    RESULTS ON STRAINS DIFFERING IN HOW MUCH NITROGEN THEY PROVIDE  ######
################################################################################

## Size of Fix+, Fix-Int and Fix- strain in single inoculations (zero soil N)
## (or in dual inoculation but competed against itself)
## in either single or split-root experiments

# select data based on single inoculations in zero soil N
single <- subset(fixation, fixation$Soil_N == 0 &
                   is.na(fixation$Competing_Strain) == TRUE)
# or 'co-inoculation' but competed against itself
dual_int <- fixation[as.character(fixation$Focal_Strain) ==
```

```
                      as.character(fixation$Competing_Strain),]
dual_int2 <- subset(dual_int, dual_int$Soil_N == 0)

# For split-root Experiment: measured 10 largest nodules on each half-root
# So to select ten largest nodules of whole root need to find the ten largest
from
# both half-roots for fair comparison to single / mixed inoculation on single
root

# split the dual inoculation results in single root and split-root
dual_single_root <- subset(dual_int2, dual_int2$Root_Type == "single")
dual_split_root <- subset(dual_int2, dual_int2$Root_Type == "split")

# For split roots: order dataframe per plant with largest nodules on top
dual_split_root2 <- arrange(dual_split_root, Plant, -Area)
# now select only the top 10 entries for each plant
dual_split_root3 <- ddply(dual_split_root2,.(Plant),function(x) head(x,10))

# combine all the data
combined <- rbind(single, dual_single_root, dual_split_root3)

# calculate the average nodule size per plant and focal strain
means <- combined %>%
  group_by(Label, Plant, Experiment, Selection, Root_Type, Soil_N, Treatment,
           Focal_Strain, Competing_Strain,
           Focal_Strain_Name, Competing_Strain_Name) %>%
  summarise(
    N_Area = sum(!is.na(Area)),
    Mean_Area = mean(Area, na.rm = TRUE),
    sd_Area = sd(Area, na.rm = TRUE),
    se_Area = sd_Area / sqrt(N_Area))  %>%
  as.data.frame()
head(means)
str(means)

current_data <- means

# plot the data: plot means per plant
p_single <- ggplot(data = current_data,
                   aes(x = Focal_Strain_Name, y = Mean_Area,
                       shape = Experiment)) +
  geom_jitter(size = 10, stroke = 4, width = 0.1) +
  scale_y_continuous(limits = c(0, 7)) +
  scale_shape_manual(values = c("Exp. 13" = 1, "Exp. 14" = 3, "Exp. 21" = 2,
                                "Exp. 12" = 4, "Exp. 16" = 5),
                     breaks = c("Exp. 13", "Exp. 14", "Exp. 21",
                                "Exp. 12", "Exp. 16")) +
  theme_bw(base_size = 38) +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        legend.key = element_blank(),
        legend.title = element_blank(),
        legend.position = "none",
        axis.text = element_text(size = 38),
        axis.line.x = element_line(size = 2, color = "black"),
        axis.line.y = element_line(size = 2, color = "black")) +
  xlab("Strain") +
  ylab(expression(paste("Nodule area (mm"^"2", ")")))
p_single
```

```r
# # Export graph
# tiff("Single inoculations.tiff",
#      width = 10, height = 10, units = 'in', res = 300, compression = 'none')
# p_single # Make plot
# dev.off()

# Analyse size of nodules as function of strain, with Experiment as random eff
ect
current_data$Focal_Strain_Name <- relevel(current_data$Focal_Strain_Name,
                                          ref = "Fix-Int")
me1 <- lme(Mean_Area ~ Focal_Strain_Name,
           random = ~ 1|Experiment, data = current_data)
anova(me1)
summary(me1)
intervals(me1)

# check model assumptions
plot(me1)
qqnorm(residuals(me1))
hist(residuals(me1), breaks = 10)

# save data on Fix+ strain for later analysis
Fix_Plus <- subset(current_data, current_data$Focal_Strain_Name == "Fix+")
# clear dataframe
rm(current_data)

################################################################################
################################################################################

## Analysis of size of intermediate-fixing strain depending on
# which other strains are present on the plant in single root
# experiments dual inoculations (Exp. 13 and Exp. 14)

# Select only results based on single root experiments
current_data <- subset(means_fixation,
                       means_fixation$Experiment == "Exp. 13" |
                         means_fixation$Experiment == "Exp. 14")

# For visualisation, also calculate the overall mean of
# intermediate-fixing nodules for each competing strain
mean_int <- subset(current_data,
                   means_fixation$Focal_Strain_Name == "Fix-Int")
mean_grand <- mean_int %>%
  group_by(Competing_Strain_Name) %>%
  summarise(Grand_Mean = mean(Mean_Area)) %>%
  as.data.frame()
mean_grand

# Make a plot with the size of intermediate-fixing nodules
# as function of competing strain
p_Fix_Int <- ggplot() +
  geom_jitter(data = subset(current_data,
                            current_data$Focal_Strain_Name == "Fix-Int"),
              aes(x = Competing_Strain_Name, y = Mean_Area,
                  shape = Experiment),
              size = 10, stroke = 4, width = 0.05) +
  geom_point(data = mean_grand,
             aes(x = Competing_Strain_Name, y = Grand_Mean),
```

```
            size = 10, stroke = 4, pch = 2, color = "#e66101") +
  scale_y_continuous(limits = c(0, 7)) +
  scale_shape_manual(values = c("Exp. 13" = 1, "Exp. 14" = 3),
                     breaks = c("Exp. 13", "Exp. 14")) +
  theme_bw(base_size = 38) +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        legend.key = element_blank(),
        legend.title = element_blank(),
        legend.position = "none",
        axis.text = element_text(size = 38),
        axis.line.x = element_line(size = 2, color = "black"),
        axis.line.y = element_line(size = 2, color = "black")) +
  xlab("Competing strain") +
  ylab(expression(paste("Nodule area (mm"^"2", ")")))
p_Fix_Int

# # Export graph
# tiff("Fix_Int size.tiff", width = 10, height = 10, units = 'in', res = 300,
# compression = 'lzw')
# p_Fix_Int # Make plot
# dev.off()

# For comparison, make a plot with the size of non-fixing and fixing nodules
# (when competed against intermediate-fixing strain).
p_comparison <- ggplot() +
  geom_jitter(data = subset(current_data,
                            current_data$Focal_Strain_Name != "Fix-Int" &
                  current_data$Competing_Strain_Name == "Fix-Int"),
              aes(x = Focal_Strain_Name, y = Mean_Area, shape = Experiment),
              size = 10, stroke = 4, width = 0.05) +
  geom_point(data = subset(mean_grand,
                           mean_grand$Competing_Strain_Name != "Fix-Int"),
             aes(x = Competing_Strain_Name, y = Grand_Mean),
             size = 10, stroke = 4, pch = 2, color = "#e66101") +
  scale_y_continuous(limits = c(0, 7)) +
  scale_shape_manual(values = c("Exp. 13" = 1, "Exp. 14" = 3),
                     breaks = c("Exp. 13", "Exp. 14")) +
  theme_bw(base_size = 38) +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        legend.key = element_blank(),
        legend.title = element_blank(),
        legend.position = "none",
        axis.text = element_text(size = 38),
        axis.line.x = element_line(size = 2, color = "black"),
        axis.line.y = element_line(size = 2, color = "black")) +
  xlab("Strain") +
  ylab(expression(paste("Nodule area (mm"^"2", ")")))
p_comparison

# # Export graph
# tiff("comparison.tiff", width = 10, height = 10, units = 'in', res = 300,
#      compression = 'lzw')
# p_comparison # Make plot
# dev.off()

## Analyze the nodule size as function of competing strain
# with Experiment as random effect
```

```
current_data$Competing_Strain_Name <- relevel(current_data$Competing_Strain_Na
me,
                                              ref = "Fix-Int")
me1 <- lme(Mean_Area ~ Competing_Strain_Name, random = ~ 1|Experiment,
           data = subset(current_data,
                         current_data$Focal_Strain_Name == "Fix-Int"))
anova(me1)
summary(me1)
intervals(me1)

# check model assumptions
plot(me1)
qqnorm(residuals(me1))
hist(residuals(me1), breaks = 10)

# Test whether intermediate-fixing nodules become
# as small as non-fixing nodules
# in the presence of the fixing strain

# select nodule area measurements for intermediate-fixing strain
# when competed against fixing strain
int_fix <- subset(current_data,
                  current_data$Focal_Strain_Name == "Fix-Int" &
                    current_data$Competing_Strain_Name == "Fix+")

# select nodule area measurements for non-fixing strain
# (competed against intermediate-fixing strain)
non_fix <- subset(current_data, current_data$Focal_Strain_Name == "Fix-")

# do t-test on the two groups. Measured on different plants so can't be paired
t.test(int_fix$Mean_Area,
       non_fix$Mean_Area,
       paired = FALSE,
       var.equal = TRUE)

# test whether difference in size between fixing
# and intermediate-fixing strain increases in co-inoculation as compared
# to single inoculation. If so, this unequivocally shows that
# resource allocation depends on competing strain
# (if fixing nodules also become smaller in co-inoculation,
# it could be argued that this is artefact of measuring 20 largest nodules
# rather than 10 largest nodules when strain is competed against itself)

# select data on nodule size intermediate-fixing strain when co-inoculated
# with itself or with fixing strain
Fix_Int <- subset(current_data,
                  current_data$Focal_Strain_Name == "Fix-Int" &
                    current_data$Competing_Strain_Name != "Fix-")
Fix_Plus_Dual <- subset(current_data,
                        current_data$Focal_Strain_Name == "Fix+")

# clear dataframe
rm(current_data)

# combine data on Fix-Int with data on Fix+
# (from single as saved from single inoculation data)
current_data <- rbind(Fix_Int, Fix_Plus, Fix_Plus_Dual)

# test whether difference in size depends on single or co-inoculation
```

```r
# first indicate whether results are from single or dual inoculation
current_data$Inoculation <-
  as.factor(ifelse(is.na(current_data$Competing_Strain_Name) == TRUE,
                                                  "single",
                    ifelse(as.character(current_data$Focal_Strain_Name) !=
                   as.character(current_data$Competing_Strain_Name),
                    "dual", "single")))
current_data$Inoculation <- relevel(current_data$Inoculation, ref = "single")

me1 <- lme(Mean_Area ~ Focal_Strain_Name + Focal_Strain_Name * Inoculation,
           random = ~ 1 | Experiment, data = current_data)
anova(me1)
summary(me1)
intervals(me1)

# check model assumptions
plot(me1)
qqnorm(residuals(me1))
hist(residuals(me1), breaks = 10)

# clear dataframe
rm(current_data)

# Check that Fix-Int and competing strain (Fix+ or Fix-)
# form +/- equal numbers of nodules
current_data <- subset(counts, (counts$Experiment == "Exp. 13" |
                         counts$Experiment == "Exp. 14") &
                         counts$Competing_Strain_Name != "Fix-Int")
current_data$Perc_Nodules_Focal_Strain <- (current_data$Nr_Nodules_Focal_Strai
n /
  (current_data$Nr_Nodules_Focal_Strain + current_data$Nr_Nodules_Competing_St
rain)) * 100

# Plot the percentage of nodules formed by Fix-Int strain as function of compe
ting strain
p_perc <- ggplot() +
  geom_jitter(data = current_data,
              aes(x = Competing_Strain_Name, y = Perc_Nodules_Focal_Strain,
                  shape = Experiment),
              size = 10, stroke = 4, width = 0.05) +
  scale_y_continuous(limits = c(0, 100)) +
  scale_shape_manual(values = c("Exp. 13" = 1, "Exp. 14" = 3),
                     breaks = c("Exp. 13", "Exp. 14")) +
  theme_bw(base_size = 38) +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        legend.key = element_blank(),
        legend.title = element_blank(),
        legend.position = "none",
        axis.text = element_text(size = 38),
        axis.line.x = element_line(size = 2, color = "black"),
        axis.line.y = element_line(size = 2, color = "black")) +
  xlab("Competing strain") +
  ylab(expression(paste("Percentage Fix-Int nodules")))
p_perc

# Analyze percentage of nodules formed by Fix-Int as function of competing str
ain
# Include experiment as random effect
```

```r
me1 <- lme(Perc_Nodules_Focal_Strain ~ Competing_Strain_Name,
           random = ~ 1|Plant, data = current_data)
anova(me1)
summary(me1)
intervals(me1)
# No indication that number of nodules formed depends on competing strain

# check model assumptions
plot(me1)
qqnorm(residuals(me1))
hist(residuals(me1), breaks = 10)

# Calculate mean +- se of percentage nodules
mean_perc <- mean(current_data$Perc_Nodules_Focal_Strain)
sd_perc <- mean(current_data$Perc_Nodules_Focal_Strain)
se_perc <- sd_perc / sqrt(length(current_data))

# clear dataframe
rm(current_data)

################################################################################
################################################################################

## Analysis of size of intermediate-fixing strain depending on which other
# strains are present on the plant in single root experiments with triple
# inoculation of fixing, non-fixing and intermediate-fixing strain (Exp. 18)

current_data <- subset(means_fixation, means_fixation$Experiment == "Exp. 18")

# plot the results
p_triple <- ggplot(data = current_data) +
  geom_jitter(aes(x = Focal_Strain_Name, y = Mean_Area),
              size = 10, stroke = 4, pch = 1, width = 0.05) +
  scale_y_continuous(limits = c(0, 7)) +
  theme_bw(base_size = 38) +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        legend.key = element_blank(),
        axis.text = element_text(size = 38),
        axis.line.x = element_line(size = 2, color = "black"),
        axis.line.y = element_line(size = 2, color = "black")) +
  xlab("Strain") +
  ylab(expression(paste("Nodule area (mm"^"2", ")")))
p_triple

# # Export graph
# tiff("three strains.tiff", width = 10, height = 10,
# units = 'in', res = 300, compression = 'lzw')
# p_triple # Make plot
# dev.off()

# Analyze nodule size as function of strain,
# in mixed-effects model to take Plant into
# account, as on each plant all of the three strains were measured.
me1 <- lme(Mean_Area ~ Focal_Strain_Name, random = ~ 1|Plant,
           data = current_data)
anova(me1)
summary(me1)
intervals(me1)
```

```r
# check model assumptions
plot(me1)
qqnorm(residuals(me1))
hist(residuals(me1), breaks = 10)

# clear dataframe
rm(current_data)
rm(combined)

################################################################################
#####                    ACETYLENE REDUCTION DATA                       #######
################################################################################

############################################################
# Notes on acetylene reduction calculations:
# STP = standard temperature and pressure:
# 273.15 K (0 degrees C, 32 degrees F) and
# an absolute pressure of exactly 100 kPa (1 bar).

# At STP, 1 mol of acetylene occupies 22.41 L and
# at 25 degrees Celsius it occupies 24.465 L
# (can verify this using P1V1 / T1 = P2V2 / T2).

# The area under a gas-chromatograph peak is proportional to
# the amount ( moles ) of compound eluted. Hence, the molar percentage
# composition of a mixture can be approximated by
# comparing relative peak areas.

# Initial concentration of acetylene is 2%
# (added 6.4 mL of 100% acetylene gas and
# total volume of bottle is 320 mL)
############################################################

# do the calculations to convert Perc_Ethylene and
# Perc_Acetylene data to how much acetylene has been reduced.
# First, calculate the initial amount (in umol) of acetylene added:
# 2% acetylene in 0.320 L bottle.
# At 25 degrees Celsius and normal pressure 1 mol of
# acetylene takes up 24.465 L. So:
results_ARA$uMol_Acetylene_Added <- (0.02 * 0.320 / 24.465) * 1000000
# Then, calculate the percentage of acetylene converted to ethylene.
# Perc_Ethylene and Perc_Acetylene give
# the percentage of ethylene and acetylene (in terms of moles)
# measured in sample,as derived from peak area by GC.
# In most cases they add up to 100%,
# in some cases there was a small percentage of something else.
results_ARA$Perc_Converted <- (results_ARA$Perc_Ethylene /
    (results_ARA$Perc_Ethylene + results_ARA$Perc_Acetylene)) * 100
# The amount (in umols) of ethylene produced (per jar per hour,
# as incubation time was one hour) is the percentage of acetylene converted
# multiplied by the initial amount of acetylene added.
results_ARA$uMol_Ethylene_Produced <- (results_ARA$Perc_Converted / 100) *
  results_ARA$uMol_Acetylene_Added
head(results_ARA)

# calculate the average over three (or in two cases two) technical replicates
# to get average value and standard error per jar.
means_ARA <- results_ARA %>%
```

```
  group_by(Plant, Experiment, Root_Type, Soil_N, Treatment,
           Focal_Strain, Competing_Strain,
           Focal_Strain_Name, Competing_Strain_Name,
           Nr_Nodules_Focal_Strain, Nr_Nodules_Competing_Strain) %>%
  summarise(
    N_ARA = sum(!is.na(uMol_Ethylene_Produced)),
    Mean_Ethylene_Produced = mean(uMol_Ethylene_Produced, na.rm = TRUE),
    sd_ARA = sd(uMol_Ethylene_Produced, na.rm = TRUE),
    se_ARA = sd_ARA / sqrt(N_ARA))  %>%
  as.data.frame()
head(means_ARA)
str(means_ARA)
table(means_ARA$Focal_Strain, means_ARA$Competing_Strain)

# Also get the data about nodule sizes
# Acetylene reductions were only done on the focal strain, while nodule size
# was measured on both half-roots (so two measurements per plant)
# Only the half-roots with "L" in the label have acetylene reduction data
# So also select only those pots
current_sizes_int <- subset(means_fixation,
                            means_fixation$Experiment == "Exp. 21")
current_sizes <- current_sizes_int[grep("L", current_sizes_int$Label), ]

# add the data on nodule size to the acetylene reduction dataset
combined <- merge(means_ARA, current_sizes[,c("Plant", "Mean_Area")])
head(combined)

# Calculate the percentage of nodules formed by the focal strain
combined$Perc_Focal_Strain <- (combined$Nr_Nodules_Focal /
    (combined$Nr_Nodules_Focal + combined$Nr_Nodules_Competing))*100

# In some cases, none or (almost) all nodules were formed by the focal strain
# Exclude these data from analysis, as it's not fair to
# say there was a competing strain in those cases
# Also exclude few cases where data on number of nodules not
# available for both half-roots
# NB this action removes Fix- vs water control treatments

########################################
# Further details on which plants removed from data:
# If Fix-Int is the focal strain:
# Plant 11 (Fix-Int against Fix+) has formed no Fix+ nodules.
# If Fix+ was the focal strain:
# All plants of Fix+ against Fix-Int have formed
# almost solely Fix+ nodules (>99%).
# It seems this cannot be explained by Fix-Int being
# a less effective strain than Fix+,
# as when Fix-Int or Fix+ were competed against OPS0365,
# approximately equal number of nodules were formed on both half-roots.
# Plant 29 has formed no nodules on the focal side.
########################################

current_data <- subset(combined, combined$Perc_Focal_Strain > 0 &
                            combined$Perc_Focal_Strain < 99)

# Make a plot with nodule size of Fix-Int
# as a function of competing strain (Fix-, Fix-Int or Fix+)
p_size <- ggplot() +
  geom_jitter(data = subset(current_data,
```

```
                            current_data$Focal_Strain_Name == "Fix-Int"),
             aes(x = Competing_Strain_Name, y = Mean_Area),
             pch = 1, size = 10, stroke = 4, width = 0.05) +
  scale_y_continuous(limits = c(0, 7)) +
  theme_bw(base_size = 38) +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        legend.key = element_blank(),
        legend.title = element_blank(),
        legend.position = "none",
        axis.text = element_text(size = 38),
        axis.line.x = element_line(size = 2, color = "black"),
        axis.line.y = element_line(size = 2, color = "black")) +
  xlab("Competing strain") +
  ylab(expression(paste("Nodule area (mm"^"2", ")")))
p_size

# # Export graph
# tiff("Exp. 21 size.tiff", width = 10, height = 10, units = 'in',
#      res = 300, compression = 'lzw')
# p_size # Make plot
# dev.off()

# Plot acetylene reduction rates as function of
# average nodule size (per half-root)
p_ARA_size <- ggplot(data = subset(current_data, current_data$N_ARA > 2 &
                           current_data$Focal_Strain_Name == "Fix-Int"),
                     aes(x = Mean_Area, y = Mean_Ethylene_Produced)) +
  geom_point(size = 10, stroke = 4, shape = 1) +
  geom_smooth(method = lm, colour = "black", fill = "black",
              size = 3, alpha = 0.6) +
  coord_cartesian(xlim = c(0, 4), ylim = c(0, 3)) +
  annotate('text', x = 1.8, y = 3,
    label = "italic(y) == -1.41 + 0.91~italic(x)~','~italic(R^2) == 0.42",
    size = 12, parse = TRUE) +
  theme_bw(base_size = 38) +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        legend.key = element_blank(),
        axis.line.x = element_line(size = 2, color = "black"),
        axis.line.y = element_line(size = 2, color = "black")) +
  xlab(expression(paste("Nodule area (mm"^"2", ")"))) +
  ylab(expression(atop(paste("Ethylene produced (",mu,"mol)"),
                       paste("half-root"^"-1", " hour"^"-1"))))
p_ARA_size

# # Export graph
# tiff("Exp. 21 size and ARA.tiff", width = 10, height = 10,
#      units = 'in', res = 300, compression = 'lzw')
# p_ARA_size # Make plot
# dev.off()

# Analyses
# Analyse the size of the intermediate-fixing strain nodules
# as function of competing strain
m1 <- lm(Mean_Area ~ Competing_Strain_Name,
         data = subset(current_data,
                       current_data$Focal_Strain_Name == "Fix-Int"))
anova(m1)
```

```r
display(m1)
summary(m1)
confint(m1)

# check model assumptions
par(mfrow = c(2,2))
plot(m1)
par(mfrow = c(1,1))

# Model ARA per half-root of intermediate-fixing strain
# as function of competing strain
m2 <- lm(Mean_Ethylene_Produced ~ Competing_Strain_Name,
         data = subset(current_data,
          current_data$Focal_Strain_Name == "Fix-Int" &
          current_data$Competing_Strain_Name != "Fix-Int"))
anova(m2)
display(m2)
summary(m2)
confint(m2)

# check model assumptions
par(mfrow = c(2,2))
plot(m2)
par(mfrow = c(1,1))
boxcox(m2)

# See whether acetylene reduction per nodule positively
# correlates with nodule size
m3 <- lm(Mean_Ethylene_Produced ~ Mean_Area,
         data = subset(current_data, current_data$N_ARA > 2 &
                         current_data$Focal_Strain_Name == "Fix-Int"))
anova(m3)
display(m3)
summary(m3)
confint(m3)

# check model assumptions
par(mfrow = c(2,2))
plot(m3)
par(mfrow = c(1,1))
boxcox(m3)

# clear dataframe
rm(current_data)

#############################################################################
#####               SOIL NITRATE EXPERIMENTS                        #######
#############################################################################

# Analysis on effect of soil N on nodule size
# select the results on nodule size for the experiments
# with ranging soil N (Exp. 12 and 16)
current_data <- subset(means_fixation,
                       means_fixation$Experiment == "Exp. 12" |
                         means_fixation$Experiment == "Exp. 16")

# calculate average value per Treatment and Experiment
# so that each point is average of 1-6 replicate plants
grand_means <- current_data %>%
```

```
  group_by(Root_Type, Soil_N, Treatment,
           Focal_Strain, Focal_Strain_Name) %>%
  summarise(
    N_grand = sum(!is.na(Mean_Area)),
    Grand_mean = mean(Mean_Area, na.rm = TRUE),
    sd_grand = sd(Mean_Area, na.rm = TRUE),
    se_grand = sd_grand / sqrt(N_grand)) %>%
  as.data.frame()
head(grand_means)
str(grand_means)

# plot the data: plot means per treatment with error bars.
# Only include data with at least three replicates
p_soilN <- ggplot(data = subset(grand_means,
                                grand_means$N_grand >= 3),
                  aes(x = Soil_N, y = Grand_mean,
       colour = Focal_Strain_Name, pch = Focal_Strain_Name)) +
  geom_point(size = 5, stroke = 4) +
  geom_errorbar(aes(ymin=Grand_mean-se_grand, ymax=Grand_mean+se_grand),
                size = 1.5, width=.01, show.legend = FALSE) +
  scale_y_continuous(limits = c(0, 4)) +
  scale_colour_manual(values = c("#5e3c99", "#e66101", "#000000")) +
  scale_shape_manual(values = c(5, 2, 1)) +
  theme_bw(base_size = 38) +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        legend.title = element_blank(),
        legend.text = element_text(),
        legend.justification=c(1.05,1.05), legend.position=c(1,1),
        legend.background = element_rect(color = "black",
                                         fill = "white", size = 1),
        legend.key.height = unit(2.5, "line"),
        axis.line = element_line(size = 2, color = "black")) +
  xlab(expression(paste("g ", KNO[3], " / plant"))) +
  ylab(expression(paste("Nodule area (mm"^"2", ")")))
p_soilN

# # Export graph
# tiff("SoilN_size.tiff", width = 10, height = 10, units = 'in',
# res = 300, compression = 'lzw')
# p_soilN # Make plot
# dev.off()

# Analyze data: fit model of nodule size as function of
# both nitrate concentration and strain,
# also include the interaction, and with Experiment as random effect
me1 <- lme(Mean_Area ~ Focal_Strain_Name + Soil_N +
             Focal_Strain_Name:Soil_N, random = ~ 1|Experiment,
           data = current_data)
anova(me1)
summary(me1)
intervals(me1)

# check model assumptions
plot(me1)
qqnorm(residuals(me1))
hist(residuals(me1), breaks = 10)

# clear dataframes
```

```r
rm(current_data)
rm(grand_means)

# Analysis on effect of soil N on nodule number
current_data <- subset(counts, counts$Experiment == "Exp. 12" |
                         counts$Experiment == "Exp. 16")

grand_means <- current_data %>%
  group_by(Root_Type, Soil_N, Treatment,
           Focal_Strain, Focal_Strain_Name) %>%
  summarise(
    N_grand = sum(!is.na(Nr_Nodules_Focal_Strain)),
    Grand_mean = mean(Nr_Nodules_Focal_Strain, na.rm = TRUE),
    sd_grand = sd(Nr_Nodules_Focal_Strain, na.rm = TRUE),
    se_grand = sd_grand / sqrt(N_grand)) %>%
  as.data.frame()
head(grand_means)
str(grand_means)

# plot the data: plot means per treatment with error bars
# Only include data with at least three replicates
p_numbers <- ggplot(data = subset(grand_means, grand_means$N_grand >= 3),
                    aes(x = Soil_N, y = Grand_mean,
          colour = Focal_Strain_Name, pch = Focal_Strain_Name)) +
  geom_point(size = 5, stroke = 4) +
  geom_errorbar(aes(ymin=Grand_mean-se_grand, ymax=Grand_mean+se_grand),
                size = 1.5, width=.01, show.legend = FALSE) +
  scale_y_continuous(limits = c(0, 320)) +
  scale_colour_manual(values = c("#5e3c99", "#e66101", "#000000")) +
  scale_shape_manual(values = c(5, 2, 1)) +
  theme_bw(base_size = 38) +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        legend.title = element_blank(),
        legend.text = element_text(),
        legend.justification=c(1.05,1.05), legend.position=c(1,1),
        legend.background = element_rect(color = "black",
                                         fill = "white", size = 1),
        legend.key.height = unit(2.5, "line"),
        axis.line = element_line(size = 2, color = "black")) +
  xlab(expression(paste("g ", KNO[3], " / plant"))) +
  ylab(expression(paste("Number of nodules")))
p_numbers

# # Export graph
# tiff("SoilN_number.tiff", width = 10, height = 10, units = 'in',
# res = 300, compression = 'lzw')
# p_numbers # Make plot
# dev.off()

# fit model of nodule number as function of both nitrate concentration and str
ain,
# also include the interaction, and with Experiment as random effect
me2 <- lme(Nr_Nodules_Focal_Strain ~ Focal_Strain_Name + Soil_N +
             Focal_Strain_Name:Soil_N, random = ~ 1|Experiment,
           data = current_data)
anova(me2)
summary(me2)
intervals(me2)
```

```r
# check model assumptions
plot(me2)
qqnorm(residuals(me2))
hist(residuals(me2), breaks = 10)

# clear dataframes
rm(current_data)
rm(grand_means)

################################################################################
#####       RESULTS ON STRAINS DIFFERING IN HOW MUCH CARBON THEY COST    #####
################################################################################

# plot the results
# (annotate with A for single inoculation and B for co-inoculation)
means_carbon$Scenario <- ifelse(means_carbon$Treatment == "Lux- vs Lux+", "B",
"A")
table(means_carbon$Scenario, means_carbon$Treatment)
# To annotate panels with A, B
ann_text <- data.frame(x = c(0),y = c(6.5),lab=c("A","B"),
                       Scenario = factor(c("A","B"),levels = c("A","B")))

# To provide different title for each graph
scenario_names <- list(
  'A'="      Single inoculation",
  'B'="      Co-inoculation"
)
scenario_labeller <- function(variable,value){
  return(scenario_names[value])
}

p_carbon <- ggplot(data = means_carbon) +
  facet_wrap( ~ Scenario, ncol = 2, scales = "free",
              labeller = scenario_labeller) +
  geom_jitter(aes(x = Focal_Strain_Name, y = Mean_Area),
              size = 10, stroke = 4, pch = 1, width = 0.05) +
  geom_text(aes(x, y, label=lab, group=NULL),data=ann_text,
            size = 25, hjust = -0.3) +
  scale_y_continuous(limits = c(0, 7)) +
  theme_bw(base_size = 38) +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        legend.key = element_blank(),
        axis.text = element_text(size = 38),
        axis.line.x = element_line(size = 2, color = "black"),
        axis.line.y = element_line(size = 2, color = "black"),
        strip.background = element_rect(colour = NA, fill = NA),
        strip.text.x = element_text(size = 38),
        panel.spacing.x=unit(3, "lines")) +
  xlab("Strain") +
  ylab(expression(paste("Nodule area (mm"^"2", ")")))
p_carbon

# # Export graph
# tiff("allocation carbon.tiff", width = 20, height = 10,
#      units = 'in', res = 300, compression = 'lzw')
# p_carbon # Make plot
# dev.off()
```

```
## Analyze the nodule size as function of focal strain, including
## the interaction with treatment (single or dual inoculation).
## Only if difference in size is larger in dual than in single inoculation
## is that evidence for sanctioning based on carbon costs.
m1 <- lm(Mean_Area ~ Focal_Strain_Name + Focal_Strain_Name * Scenario,
         data = means_carbon)
anova(m1)
summary(m1)
confint(m1)

# check model assumptions
par(mfrow = c(2,2))
plot(m1)
par(mfrow = c(1,1))

# do t-test to obtain 95% CI of difference in nodule size
# in single inoculation
# (unpaired because data from different plants)
t.test(means_carbon$Mean_Area[means_carbon$Treatment == "Lux-"],
       means_carbon$Mean_Area[means_carbon$Treatment == "Lux+"], var.equal = T
RUE)

# do paired t-test to obtain 95% CI of difference in nodule size in co-inocula
tion
# (paired t-test to pair Lux- and Lux+ of each plant)

# first get data in correct format
# reshape the data to wide format
means_carbon$Name <- ifelse(means_carbon$Focal_Strain_Name == "Lux-",
                            "Lux_Min", "Lux_Plus")
table(means_carbon$Focal_Strain_Name, means_carbon$Name)
m_means <- melt(means_carbon,
                id.vars = c("Plant", "Name", "Treatment"),
                measure.vars = "Mean_Area")
wide_means <- cast(m_means, Plant ~ variable + Name)

# do paired t-test (only on co-inoculated plants)
test <- subset(wide_means, as.numeric(wide_means$Plant) > 14)
t.test(test$Mean_Area_Lux_Min,
       test$Mean_Area_Lux_Plus,
       paired = TRUE)

##############################################################################
######            RESULTS BASED ON RANDOMLY SELECTED NODULES          #######
##############################################################################

# clear dataframes
rm(list = ls())

# read in the results about nodule sizes and check data type
# (factor, numerical etc.)
sizes <- read.csv("nodule sizes.csv")
head(sizes)
str(sizes)
sizes$Plant <- as.factor(sizes$Plant)
str(sizes)

# For results based on strains differing in how much nitrogen they provide:
```

```
fixation <- subset(sizes, (sizes$Focal_Strain_Name == "Fix-" |
                             sizes$Focal_Strain_Name == "Fix-Int" |
                             sizes$Focal_Strain_Name == "Fix+") &
                     sizes$Selection == "random")

# calculate the average nodule size per plant and focal strain
means_fixation <- fixation %>%
  group_by(Label, Plant, Experiment, Selection, Root_Type, Soil_N, Treatment,
           Focal_Strain, Competing_Strain,
           Focal_Strain_Name, Competing_Strain_Name) %>%
  summarise(
    N_Area = sum(!is.na(Area)),
    Mean_Area = mean(Area, na.rm = TRUE),
    sd_Area = sd(Area, na.rm = TRUE),
    se_Area = sd_Area / sqrt(N_Area))  %>%
  as.data.frame()
head(means_fixation)
str(means_fixation)

## Size of Fix+, Fix-Int and Fix- strain in single inoculations (zero soil N)
## (or in dual inoculation but competed against itself)

# select data based on single inoculations in zero soil N
single <- subset(fixation, fixation$Soil_N == 0 &
                   is.na(fixation$Competing_Strain) == TRUE)
# or 'co-inoculation' but competed against itself
dual_int <- fixation[as.character(fixation$Focal_Strain) ==
                       as.character(fixation$Competing_Strain),]
dual_int2 <- subset(dual_int, dual_int$Soil_N == 0)

# combine the data
combined <- rbind(single, dual_int2)

# calculate the average nodule size per plant and focal strain
means <- combined %>%
  group_by(Label, Plant, Experiment, Selection, Root_Type, Soil_N, Treatment,
           Focal_Strain, Competing_Strain,
           Focal_Strain_Name, Competing_Strain_Name) %>%
  summarise(
    N_Area = sum(!is.na(Area)),
    Mean_Area = mean(Area, na.rm = TRUE),
    sd_Area = sd(Area, na.rm = TRUE),
    se_Area = sd_Area / sqrt(N_Area))  %>%
  as.data.frame()
head(means)
str(means)

current_data <- means

# plot the data: plot means per plant
p_single <- ggplot(data = current_data,
                   aes(x = Focal_Strain_Name,
                       y = Mean_Area, shape = Experiment)) +
  geom_jitter(size = 10, stroke = 4, width = 0.1) +
  scale_y_continuous(limits = c(0, 4)) +
  scale_shape_manual(values = c("Exp. 13" = 1, "Exp. 14" = 3, "Exp. 21" = 2,
                                "Exp. 12" = 4, "Exp. 16" = 5),
                     breaks = c("Exp. 13", "Exp. 14",
                                "Exp. 21", "Exp. 12", "Exp. 16")) +
```

```
  theme_bw(base_size = 38) +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        legend.key = element_blank(),
        legend.title = element_blank(),
        legend.position = "none",
        axis.text = element_text(size = 38),
        axis.line.x = element_line(size = 2, color = "black"),
        axis.line.y = element_line(size = 2, color = "black")) +
  xlab("Strain") +
  ylab(expression(paste("Nodule area (mm"^"2", ")")))
p_single

# # Export graph
# tiff("Single inoculations_random.tiff",
#      width = 10, height = 10, units = 'in', res = 300, compression = 'none')
# p_single # Make plot
# dev.off()

# Analyse size of nodules as function of strain,
# with Experiment as random effect
current_data$Focal_Strain_Name <- relevel(current_data$Focal_Strain_Name,
                                          ref = "Fix-Int")
me1 <- lme(Mean_Area ~ Focal_Strain_Name, random = ~ 1|Experiment,
           data = current_data)
anova(me1)
summary(me1)
intervals(me1)

# check model assumptions
plot(me1)
qqnorm(residuals(me1))
hist(residuals(me1), breaks = 10)

# save data on Fix+ strain for later analysis
Fix_Plus <- subset(current_data, current_data$Focal_Strain_Name == "Fix+")
# clear dataframe
rm(current_data)

##############################################################################
##############################################################################
## Analysis of size of intermediate-fixing strain depending on which other str
ains
## are present on the plant in single root experiments (Exp. 13 and Exp. 14)

# Select only results based on single root experiments
current_data <- subset(means_fixation, means_fixation$Experiment == "Exp. 13"
|
                       means_fixation$Experiment == "Exp. 14")
current_data$Competing_Strain_Name <- factor(current_data$Competing_Strain_Nam
e,
                                levels = c("Fix-", "Fix-Int", "Fix+"))

# For visualisation, also calculate the overall mean of
# intermediate-fixing nodules for each competing strain
mean_int <- subset(current_data,
                   means_fixation$Focal_Strain_Name == "Fix-Int")
mean_grand <- mean_int %>%
  group_by(Competing_Strain_Name) %>%
```

```
  summarise(Grand_Mean = mean(Mean_Area)) %>%
  as.data.frame()
mean_grand

# Make a plot with the size of intermediate-fixing nodules
# as function of competing strain
p_Fix_Int <- ggplot() +
  geom_jitter(data = subset(current_data,
                            current_data$Focal_Strain_Name == "Fix-Int"),
              aes(x = Competing_Strain_Name, y = Mean_Area,
                  shape = Experiment),
              size = 10, stroke = 4, width = 0.05) +
  geom_point(data = mean_grand,
             aes(x = Competing_Strain_Name, y = Grand_Mean),
             size = 10, stroke = 4, pch = 2, color = "#e66101") +
  scale_y_continuous(limits = c(0, 4)) +
  scale_shape_manual(values = c("Exp. 13" = 1, "Exp. 14" = 3),
                     breaks = c("Exp. 13", "Exp. 14")) +
  theme_bw(base_size = 38) +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        legend.key = element_blank(),
        legend.title = element_blank(),
        legend.position = "none",
        axis.text = element_text(size = 38),
        axis.line.x = element_line(size = 2, color = "black"),
        axis.line.y = element_line(size = 2, color = "black")) +
  xlab("Competing strain") +
  ylab(expression(paste("Nodule area (mm"^"2", ")")))
p_Fix_Int

# # Export graph
# tiff("Fix_Int size_random.tiff", width = 10, height = 10,
# units = 'in', res = 300, compression = 'lzw')
# p_Fix_Int # Make plot
# dev.off()

# For comparison, make a plot with the size of non-fixing and fixing nodules
# (when competed against intermediate-fixing strain).
p_comparison <- ggplot() +
  geom_jitter(data = subset(current_data,
                            current_data$Focal_Strain_Name != "Fix-Int" &
                              current_data$Competing_Strain_Name == "Fix-Int")
,
              aes(x = Focal_Strain_Name, y = Mean_Area, shape = Experiment),
              size = 10, stroke = 4, width = 0.05) +
  geom_point(data = subset(mean_grand,
                           mean_grand$Competing_Strain_Name != "Fix-Int"),
             aes(x = Competing_Strain_Name, y = Grand_Mean),
             size = 10, stroke = 4, pch = 2, color = "#e66101") +
  scale_y_continuous(limits = c(0, 4)) +
  scale_shape_manual(values = c("Exp. 13" = 1, "Exp. 14" = 3),
                     breaks = c("Exp. 13", "Exp. 14")) +
  theme_bw(base_size = 38) +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        legend.key = element_blank(),
        legend.title = element_blank(),
        legend.position = "none",
```

```
        axis.text = element_text(size = 38),
        axis.line.x = element_line(size = 2, color = "black"),
        axis.line.y = element_line(size = 2, color = "black")) +
  xlab("Strain") +
  ylab(expression(paste("Nodule area (mm"^"2", ")")))
p_comparison

# # Export graph
# tiff("comparison_random.tiff", width = 10, height = 10, units = 'in',
# res = 300, compression = 'lzw')
# p_comparison # Make plot
# dev.off()

## Analyze the nodule size as function of competing strain
# with Experiment as random effect
current_data$Competing_Strain_Name <- relevel(current_data$Competing_Strain_Na
me,
                                              ref = "Fix-Int")
m1 <- lm(Mean_Area ~ Competing_Strain_Name,
          data = subset(current_data,
                        current_data$Focal_Strain_Name == "Fix-Int"))
anova(m1)
summary(m1)
confint(m1)

# check model assumptions
par(mfrow = c(2,2))
plot(m1)
par(mfrow = c(1,1))


################################################################################
#####                    SOIL NITRATE EXPERIMENTS                       #######
################################################################################


# Analysis on effect of soil N on nodule size
# select the results on nodule size for
# the experiments with ranging soil N (Exp. 12 and 16)
current_data <- subset(means_fixation,
                       means_fixation$Experiment == "Exp. 12" |
                         means_fixation$Experiment == "Exp. 16")

# calculate average value per Treatment and Experiment
# so that each point is average of 1-6 replicate plants
grand_means <- current_data %>%
  group_by(Root_Type, Soil_N, Treatment,
           Focal_Strain, Focal_Strain_Name) %>%
  summarise(
    N_grand = sum(!is.na(Mean_Area)),
    Grand_mean = mean(Mean_Area, na.rm = TRUE),
    sd_grand = sd(Mean_Area, na.rm = TRUE),
    se_grand = sd_grand / sqrt(N_grand)) %>%
  as.data.frame()
head(grand_means)
str(grand_means)

# plot the data: plot means per treatment with error bars.
# Only include data with at least three replicates
```

```r
p_soilN <- ggplot(data = subset(grand_means,
                                grand_means$N_grand >= 3),
                  aes(x = Soil_N, y = Grand_mean,
        colour = Focal_Strain_Name, pch = Focal_Strain_Name)) +
  geom_point(size = 5, stroke = 4) +
  geom_errorbar(aes(ymin=Grand_mean-se_grand, ymax=Grand_mean+se_grand),
                size = 1.5, width=.01, show.legend = FALSE) +
  scale_y_continuous(limits = c(0, 2)) +
  scale_colour_manual(values = c("#5e3c99", "#e66101", "#000000")) +
  scale_shape_manual(values = c(5, 2, 1)) +
  theme_bw(base_size = 38) +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        legend.title = element_blank(),
        legend.text = element_text(),
        legend.justification=c(1.05,1.05), legend.position=c(1,1),
        legend.background = element_rect(color = "black",
                                         fill = "white", size = 1),
        legend.key.height = unit(2.5, "line"),
        axis.line = element_line(size = 2, color = "black")) +
  xlab(expression(paste("g ", KNO[3], " / plant"))) +
  ylab(expression(paste("Nodule area (mm"^"2", ")")))
p_soilN

# # Export graph
# tiff("SoilN_size_random.tiff", width = 10, height = 10, units = 'in',
# res = 300, compression = 'lzw')
# p_soilN # Make plot
# dev.off()

# Analyze data: fit model of nodule size as function of
# both nitrate concentration and strain,
# also include the interaction, and with Experiment as random effect
me1 <- lme(Mean_Area ~ Focal_Strain_Name + Soil_N +
             Focal_Strain_Name:Soil_N, random = ~ 1|Experiment,
           data = current_data)
anova(me1)
summary(me1)
intervals(me1)

# check model assumptions
plot(me1)
qqnorm(residuals(me1))
hist(residuals(me1), breaks = 10)

# clear dataframes
rm(current_data)
rm(grand_means)

##############################################################################
##############################################################################
##############################################################################
```