A molecular toolkit for high-throughput identification of competitive and effective rhizobial strains

Marcela Alejandra Mendoza Suárez

Balliol College, University of Oxford



A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2018

DECLARATION

I declare that this thesis was composed by myself and that the work contained herein is my own except where explicitly stated in the text. The work has not been submitted for any degree or professional qualification except as specified.

Marcela A. Mendoza-Suárez

ACKNOWLEDGEMENTS

I would like to thank my supervisor Professor Philip Poole, not only for all the support, guidance and teaching, but also for the freedom and trust so I could conceive and learn on my own. Phil, thanks for encouraging me, inspiring me, and for keeping me on the right path when I needed it. None of this would have been possible without you.

I would like to thank Dr Alison East for all her support and advice. Alison, you always helped me to explain my ideas in a simple, eloquent and elegant style. You gave me the courage to say out loud my ideas in this foreign language and if you hadn't supported me since the first day that we worked together at JIC, I would never have gotten to this place today.

I would like to thank Dr Carmen Sánchez-Cañizares and Dr Barney Geddes for training me in the lab and for your true friendship. Thanks for all the explanations, discussions and for sharing with me all your knowledge without expecting anything in return. When I met you, I knew immediately want kind of researcher I wanted to be, a combination of you both! Dr Beatriz Jorrín, you arrived to the lab in the precise moment I needed you the most. Thanks to your tips and tricks I was able to work with all 96-well plates without dying in the attempt. Thanks for teaching me the baby steps of bioinformatics that helped prevent me from going crazy, but most especially, thanks for your friendship. Dr Charlotte Kirchhelle, thanks for your contribution with confocal microscopy and image analysis. My most sincere thanks to Dr Ricardo Ramírez-González from JIC, for your friendship and for sharing with me all your knowledge about Biopython.

Thanks to all the current and former members of the Poole lab for your solidarity, friendship and advice, I know I'm not an easy person, so I really appreciate that you tolerated me these four years. I'm also grateful to Liam's group, and Helen, Lida and Maria for being such good colleagues. I would also like to thank all the admin staff at Plant Sciences: Gem, Roni, Rob, John and others that always helped me in the friendliest way.

I would like to thank the Mexican Government for funding this project through my CONACYT-I2T2 scholarship and to Balliol College for granted me the Foley-Béjar scholarship and additional funding throughout my DPhil. I would also like to thank the University of Oxford for providing me with the disability support needed for this project.

I want to thank all my friends who made my time in the UK so wonderful: Ivo, Heli, Bernardo, Rita María, Zuleyka, Barbara, Pri, Anna, Adrian, Jimena and to my Norwich family: Oscar, Ricardo, Rocio and Eleni.

Finally, I would like to thank my best friend, my confidant and my love, Leila Walker, you were the fundamental piece in this journey. Even in the most difficult moments, you kept me sane and happy, you made me laugh and dream for more…I didn't know why I ended up on this side of the world until the day that I met you…I thought I came to Oxford to get a degree, but what I got was my soul mate.

I dedicate this thesis to all my family, who have supported me in every step of my life, to my grandmother Alicia, for your unconditional love and prays, but especially to my best role model, the person that I admire and I love the most, my mother…paso a pasito...

Abstract

Legumes tend to become nodulated by highly competitive but not necessarily effective rhizobial strains, resulting in less optimal plant growth. Research focused on the identification of both competitive and highly effective nitrogen-fixing strains has been limited because, so far, the only way to test competitive rhizobial strains has been to isolate individual strains and compare them one at a time in large-scale plant growth assays.

We have developed a library of novel reporter plasmids (Plasmid IDs) compatible with Golden Gate cloning which consist of i) a promoter-probe vector (pOGG026) suitable for studies where antibiotic selection is not possible, ii) the maker gene sfGFP under the control of a synthetic consensus promoter (psNifH) to drive nodule-specific expression in biovars and strains of *Rhizobium leguminosarum* and which can be used as a proxy for nitrogenase activity assays, and iii) a 12-nucleotide specific sequence to mark *R. leguminosarum* strains.

Additionally, we developed reporter plasmids that express biomarkers, such as *gusA*, *celB* and mCherry suitable for rhizobial screening in the rhizosphere. The plasmid pOGG026 allowed us to develop a new plasmid-based system to mark *Rhizobium* strains with GusA/CelB and evaluate their competitiveness.

Results demonstrated the proof of concept of Plasmid IDs as a tool for high-throughput identification of competitive and effective rhizobial strains.

The application of this tool in a competition assay among eighty-four Tagged-strains, in non-sterilised agricultural soil containing indigenous rhizobia, allowed us to identify an effective and super competitive strain, in addition to mixed nodules occupied by up to four Tagged-strains. This tool has the potential to revolutionise the search for elite indigenous rhizobia, by reducing the costs involved in identifying competitive and effective rhizobial strains.

ACKNOWLEDGEMENTS II

ABSTRACT IV

LIST OF ABBREVIATIONS IX

CHAPTER 1: GENERAL INTRODUCTION - 1 -

NITROGEN CYCLE - 1 -
RHIZOSPHERE - 3 -
BIOLOGICAL NITROGEN FIXATION (BNF) AND IT IS IMPORTANCE IN AGRICULTURE - 4 -
RHIZOBIA - 5 -
LEGUMES - 7 -
THE RHIZOBIA-LEGUME SYMBIOSIS - 8 -
ESTABLISHMENT - 8 -
INDETERMINATE NODULES - 10 -
DETERMINATE NODULES - 11 -
MIXED NODULES - 12 -
HOST SPECIFICITY - 13 -
RHIZOBIAL ADAPTATION - 14 -
RHIZOBIAL COMPETITIVENESS - 14 -
N_2-FIXING EFFECTIVENESS AND INEFFECTIVENESS - 15 -
RHIZOBIAL IDENTIFICATION - 15 -
METHODS USED TO STUDY RHIZOBIAL COMPETITION - 17 -
COMMERCIAL INOCULANTS - 18 -
AIM OF THE PROJECT - 19 -

CHAPTER 2: GENERAL MATERIALS AND METHODS - 21 -

MEDIA, ANTIBIOTICS AND OTHER CHEMICALS - 21 -
BACTERIAL STRAINS, PLASMIDS AND PRIMERS - 23 -
MOLECULAR TECHNIQUES - 47 -
PLANT EXPERIMENTS - 54 -
IMAGE ACQUISITION - 61 -
COMPUTATIONAL METHODS - 63 -

CHAPTER 3: DEVELOPMENT AND VALIDATION OF A VECTOR FOR HIGH-THROUGHPUT CLONING IN RHIZOBIA - 64 -

INTRODUCTION - 64 -
RESULTS AND DISCUSSION - 65 -
DESIGN AND CONSTRUCTION OF POGG026 - 65 -

HIGH-THROUGHPUT CLONING IN POGG026 USING GOLDEN GATE ASSEMBLY - 70 -
ASSESSING MARKER GENE DETECTION IN SYMBIOSIS USING VECTOR POGG026 - 75 -
USING THE PLASMID-BASED MARKER SYSTEM TO ASSESS RHIZOBIAL COMPETITIVENESS - 78 -
CONCLUSIONS **- 82 -**

CHAPTER 4: DESIGN, VALIDATION AND APPLICATION OF A SYNTHETIC CONSENSUS PROMOTER - 84 -

INTRODUCTION **- 84 -**
RESULTS AND DISCUSSION **- 85 -**
DESIGN OF A SYNTHETIC CONSENSUS PROMOTER TO DRIVE NODULE-SPECIFIC EXPRESSION ACROSS A VARIETY OF RHIZOBIUM *LEGUMINOSARUM* BIOVARS AND STRAINS - 85 -
ASSESSMENT OF THE PERFORMANCE OF PSNIFH TO DRIVE STRONG, NODULE-SPECIFIC EXPRESSION ACROSS RHIZOBIAL STRAINS - 90 -
ASSESSMENT OF THE CONSENSUS SYNTHETIC PROMOTER PSNIFH IN COMPETITION FOR NODULE FORMATION ASSAYS USING ISOGENIC MARKED AND UNMARKED RHIZOBIAL STRAINS - 107 -
DETECTION OF SFGFP UNDER CONTROL OF SYNTHETIC CONSENSUS PROMOTER PSNIFH AS A PROXY FOR NITROGENASE ACTIVITY ASSAYS - 110 -
CONCLUSION **- 119 -**

CHAPTER 5: DESIGN AND TESTING OF TAGGING BACTERIAL STRAINS USING PLASMID IDS - 120 -

INTRODUCTION **- 120 -**
RESULTS AND DISCUSSION **- 122 -**
DESIGN OF A PLASMID ID FOR BACTERIAL IDENTIFICATION AT STRAIN LEVEL - 122 -
USE OF A PLASMID ID FOR HIGH-THROUGHPUT IDENTIFICATION OF RHIZOBIAL ISOLATES FOR NODULE OCCUPANCY - 128 -
PRIMER DESIGN FOR MULTIPLEX ION TORRENT SEQUENCING STRATEGY - 130 -
PLASMID ID DATA ANALYSIS BY MISEQ SEQUENCING - 137 -
CONCLUSIONS **- 142 -**

CHAPTER 6: APPLICATION OF A MOLECULAR TOOL FOR A HIGH-THROUGHPUT IDENTIFICATION OF COMPETITIVE AND EFFECTIVE RHIZOBIAL STRAINS - 144 -

INTRODUCTION **- 144 -**
RESULTS AND DISCUSSION **- 146 -**
ANALYSIS OF SELECTED SOIL - 146 -
STRAINS WELL-STUDIED IN OUR LAB IN COMPETITION ASSAYS - 147 -
TAGGING RLV-LIBRARY WITH PLASMID IDS - 150 -
COMPETITION ASSAYS AMONG TAGGED-STRAINS IN PRESENCE OF INDIGENOUS RHIZOBIA - 154 -
RESULTS FROM A SINGLE CLONE INFECTION - 157 -

NODULES CONTAINING MORE THAN ONE STRAIN (MIXED INFECTION) - 160 -
CONCLUSIONS **- 165 -**

CHAPTER 7: GENERAL DISCUSSION AND FUTURE PERSPECTIVES - 167 -

REFERENCES - 177 -

SUPPLEMENTAL MATERIAL - 201 -

List of abbreviations

(Excluding chemical symbols and International Units)

Amp	Ampicillin
Amp^r	Ampicillin resistant
ALA	5-aminolevulinic acid
ARA	Acetylene reduction assay
BNF	Biological nitrogen fixation
BLAST	Basic local alignment search tool
bp	base pairs
BSA	Bovine serum albumin
CDS	Coding sequence
Cfu	Colony-forming units
Chl	Chloramphenicol
Chl^r	Chloramphenicol resistant
CPS	Counts per second
cv	Cultivar
dpi	days post inoculation
dNTPs	Deoxyribonuclotide triphosphates
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DWA	Distilled water agar
EDTA	Ethylenediaminetetraacetic acid
Fix^+	Nitrogen fixing phenotype
Fix^-	Non-nitrogen fixing phenotype

GC	Gas chromatography
GFP	Green fluorescent protein
Gm	Gentamycin
Gm^r	Gentamycin resistant
gus	β- glucuronidase
IGR	Intergenic region
INSeq	Insertion sequencing
IT	Infection thread
LB	Luria Bertani broth
kb	Kilobase
Kan	Kanamycin
Kan^r	Kanamycin resistant
Log	Logarithm
Magenta-glcA	5-bromo-6-chloro-3-indolyl-β-D-glucuronide acid
MPN	Most Probable Number
NCR	Nodule-specific cysteine rich peptides
NF	Nod Factor
Nm	Neomycin
Nm^r	Neomycin resistant
Nit	Nitrofurantoin
Nit^r	Nitrofurantoin resistance
nt	Nucleotides
^oC	Degrees (Celsius)
OD	Optical density
OD_{600}	Optical density at 600nm

ONPG	Ortho-Nitrophenyl-β-galactoside
ORF	Open reading frame
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEG	Poly (ethylene glycol)
PEP	Phosphoenolpyruvate
^{r}	Resistant
Rf	Rifampicin
Rf^{r}	Rifampicin resistant
Rl	*Rhizobium leguminosarum*
Rlv	*Rhizobium leguminosarum* bv. *viciae*
RNA	Ribonucleic acid
rpm	Revolutions per minute
RT	Room temperature
SDS	Sodium dodecyl sulphate
SEM	Standard error of the mean
sfGFP	Superfolder green fluorescent protein
sp.	*specie*
Sp	Spectinomycin
Sp^{r}	Spectinomycin resistant
Str	Streptomycin
Str^{r}	Streptomycin resistant
Tet	Tetracycline
Tet^{r}	Tetracycline resistant
Tm	Melting temperature

TY	Tryptone yeast
UAS	Upstream activator sequence
UMS	Universal Minimal Salts
UV	Ultraviolet
v/v	Volume of solute/volume of solution
WT	Wild-type
w/v	Mass of solute/volume of solution
X-glcA	5-bromo-4-chloro-3-indolyl-β-D-glucuronide
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside

Chapter 1: General introduction

Nitrogen cycle

Nitrogen is essential for all organisms on earth, given its role in the synthesis of proteins and nucleic acids. After water (H_2O) and carbon dioxide (CO_2), nitrogen (N), phosphorus (P), and potassium (K) are considered the most important nutrients but N is the most limiting for plant growth (Lupwayi *et al.*, 2006). The atmosphere contains approximately 80% nitrogen gas (N_2), but unfortunately it is unavailable for use in this form by many organisms because of the strength of the triple bond that holds together the N atoms. A limited number of species of Bacteria and Archaea have evolved the ability to fix N_2 to forms such as ammonia (NH_3) or ammonium (NH_4^+) which are an important source of nitrogen for many plant species (Galloway *et al.*, 2004). The limited soil nitrogen supply and the dependence of plants on this element have created a vast industry dedicated to producing synthetic nitrogen-based fertilisers (N-fertilisers) in order to obtain high yields of crops (Butterbach-Bahl *et al.*, 2011)

We cannot discuss yields of crops without considering the Haber-Bosch process, which mimics biological fixation by taking N from the air and making the chemical reaction with hydrogen in the presence of iron at high pressures and temperatures. The discovery of this process in the early 1900s[1] has profoundly increased the quantity of nitrogen that is transformed from atmospheric N_2 to NH_3 and triggered what we now know as the *green revolution*, which allowed rapid agricultural development and consequently an expansion in the global population (Galloway *et al.*, 2004).

[1] Fritz Haber and Carl Bosch were awarded the Nobel Prize in Chemistry in 1919 for this discovery.

With the exponential growth of global human populations and consequently the high demand on food production, the consumption of total N-fertiliser, by far the largest fertiliser base, has not reduced. On the contrary, it will increase even more since the demands of feeding a world population are expected to double by 2050 (Vijayakumar, 2018). The World Fertiliser Trends and Outlook to 2018 report, from the Food and Agriculture Organization of the United Nations (FAO), found that just in the last decade, consumption of total N-fertiliser has had an annual growth of 1.4%. Total N-fertiliser demand in 2013 was estimated to be 111.4 millions tonnes (Tg), and is estimated to reach 119.4 Tg by the end of 2018 (Food and Agriculture Organization of the United Nations., 2015).

Unfortunately, all the effort and consumption of energy, not to mention significant costs to the environment, of producing synthetic nitrogen with the aim of increasing crop yields, is not fully realised, since less than the 50% of fertilisers applied in agriculture are able to be utilised by plants (Hu *et al.*, 2010). Resultant leaching from fields contributes to nitrate contamination of soils and ground water. This leads to consequences for human health and creates environmental problems that are becoming increasingly serious; eutrophication of water systems, impacts on biodiversity and vegetation, acidification of soils and greenhouse gas (GHG) production (Maathuis, 2009; Hu *et al.*, 2010; Islam *et al.*, 2011; Santi *et al.*, 2013).

Additionally, due to the huge costs of this product, low-income countries have not enjoyed the benefits of easy access to chemical fertilisers (Conway & Barbie, 1988; Tscharntke *et al.*, 2012). This has reinforced a social and economic gap between low-

income countries and countries that can afford it. Therefore, use of synthetic N-fertilisers compromises the achievement of agricultural sustainability and it is necessary to find alternatives to tackle this problem (Tilman *et al.*, 2002; Graham & Vance, 2003).

Rhizosphere

In the early 1900s, Lorenz Hiltner coined the term rhizosphere, to describe the soil immediately surrounding the roots that is under the influence of the plant (Hartmann *et al.*, 2008). However, more specifically, the rhizosphere is characterised as the local environment that is influenced by living roots, and many interactions between the soil, microorganisms and the plant take place in this area (Bais *et al.*, 2006). In the rhizosphere, plant roots not only need to compete with other neighbouring plant species for space, water and mineral nutrients, but they also compete with a diversity of microbes, including nematodes, fungi and invertebrates for organic material (Ryan *et al.*, 2001; Franche *et al.*, 2008; Turner *et al.*, 2013). The interactions that roots can have within the rhizosphere can either be symbiotic, such as the association of mycorrhizal fungi and nitrogen-fixing bacteria; or negative, such as the association with parasitic plants, pathogenic microorganisms and insects. As a consequence of these interactions, a large amount of root exudates, organic and inorganic substances, are exchanged between the roots and soil (Turner *et al.*, 2013). These exudates act as inhibitors of competing plant species, provide resistance to pests and last, but not least, act as messengers attracting and initiating symbiotic interactions with the plant

(Walker *et al.*, 2003; Bertin *et al.*, 2003; Bais *et al.*, 2006; Murray, 2011; Pini *et al.*, 2017).

Biological Nitrogen Fixation (BNF) and it is importance in agriculture

Only some prokaryotes are able to use atmospheric nitrogen through a process known as Biological Nitrogen Fixation (BNF). These bacteria are called diazotrophs and encode nitrogenase, the enzyme complex that catalyses the conversion of N_2 from the atmosphere into NH_3. Nitrogen reduction is a complicated reaction. The result of net reduction of molecular nitrogen to ammonia is generally accounted for by the following energetically expensive reaction (Dresler-Nurmi *et al.*, 2009; Lee *et al.*, 2014):

$$N_2 + 16\,ATP + 8e^- + 8H^+ \rightarrow 2NH_3 + H_2 + 16\,ADP + 16P_i$$

Diazotrophic bacteria are considered a major group of Plant Growth-Promoting Rhizobacteria (PGPR). They include both free-living and symbiotic organisms. BNF, which provides about 65% of the nitrogen available on earth, is the greatest contributor of fixed nitrogen to the nitrogen cycle. It is estimated to contribute 110 Tg of nitrogen per year (Tg N/year) on land and 140 Tg N/year in oceans, with the *rhizobium*-legume symbiosis being the largest contributor of fixed nitrogen (Dixon & Kahn, 2004; Galloway *et al.*, 2004; Franche *et al.*, 2008; Meena & Lal, 2018; Vijayakumar, 2018).

BNF is of great importance in the nitrogen cycle and has been implemented in agriculture for many years, especially by farmers that cannot afford chemical fertilisers and in recent years has played an important role in organic crop systems. BNF in

agriculture is important since a bioavailable nitrogen is generated on the roots of legume crops. Other non-legume crops grown in rotation or at the same time benefit from the input of fixed nitrogen. This benefit can be perceived by the increased yields compared to areas without chemical fertilisers or a BNF system (O'Hara, 1998). The use of BNF in agriculture ranges from seasonal crop-rotation with legumes, simultaneous legume and non-legume crops, to the most studied case of rhizobial inoculation of legume crops (Zahran, 1999). The use of BNF in legumes can contribute to reducing GHG emissions per unit area by five to seven times compared to other crops (Bohlool *et al.*, 1992; Peoples *et al.*, 1995; Vijayakumar, 2018).

Rhizobia

The diazotrophic bacteria involved in the processes of nitrogen fixation can be divided into three groups: free living (e.g. *Azotobacter*), associative symbiotic (e.g. *Azospirillum*) and symbiotic diazotrophs (e.g. *Frankia* and rhizobia). Only the symbiotic diazotrophs are capable of infecting the plant root and inducing the plant to form nodules where atmospheric nitrogen fixation takes place. *Frankia* species (Gram positive) typically have a symbiosis with non-leguminous plants, while rhizobia (Gram negative) establish a symbiosis mainly with a wide variety of plant species belonging to the family *Leguminosae* (*Fabaceae*). Legumes include numerous important crops such as *Phaseolus vulgaris* (common bean), *Pisum sativum* (pea), *Glycine max* (soybean), *Trifolium* spp. (clover), *Medicago* spp. (alfalfa), and *Arachis hypogaea* (peanut). The exception to this being the non-legume *Parasponia* (van Velzen *et al.*, 2018). Rhizobia can inhabit both soil and nodules. When they grow outside their plant host they need

to survive in the rhizosphere for years as non-symbiotic rhizobia, although the adaptations that allow a successful life in the soil environment are not fully understood yet. Despite being unable to fix nitrogen on their own in the soil, rhizobia are able to do so in symbiosis with the plant host (Bais *et al.*, 2006; Dresler-Nurmi *et al.*, 2009; Santi *et al.*, 2013; Poole *et al.*, 2018).

Rhizobia are phylogenetically disparate α- and β-proteo-bacteria (Masson-Boivin *et al.*, 2009) and there have been several modifications in the taxonomy and names in the genera and species (Young, 1996; Young *et al.*, 2006; Martinez *et al.*, 2008; de Lajudie & Young, 2018). Currently rhizobia do not form a single taxonomic cluster, they are members of the α and β subgroup of the *Proteobacteria* lineages. Although, more known examples are confined to the α-*proteobacteria* class. For practical purposes, the α proteobacterial rhizobia can be divided into three clusters: the fast- or moderately fast-growing genera *Rhizobium*, *Allorhizobium, Sinorhizobium* and *Mesorhizobium*, the slow-growing *Bradyrhizobium*, and the stem-nodulating *Azorhizobium*, and the β-*proteobacteria* rhizobia consist of two genera: *Burkholderia* and *Cupriavidus* (Dresler-Nurmi *et al.*, 2009; Masson-Boivin *et al.*, 2009; Bontemps *et al.*, 2010). The similarities in the rhizobia genomes so far published are that they have a large genome (5.4-9.2 Mb), and contain abundant transport, metabolic and regulatory genes. Two widely distributed sets of genes encode these functions in most rhizobia: the nodulation (*nod*) and nitrogen fixation (*nif*) genes (Masson-Boivin *et al.*, 2009). Because of their relevance to agriculture, most research has concentrated on rhizobia that nodulate crop and forage legumes such as pea nodulated by *Rhizobium leguminosarum* biovar *viciae* (Rlv), clover by *Rl* biovar *trifolii* in clover, common bean

by Rl biovar *phaseoli* and *R. etli*, soybean *by Bradyrhizobium japonicum* and *Sinorhizobium fredii*, and alfalfa *by S. meliloti*.

Legumes

Legumes are cultivated throughout the world and are a principal crop, valued as a relatively inexpensive source of protein and energy. They have been used as an important source of protein in low-income countries because the dry mature seeds (pulses) can contribute 33% of the dietary protein nitrogen needs of humans and they have high levels of lysine. The primary human dietary legumes are: common bean, pea, *Cicer arietinum* (chickpea), *Vicia faba* (broad bean), *Cajanus cajan* (pigeon pea), *Vigna unguiculata* (cowpea) and *Lens esculenta* (lentil) (McIntosh & Topping, 2000; Graham & Vance, 2003). Their use is also important for animal nutrition, for example *Vigna radiate* (mung bean), broad bean, soybean, peanuts and other variety of peas (Graham & Vance, 2003; NRA, CIRAD, 2015). In addition to traditional food and forage consumptions, the use of legumes has been diversified. In the last couple of decades, the food industry has discovered novel uses for specific legumes as dairy-like products, additives in flour, pasta, snacks and confectionary. Legumes have been used industrially to prepare biodegradable plastics, oils, gums, inks, textiles, papers, pill formulation and more recently for biodiesel production (García, *et al.,* 1998; Graham & Vance, 2003; Morris, 1997).

The rhizobia-legume symbiosis

The three subfamilies of Fabaceae (Caesalpinioideae, Mimosoideae, Papilionoideae) contain species that form root nodules. It has long been known that the nodulation ratio is different in species of the three subfamilies: rare in Caesalpinioideae, common in Mimosoideae and very common in Papilionoideae (Pueppke & Broughton, 1999; Sprent & James, 2007). Rhizobia and their host must recognize each other for the nodulation process to begin and nodule morphology depends on the species of legume. There is no general consensus on the morphological types of legume nodule and their internal structure. Some authors use a classification of up to five types of nodules, *Aeschynomenoid*, *Desmodioid*, *Indeterminate*, *Lupinoid*, and *Primitive* (Sprent, 2008), but for practical purposes it is more common to use the classification of *Determinate* and *Indeterminate* nodules.

Establishment

The establishment of the symbiosis between rhizobia and legumes is a process with multiple stages, which is initiated by pre-infection events in the rhizosphere (for reviews, refer to: Oldroyd *et al.*, 2011; Oldroyd, 2013; Downie, 2014; and Poole *et al.*, 2018). In broad terms (Fig. 1.1), rhizobia respond to plant root exudates, mostly flavonoids, and are attracted to the legume's roots, which induce the rhizobia to produce one or more lipooligosaccharides (LCO), also known as Nod factors (NF). Different NF are produced by different rhizobial species, which is important for the specificity of interactions between different rhizobia and host legume species. Subsequently, the rhizobia attach to the plant root surface to young root hairs.

Rhizobia often enter through root hairs, although infection through epidermal cracks is possible. They grow and divide inside tubules called infection threads (IT). The IT progression agrees with cell division; cortical cells divide to form a nodule primordium and the location in the root cortex depends on the individual plant. In this stage, plant and bacteria continue with their dialogue through cysteine-rich (NCR) peptides, nodulins and flavonoids. The final step is the establishment of bacteroids in the central zone of the nodule, which provides a suitable environment for N_2 fixation since supply of carbon (as dicarboxylates such as succinate and malate) is provided to the bacteria and a low O_2 environment protects the nitrogenase. Depending on the legume species, nodules can be either indeterminate or determinate.

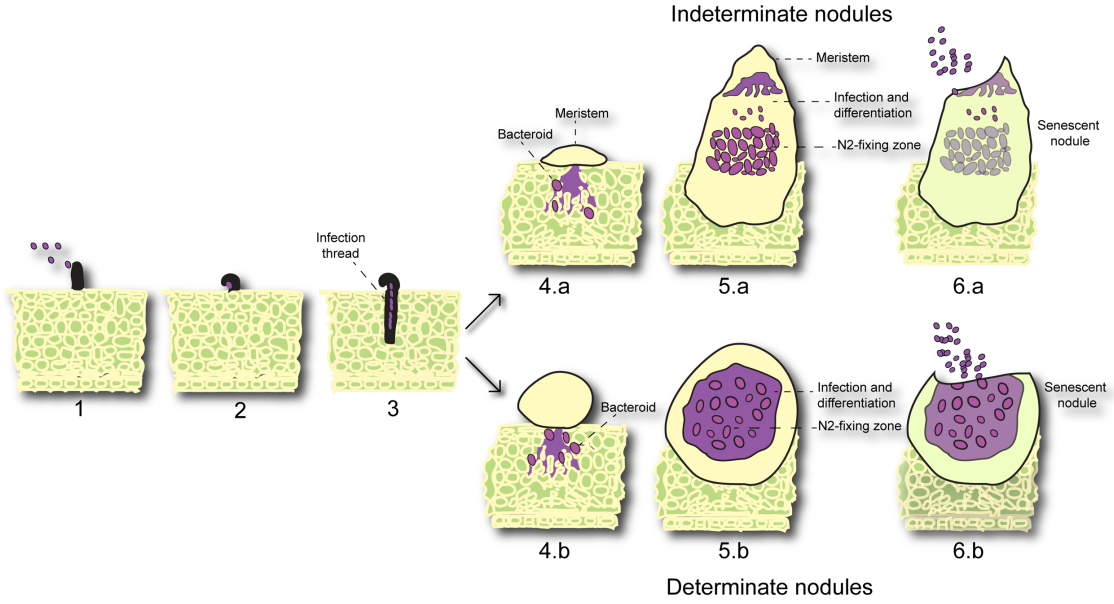


Figure 1.1 Schematic developmental stages of indeterminate and determinate legume nodules. 1) Flavonoids secreted by root hairs attract rhizobia and produce NFs. 2) Secreted NF cause root hairs to curl, 3) trapping the rhizobia and IT initiates. 4.a) In indeterminate nodules, a meristem keeps growing and bacteroids are

released from the IT, leading to successive zones of rhizobial invasion and differentiation. 5.a) Mature indeterminate nodule with distinct zones. In contrast, determinate nodules 4.b) do not have a persistent meristem and mature nodules 5.b) has a homogenous infection/differentiation zone and N_2-fixing zone. 6a-b) Viable bacteria can be released from a senescent nodule returning to the rhizosphere for future invasion.

Indeterminate nodules

Indeterminate nodules have a growing tip meristem and the mature modules tend to have an elongated, cylindrical and branched form. In the indeterminate nodule in Figure 1.2 we can find four distinct zones. **Zone I** is the growing tip of the meristem, **zone II** is the area where rhizobia infect the root plant cells, **zone III** is the area where bacteroids are established and nitrogen fixation is carried out, and **zone IV** is the area where older infected root plant cells and bacteroids are senescing. It is know that indeterminate nodules also contain a **zone V**, where a new batch of cells is released from remaining IT, leading to the reinvasion of plant cells that have completely senesced (Puppo *et al.*, 2004; Poole *et al.*, 2018). Indeterminate nodules are found in the majority of temperate species such as alfalfa, pea and clover (Timmers *et al.*, 2000; Terpolilli *et al.*, 2012).

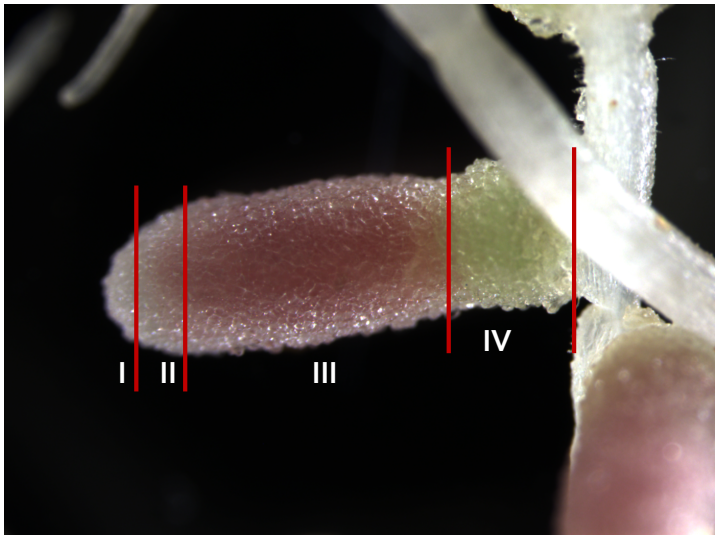


Figure 1.2 Microscopic view of a six-week-old pea nodule, used as an example of an indeterminate nodule. Lines emphasize four distinct zones. Zone I, Tip of the meristem. Zone II, Infection and differentiation. Zone III, pink colour due to expression of leghaemoglobin indicating an efficient symbiosis. Zone IV Senescence area, green colour indicating an alteration of leghaemoglobin.

Determinate nodules

Determinate nodules originate in the outer or middle cortex of the root and tend to have a spherical shape without showing a persistent meristem (Figure 1.3). The number of cells in young and old nodules does not change. These nodules do not show a clear developmental gradient, having a homologous population of bacteroids and therefore one homogenous N_2-fixing zone. Since there is no regeneration of N_2-fixing bacteroids, fixation of N_2 is temporary and the life of the nodule ends with senescence, starting in the centre of the nodule. Legumes that form determinate nodules are predominately tropical and subtropical species, for example common bean and

soybean, but also include some temperate species such as *Lotus japonicus* (Maunoury *et al.*, 2008; Ferguson *et al.*, 2010).



Figure 1.3 Microscopic view of six-week-old common bean nodules used as examples of determinate. Spherical shaped nodule without showing a persistent meristem.

Mixed nodules

For a long time it was accepted that nodules were formed by single clone infection, but experimental results proved, several decades ago, that both indeterminate and determinate nodules can contain more than one rhizobial strain (Lindemann *et al.*, 1974; Johnston & Beringer, 1975; Dowling & Broughton, 1986). The frequency of double (Figure 1.4) or multiple strain occupancy is enough that it should be carefully considered in rhizobial studies, particularly in competition assays (May & Bohlool, 1983; Moawad & Schmidt, 1987).

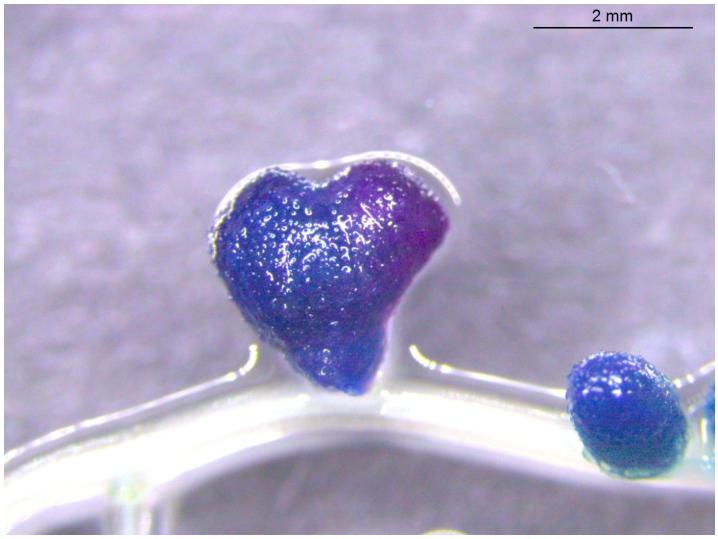


Figure 1.4 Microscopic view of an indeterminate mixed nodule. Pink colour due to Rlv UPM791 expressing *gusA* marker gene and blue colour due to Rlv3841 expressing *celB* marker gene.

Host specificity

Host specificity plays an important role in the establishment of a successful legume-rhizobia symbiosis. Some rhizobial strains are very selective and have a narrow host range, such as *R. leguminosarum* bv. *trifolii* strains that can only nodulate plants of the genus *Trifolium* (Broughton & Perret, 1999; Poole *et al.*, 2018). Other studies have identified that some rhizobial strains can have more than one plant partner, as in the case of *Ensifer* sp. NGR234 that can nodulate over a hundred legumes, as well as the non-legume Parasponia. However, this promiscuity does not necessary lead all the time to effective nitrogen-fixing nodules (Pueppke & Broughton, 1999).

Rhizobial adaptation

Rhizobial strains need to survive and grow as free-living in the soil environment and during rhizosphere colonisation for long periods. The resulting adaptation of rhizobial strains to hostile soil factors, i.e. saline conditions, drought or even presence of heavy metals, is a pre-requisite to success in competition among their pairs. Although the adaptations that allow a successful life in the soil environment are not fully understood yet, soil pH has been shown to affect the competitiveness of *R. leguminosarum* strains (Vargas & Graham, 1989; Frey & Blum, 1994; Anyango *et al.*, 1998).

Rhizobial competitiveness

Rhizobial competitiveness refers to the performance of a specific strain in terms of growth, survival and nodule formation on a particular legume host, in relation to the performance of another strain in the same environment. To predict the competitiveness of one strain it is essential to know the quantitative relationship between the proportion of that specific strain in the inoculum and the proportion of nodules formed. Expressed in a different way, if two strains are equality competitive an equal relationship is expected between the percentage of nodules formed from each strain and their percentage of cells in the inoculum. (Amarger & Lobreau, 1982; Sessitsch *et al.*, 1997; Onishchuk *et al.*, 2017).

N_2-fixing effectiveness and ineffectiveness

Rhizobial strains have the ability to establish inside the roots and/or stem nodules of legume plants. Symbiotic effective rhizobial strains are those which fix atmospheric N_2 to their host, and a strain with N_2-fixing effectiveness, throughout this thesis simply termed 'effectiveness', refers to the capacity of a specific strain to fix a higher quantity of nitrogen compared to other strains in a particular legume host. Rhizobia that are able to infect nodules whilst fixing very little or no N_2 for the host are also found, and we can call these ineffective associations (Giller, 2001). Effectiveness and ineffectiveness are not constant across different hosts; in another words, one strain can be effective in one host and ineffective in another.

We can find legumes nodulated by highly competitive but not necessarily effective rhizobial strains, resulting in less than optimal plant growth (Pueppke & Broughton, 1999). If we want to harness the potential of the rhizobia-legume symbiosis with the aim of achieving sustainable agriculture, it will be necessary to first develop a better means of identifying strains that possess both features, competitiveness *and* effectiveness.

Rhizobial identification

It is complicated to directly observe the interaction between bacteria and plants but the use of bacterial markers, which give an easily detected signal on expression, can help with this task. Examples include genes encoding catechol 2,3-dioxygenase (*xylE),* ß-galactosidase *(lacZ),* ß-glucuronidase (*gusA*) and the thermostable ß-glucosidase (*celB*). The importance of these reporter genes is that they allow closely related strains

of bacteria to be told apart, which means that an individual marked strain can be distinguished from non-marked strains (Wilson, 1995).

Another example of reporter system used in the rhizobial-legume system it is the operon *luxCDABE* (Fan *et al.*, 2008) which has been incorporated to a very stable vector that can be used even in the absence of antibiotic selection. This *lux* reporter plasmid was used to identify rhizobial attachment to roots by Frederix *et al.*, 2014 and subsequently, Pini *et al.*, 2017 developed a collection of bioreporters to analyse the process of colonisation and symbiosis revealing compounds important in the rhizobium-legume association.

Fluorescent proteins such superfolder green florescent protein (sfGFP) or the red fluorescent protein mCherry have helped to study plant-bacterial interaction and have permitted high-throughput screenings (Pédelacq *et al.*, 2006; Singer *et al.*, 2010). Having these kinds of bioreporters in plasmids streamlines the process of observing the interaction between bacteria and plants, and you can follow and recover bacteria throughout the assay. Another advantage of this system is that specific promoter[2] can be used to follow the expression under different environmental conditions or precise steps of the rhizobium-legume association (Pini *et al.*, 2017).

Unmarked-strain identification can be possible by the use of naturally occurring and induced antibiotic resistance (Poysti *et al.*, 2007; Geddes & Oresnik, 2012) but it does not lend itself to rapid analysis of a large number of samples since individual analysis of nodules is required. Metagenomics approaches for high-throughput identification of

[2] A promoter is a region of DNA that initiates transcription of a particular gene or operon.

strains do exist, such as the small-subunit ribosomal RNA (16S rRNA) gene locus that is highly conserved across microbes (Pace *et al.*, 1986), but unfortunately use of this method is not possible here since an inability to distinguish between closely related strains, as they all have the same or very similar 16S rRNA, would result in low sensitivity.

Methods used to study rhizobial competition

In the case of the rhizobial-legume system, *gusA* and *celB* have had very good results as bioreporters to follow inoculation with one or two rhizobial strains (Sessitsch *et al.*, 1998; Fan *et al.*, 2008). Combining these two systems has brought advantages in the evaluation of competitiveness between rhizobial species because it allows efficient scoring of single or double nodule infections without requiring sophisticated equipment (Sánchez-Cañizares & Palacios, 2013). The disadvantages are that this method is highly time-consuming since chromosomal integration of *gusA* and *celB* is necessary, it is not possible to recover the bacteria inside the nodule since a harsh fixation process and enzymatic reaction is necessary to distinguish nodule occupancy, and it is only possible to assess two strains at the same time.

Methods such as phage typing (Schwinghamer & Brockwell, 1978; Bromfield *et al.*, 1987), intrinsic antibiotic resistance (IAR), which can naturally occur or be induced (Stein *et al.*, 1982; Young & Chao, 1989), or the use of PCR-fingerprints which target specific plasmids profiles or genes (Laguerre *et al.*, 1996; Thies *et al.*, 2001; Lakzian *et al.*, 2002; Duodu *et al.*, 2009; Mishra *et al.*, 2012), have the ability to monitor strain-level changes. However, even in the most recent studies, only up to six strains have

been used (Ji *et al.*, 2017; Lardi *et al.*, 2017) and methods that allow for the assessment of more strains at once and are less laborious are therefore still needed.

Commercial inoculants

The *Rhizobium*-legume symbioses have been studied for more than hundred years and even the first industrial production of rhizobial inoculant had begun by the end of the 19th century (Parnell *et al.*, 2016). This symbiotic associations between leguminous plants and root nodule bacteria have been estimated to fix approximately 80% of the biologically-fixed nitrogen in agricultural areas (O'Hara, 1998). However, use of *Rhizobium*-legume symbioses as inoculants is vulnerable because of an important problem in microbial ecology concerning the efficacy of nodulation, known as the "rhizobial competition problem". Under controlled conditions, specific strains of rhizobia increase nitrogen fixation when they are applied to legumes, nevertheless, once these strains are applied to legumes under real agricultural conditions, they often fail to nodulated legumes compared to indigenous strains due to the soil conditions, rooting depth, humidity, temperature and the inoculant formulation (López-García *et al.*, 2002; Naeem *et al.*, 2004; Yates *et al.*, 2011; Drew *et al.*, 2012). This competition problem might be tackled if it were possible to identify indigenous strains from a specific soil that has both features, competitiveness and effectiveness, and then used them as tailormade inoculants.

Aim of the project

Despite intensive research for many years in the *Rhizobium*-legume symbioses, we are still at the beginning of understanding the factors that influence the survival, persistence, efficiency and competition of rhizobial strains in various environmental conditions. The principal problem that has limited this understanding is the complex procedure in nodule competition assays whereby it is necessary to manually remove root nodules, crush them to extract the strain and then conduct a strain differentiation test on antibiotic plates (Oresnik *et al.*, 1999; Poysti *et al.*, 2007; Geddes & Oresnik, 2012). Techniques such as acetylene reduction assays (ARA) for effectiveness, and the combination of *gusA* and *celB* for scoring competitiveness have helped us in lab assays, but not to the extent required due to the limited numbers of strains that can be assessed at the same time (Hardy *et al.*, 1968; Sánchez-Cañizares & Palacios, 2013).

Fortunately, NGS and multi-fragment assembly techniques, such as Golden Gate Cloning, offer a flexible and cost-effective strategy for high-throughput assays to aid identification of competitive and effective rhizobial strains in realistic conditions. This can be made possible through the use of biomarkers and barcodes allowing the genetic analysis of hundreds of field isolates at the same time (Engler *et al.*, 2008; Engler & Marillonnet, 2011; Fadrosh *et al.*, 2014).

Combining these two approaches, the principal aim of this project is to **engineer a molecular tool for high-throughput identification of competitive and effective rhizobial strains.**

The strategy that we will follow to engineer this tool is: i) to develop a broad host range vector which demonstrates excellent stability in rhizobia, even in the absence of antibiotics. This vector should be compatible with Golden Gate cloning to build in a high-throughput approach for the large number of plasmids required in this project; ii) the use of a reporter gene driven by nodule-specific expression in biovars and strains of *Rhizobium leguminosarum* which will be used to directly observe plant-bacterial interactions, assess competitiveness of strains without reducing their fitness and obtain information about their nitrogen activity rates; iii) will include unique synthetic nucleotide sequences, in order that multiple rhizobial strains can be marked and identified from one another for nodule occupancy in a single experiment through the use of next generation sequencing (NGS).

This tool has the potential to help pre-evaluate strains as possible inoculants on a selected legume crop and local soil before launching large-scale field use.

Developing better inoculants may contribute to obtaining high yields from legume crops and consequently contributing to the minimum necessary global food demand via sustainable agriculture.

Chapter 2: General materials and methods

Media, antibiotics and other chemicals

Media

Escherichia coli strains were grown at 37°C in Luria Bertani broth (LB) (10g/l tryptone, 5 g/l yeast extract, 5 g/l NaCl) (Sambrook *et al.*, 1989). For solid media, 1.4 w/v agar was added.

Rhizobial strains were grown at 28°C in Tryptone yeast (TY) media (5 g/l tryptone, 3 g/l yeast extract, 6 mM CaCl2) (Beringer, 1974) or Universal Minimal Salts (UMS) (Poole *et al.*, 1994) (0.5 mM K_2HPO_4, 0.5 g/l $MgSO_4.7H_2O$, 0.2 g/l NaCl, 4.19 g/l MOPS, pH adjusted to 7.0) to which, after autoclaving, was added the following filter sterilised chemicals: 1 ml of iron stock solution (12 g/l $FeSO_4.7H_2O$), 1 ml of calcium stock solution (75 g/l $CaCl_2.2H_2O$) and 1 ml of a stock solution containing 1 g/l thiamine hydrochloride, 2 g/l D-Pantothenic acid calcium salt, 100 mg/l biotin, 10 mM glucose, 10 mM NH_4Cl and 1 µl of 1000x vitamin solution (0.375 g/l EDTA-Na_2, 0.16 g/l $ZnSO_4.7H_2O$, 0.2 g/l, $NaMoO_4$, 0.25 g/l H_3BO_3, 0.2 g/l $MnSO_4.4H_2O$, 0.02 g/l $CuSO_4.5H_2O$, 1 g/l $CoCl_2.6H_2O$). For solid media 1.75 w/v agar was added.

Antibiotics

Table 2. 1 Working concentration (µg/ml) of antibiotics used in this work

	E. coli	Rhizobial strains
Ampicillin (Amp)	100	
Chloramphenicol (Chl)	10	20
Gentamycin (Gm)	10	20
Kanamycin (Kan)	20	40
Neomycin (Nm)	20	40
Nitrofurantoin (Nit)	10	10
Rifampicin (Rf)	N/A	50
Spectinomycin (Sp)	50	N/A
Streptomycin (Str)	25	500
Tetracycline (Tet)	5	5

Other chemicals

For *E. coli* ST18 strains 5-aminolevulinic acid (ALA) was added at 50 µg/L

For blue- white screening of *E. coli* X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) was added to LB agar at 40 µg/ml. For blue screening of *Rhizobium gusA*-encoding strains, X-glcA (5-bromo-4-chloro-3-indolyl-β-D-glucuronide) was added to agar plates at 40 µg/ml.

For blue screening of *Rhizobium gusA*-encoding strains, X-glcA (5-bromo-4-chloro-3-indolyl-β-D-glucuronide) was added to agar plates at 40 µg/ml.

Distilled water agar (DWA) was prepared by adding 1% w/v agar to distilled (D-)H_2O and autoclaved.

Bacterial strains, plasmids and primers

Table 2. 2 Strains used in this study

Strain	Details	Source
***E. coli* strains**		
α-Select Gold	Competent cells. Genotype: F - *deoR endA1 recA1 relA1 gyrA96 hsdR17(rk -, mk +) supE44 thi-1 phoA Δ(lacZYA argF)U169 Φ80lacZΔM15*λ -	Bioline
DH5α	Competent cells. Genotype: F- *Φ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17(rk-, mk+) phoA supE44 thi-1 gyrA96 relA1* λ-	Invitrogen
ST18	S17 *Δpir ΔhemA*	Thoma and Schobert, 2009
Rhizobial strains		
Rlv3841	*Rhizobium leguminosarum* bv. *viciae,* derivative of strain 300, Str^r	Johnston and Beringer (1975)
RlvA34	*R. leguminosarum* bv. *viciae*, Str^r	Downie et al. (1983)
CFN42	*Rhizobium etli*	Quinto et al. (1982)
CIAT899	*Rhizobium tropici*, Rf^r	Graham et al. (1982)
4292	*R. leguminosarum* bv. *phaseoli*, Rf^r	Downie et al. (1983)
1021	*Sinorhizobium meliloti*	Meade et al. (1982)
WSM1325	*R. leguminosarum* bv. *trifolii*	Reeve et al. (2010b)
WSM419	*Sinorhizobium medicae*	Reeve et al. (2010a)
WSM1022	*Sinorhizobium meliloti*	Terpolilli et al. (2013)
HH103	*Sinorhizobium fredii*	Weidner et al. (2012)
WSM1521	Host of origin: *Lathyrus*. Effectiveness class in pea: high, Nit^r	John Howieson
54 n4	Host of origin: *P. sativum*. Effectiveness class in pea: high, Nit^r	John Howieson
60 n1	Host of origin: *P. sativum*. Effectiveness class in pea: medium, Nit^r	John Howieson
71 n4	Host of origin: *P. sativum*. Effectiveness class in pea: medium, Nit^r	John Howieson
SARDI 962	Host of origin: *P. sativum*. Effectiveness class in pea: low, Nit^r	John Howieson
SU303	Host of origin: unknown. Effectiveness class in pea: high, Nit^r	John Howieson
WSM1455	Host of origin: *V. faba*. Effectiveness class in pea: medium, Nit^r	John Howieson
WSM1475	Host of origin: *V. faba*. Effectiveness class in pea: medium, Nit^r	John Howieson

WSM1480	Host of origin: *V. faba*. Effectiveness class in pea: medium, Nit^r	John Howieson
WSM1481	Host of origin: *Vicia* spp. Effectiveness class in pea: low, Nit^r	John Howieson
WSM1488	Host of origin: *V. faba*. Effectiveness class in pea: low, Nit^r	John Howieson
WSM1529	Host of origin: *V. faba*. Effectiveness class in pea: high, Nit^r	John Howieson
WSM4458	Host of origin: *P. sativum*. Effectiveness class in pea: low, Nit^r	John Howieson
WSM4459	Host of origin: *P. sativum*. Effectiveness class in pea: low, Nit^r	John Howieson
WSM4460	Host of origin: *P. sativum*. Effectiveness class in pea: medium, Nit^r	John Howieson
WSM4461	Host of origin: *P. sativum*. Effectiveness class in pea: high, Nit^r	John Howieson
WSM4462	Host of origin: *P. sativum*. Effectiveness class in pea: low, Nit^r	John Howieson
WSM4645	Host of origin: *P. sativum*. Effectiveness class in pea: high, Nit^r	John Howieson
WSM937	Effectiveness class in pea: high, Nit^r	John Howieson
3841*celB*	*R. leguminosarum* bv. *viciae*, Str^r	Sánchez-Cañizares & Palacios, (2013)
UPM791*gusA*	*R. leguminosarum* bv. *viciae*, Str^r	Sánchez-Cañizares & Palacios, (2013)
UPM791	*R. leguminosarum* bv. *viciae*, Str^r	Sánchez-Cañizares & Palacios, (2013)
VSX1	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar *et al.*, (2015)
VSX2	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)
VSX3	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)
VSX4	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)
VSX5	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)
VSX6	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)
VSX7	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)
VSX8	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)
VSX9	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)
VSX10	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)
VSX11	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)
VSX14	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)
VSX15	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)
VSX16	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)
VSX17	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)
VSX18	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)

VSX19	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)
VSX20	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)
VSX21	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)
VSX22	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)
VSX23	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)
VSX24	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)
VSX25	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)
VSX26	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)
VSX27	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)
VSX28	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)
VSX29	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)
VSX30	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)
VSX31	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)
VSX32	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)
VSX33	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)
VSX34	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)
VSX35	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)
VSX36	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)
VSX37	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)
VSX38	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)
VSX39	*R. leguminosarum* bv. *viciae*, Nit^r	Kumar et al., (2015)
13	*Rhizobium* sp., Nit^r	Euan James
23_1	*Rhizobium* sp., Nit^r	Euan James
24	*Rhizobium* sp., Nit^r	Euan James
263	*Rhizobium* sp., Nit^r	Euan James
2B-1	*Rhizobium* sp., Nit^r	Euan James
364	*Rhizobium* sp., Nit^r	Euan James
367	*Rhizobium* sp., Nit^r	Euan James
370	*Rhizobium* sp., Nit^r	Euan James
387	*Rhizobium* sp., Nit^r	Euan James
388	*Rhizobium* sp., Nit^r	Euan James
42	*Rhizobium* sp., Nit^r	Euan James
535	*Rhizobium* sp., Nit^r	Euan James
536	*Rhizobium* sp., Nit^r	Euan James
585	*Rhizobium* sp., Nit^r	Euan James
CellTech	*Rhizobium* sp., Nit^r	Euan James
Corus 1+2 1-1	*Rhizobium* sp., Nit^r	Euan James
Corus 16 2-2	*Rhizobium* sp., Nit^r	Euan James
E2-1A	*Rhizobium* sp., Nit^r	Euan James
E2-1B	*Rhizobium* sp., Nit^r	Euan James

JED2	*Rhizobium* sp., Nit^r	Euan James
LA-2A	*Rhizobium* sp., Nit^r	Euan James
LegTech	*Rhizobium* sp., Nit^r	Euan James
Magnus 1+2 (2)	*Rhizobium* sp., Nit^r	Euan James
Magnus 16 (2)	*Rhizobium* sp., Nit^r	Euan James
PB3-3	*Rhizobium* sp., Nit^r	Euan James
PB6-3B	*Rhizobium* sp., Nit^r	Euan James
PEA NW	*Rhizobium* sp., Nit^r	Euan James
VF2	*Rhizobium* sp., Nit^r	Euan James
VF5	*Rhizobium* sp., Nit^r	Euan James
G004	Isolated using legume host: *P. sativum*, Nit^r	Jorrin & Imperial, (2015)
G007	Isolated using legume host: *P. sativum*, Nit^r	Jorrin & Imperial, (2015)
G008	Isolated using legume host: *P. sativum*, Nit^r	Jorrin & Imperial, (2015)
G011	Isolated using legume host: *P. sativum*, Nit^r	Jorrin & Imperial, (2015)
G016	Isolated using legume host: *P. sativum*, Nit^r	Jorrin & Imperial, (2015)
G028	Isolated using legume host: *P. sativum*, Nit^r	Jorrin & Imperial, (2015)
G051	Isolated using legume host: *P. sativum*, Nit^r	Jorrin & Imperial, (2015)
G067	Isolated using legume host: *P. sativum*, Nit^r	Jorrin & Imperial, (2015)
G073	Isolated using legume host: *P. sativum*, Nit^r	Jorrin & Imperial, (2015)
G077	Isolated using legume host: *P. sativum*, Nit^r	Jorrin & Imperial, (2015)
G083	Isolated using legume host: *P. sativum*, Nit^r	Jorrin & Imperial, (2015)
G088	Isolated using legume host: *P. sativum*, Nit^r	Jorrin & Imperial, (2015)
G093	Isolated using legume host: *P. sativum*, Nit^r	Jorrin & Imperial, (2015)
G094	Isolated using legume host: *P. sativum*, Nit^r	Jorrin & Imperial, (2015)
G099	Isolated using legume host: *P. sativum*, Nit^r	Jorrin & Imperial, (2015)
G108	Isolated using legume host: *P. sativum*, Nit^r	Jorrin & Imperial, (2015)
G109	Isolated using legume host: *P. sativum*, Nit^r	Jorrin & Imperial, (2015)
H005	Isolated using legume host: *V. faba*, Nit^r	Jorrin & Imperial, (2015)
H009	Isolated using legume host: *V. faba*, Nit^r	Jorrin & Imperial, (2015)
H011	Isolated using legume host: *V. faba*, Nit^r	Jorrin & Imperial, (2015)
H012	Isolated using legume host: *V. faba*, Nit^r	Jorrin & Imperial, (2015)
H031	Isolated using legume host: *V. faba*, Nit^r	Jorrin & Imperial, (2015)
H082	Isolated using legume host: *V. faba*, Nit^r	Jorrin & Imperial, (2015)
H127	Isolated using legume host: *V. faba*, Nit^r	Jorrin & Imperial, (2015)
H130	Isolated using legume host: *V. faba*, Nit^r	Jorrin & Imperial, (2015)
H174	Isolated using legume host: *V. faba*, Nit^r	Jorrin & Imperial, (2015)
H178	Isolated using legume host: *V. faba*, Nit^r	Jorrin & Imperial, (2015)
L002	Isolated using legume host: *L. culinaris*, Nit^r	Jorrin & Imperial, (2015)
L008	Isolated using legume host: *L. culinaris*, Nit^r	Jorrin & Imperial, (2015)
L010	Isolated using legume host: *L. culinaris*, Nit^r	Jorrin & Imperial, (2015)

L018	Isolated using legume host: *L. culinaris*, Nit^r	Jorrin & Imperial, (2015)
L019	Isolated using legume host: *L. culinaris*, Nit^r	Jorrin & Imperial, (2015)
L070	Isolated using legume host: *L. culinaris*, Nit^r	Jorrin & Imperial, (2015)
L074	Isolated using legume host: *L. culinaris*, Nit^r	Jorrin & Imperial, (2015)
L079	Isolated using legume host: *L. culinaris*, Nit^r	Jorrin & Imperial, (2015)
L082	Isolated using legume host: *L. culinaris*, Nit^r	Jorrin & Imperial, (2015)
L102	Isolated using legume host: *L. culinaris*, Nit^r	Jorrin & Imperial, (2015)
L104	Isolated using legume host: *L. culinaris*, Nit^r	Jorrin & Imperial, (2015)
L111	Isolated using legume host: *L. culinaris*, Nit^r	Jorrin & Imperial, (2015)
L117	Isolated using legume host: *L. culinaris*, Nit^r	Jorrin & Imperial, (2015)
V002	Isolated using legume host: *V. sativa*, Nit^r	Jorrin & Imperial, (2015)
V004	Isolated using legume host: *V. sativa*, Nit^r	Jorrin & Imperial, (2015)
V006	Isolated using legume host: *V. sativa*, Nit^r	Jorrin & Imperial, (2015)
V008	Isolated using legume host: *V. sativa*, Nit^r	Jorrin & Imperial, (2015)
V010	Isolated using legume host: *V. sativa*, Nit^r	Jorrin & Imperial, (2015)
V014	Isolated using legume host: *V. sativa*, Nit^r	Jorrin & Imperial, (2015)
V030	Isolated using legume host: *V. sativa*, Nit^r	Jorrin & Imperial, (2015)
V043	Isolated using legume host: *V. sativa*, Nit^r	Jorrin & Imperial, (2015)
V050	Isolated using legume host: *V. sativa*, Nit^r	Jorrin & Imperial, (2015)
V057	Isolated using legume host: *V. sativa*, Nit^r	Jorrin & Imperial, (2015)
V060	Isolated using legume host: *V. sativa*, Nit^r	Jorrin & Imperial, (2015)
V067	Isolated using legume host: *V. sativa*, Nit^r	Jorrin & Imperial, (2015)
V068	Isolated using legume host: *V. sativa*, Nit^r	Jorrin & Imperial, (2015)
V069	Isolated using legume host: *V. sativa*, Nit^r	Jorrin & Imperial, (2015)
V074	Isolated using legume host: *V. sativa*, Nit^r	Jorrin & Imperial, (2015)
V100	Isolated using legume host: *V. sativa*, Nit^r	Jorrin & Imperial, (2015)
V116	Isolated using legume host: *V. sativa*, Nit^r	Jorrin & Imperial, (2015)
OPS0657	Rlv3841[pOPS0253], Kan^r/Nm^r, Str^r ^{r}	This work
OPS0658	Rlv3841[pOPS0254], Kan^r/Nm^r, Str^r	This work
OPS0659	Rlv3841[pOPS0261], Kan^r/Nm^r, Str^r	This work
OPS0660	Rlv3841[pOPS0263], Kan^r/Nm^r, Str^r	This work
OPS0717	HH103[pOPS0262], Kan^r/Nm^r, Nit^r	This work
OPS0718	HH103[pOPS0263], Kan^r/Nm^r, Nit^r	This work
OPS0719	CFN42[pOPS0262], Kan^r/Nm^r, Nit^r	This work
OPS0720	CFN42[pOPS0263], Kan^r/Nm^r, Nit^r	This work
OPS0721	WSM419[pOPS0262], Kan^r/Nm^r, Chl^r	This work
OPS0722	WSM419[pOPS0263], Kan^r/Nm^r, Chl^r	This work
OPS0723	Sm1022[pOPS0262], Kan^r/Nm^r, Chl^r	This work
OPS0724	Sm1022[pOPS0263], Kan^r/Nm^r, Chl^r	This work
OPS0725	Sm1021[pOPS0262], Kan^r/Nm^r, Str^r	This work

OPS0726	Sm1021[pOPS0263], Kan^r/Nm^r, Str^r	This work
OPS0729	RlvA34[pOPS0262], Kan^r/Nm^r, Str^r	This work
OPS0730	RlvA34[pOPS0263], Kan^r/Nm^r, Str^r	This work
OPS0731	Rl bv. *phaseoli* 4292[pOPS0262], Kan^r/Nm^r, Rf^r	This work
OPS0732	Rl bv. *phaseoli* 4292[pOPS0263], Kan^r/Nm^r, Rf^r	This work
OPS0733	CIAT899[pOPS0262], Kan^r/Nm^r, Rf^r	This work
OPS0734	CIAT899[pOPS0263], Kan^r/Nm^r, Rf^r	This work
OPS0736	WSM1325[pOPS0262], Kan^r/Nm^r, Nit^r	This work
OPS0737	WSM1325[pOPS0263], Kan^r/Nm^r, Nit^r	This work
OPS0738	Rlv3841[pOPS0262], Kan^r/Nm^r, Str^r	This work
OPS0816	Rlv3841[pOPS0314], Kan^r/Nm^r, Str^r	This work
OPS0841	WSM1488[pIJ11268], Tet^r, Nit^r	This work
OPS0842	WSM1488[pLMB640], Tet^r, Nit^r	This work
OPS0843	WSM1488[pIJ11282], Tet^r, Nit^r	This work
OPS0844	WSM1481[pIJ11268], Tet^r, Nit^r	This work
OPS0845	WSM1481[pLMB640], Tet^r, Nit^r	This work
OPS0846	WSM1481[pIJ11282], Tet^r, Nit^r	This work
OPS0851	UPM791[pIJ11268], Tet^r, Str^r	This work
OPS0852	UPM791[pLMB640], Tet^r, Str^r	This work
OPS0853	UPM791[pIJ11282], Tet^r, Str^r	This work
OPS0898	Rlv3841[pOPS0384], Kan^r/Nm^r, Str^r	This work
OPS0899	Rlv3841[pOPS0378], Kan^r/Nm^r, Str^r	This work
OPS0900	Rlv3841[pOPS0380], Kan^r/Nm^r, Str^r	This work
OPS0901	Rlv3841[pOPS0382], Kan^r/Nm^r, Str^r	This work
OPS0902	Rlv3841[pOPS0383], Kan^r/Nm^r, Str^r	This work
OPS0903	Rlv3841[pOPS0377], Kan^r/Nm^r, Str^r	This work
OPS0904	Rlv3841[pOPS0379], Kan^r/Nm^r, Str^r	This work
OPS0905	Rlv3841[pOPS0381], Kan^r/Nm^r, Str^r	This work
OPS1086	Rlv3841[pOPS0750], Kan^r/Nm^r, Str^r	This work
OPS1087	UPM791[pOPS0750], Kan^r/Nm^r, Str^r	This work
OPS1511	Rlv3841[pOPS0385], Kan^r/Nm^r, Str^r	This work
OPS1511	Rlv3841[pOPS0385], Kan^r/Nm^r, Str^r	This work
OPS1512	Rlv3841[pOPS0387], Kan^r/Nm^r, Str^r	This work
OPS1513	Rlv3841[pOPS0386], Kan^r/Nm^r, Str^r	This work
OPS1514	UPM791gusA[pOPS0548], Kan^r/Nm^r, Str^r	This work
OPS1515	Rlv3841celB[pOPS0564], Kan^r/Nm^r, Str^r	This work
OPS1516	Rlv3841[pOPS0491], Kan^r/Nm^r, Str^r	This work
OPS1517	Rlv3841[pOPS0503], Kan^r/Nm^r, Str^r	This work
OPS1518	UPM791[pOPS0491], Kan^r/Nm^r, Str^r	This work

OPS1519	UPM791[pOPS0503], Kan^r/Nm^r, Str^r	This work
OPS1520	WSM1521[pOPS0383], Kan^r/Nm^r, Str^r	This work
OPS1521	WSM1481[pOPS0383], Kan^r/Nm^r, Str^r	This work
OPS1522	WSM1475[pOPS0515], Kan^r/Nm^r, Str^r	This work
OPS1523	WSM1529[pOPS0539], Kan^r/Nm^r, Str^r	This work
OPS1524	WSM1521[pOPS0504], Kan^r/Nm^r, Str^r	This work
OPS1525	WSM1481[pOPS0527], Kan^r/Nm^r, Str^r	This work
OPS1526	Rlv SU303[pOPS0383], Kan^r/Nm^r, Str^r	This work
13 [pOPS0515]	LB-ID 13[pOPS0515], Kan^r/Nm^r, Nit^r	This work
23--1 [pOPS0522]	LB-ID 23--1[pOPS0522], Kan^r/Nm^r, Nit^r	This work
24 [pOPS0539]	LB-ID 24[pOPS0539], Kan^r/Nm^r, Nit^r	This work
2B-1 [pOPS0570]	LB-ID 2B-1[pOPS0570], Kan^r/Nm^r, Nit^r	This work
364 [pOPS0505]	LB-ID 364[pOPS0505], Kan^r/Nm^r, Nit^r	This work
367 [pOPS0545]	LB-ID 367[pOPS0545], Kan^r/Nm^r, Nit^r	This work
387 [pOPS0534]	LB-ID 387[pOPS0534], Kan^r/Nm^r, Nit^r	This work
536 [pOPS0542]	LB-ID 536[pOPS0542], Kan^r/Nm^r, Nit^r	This work
CellTech [pOPS0503]	LB-ID CellTech[pOPS0503], Kan^r/Nm^r, Nit^r	This work
CORUS 1+2 [pOPS0509]	LB-ID CORUS 1+2[pOPS0509], Kan^r/Nm^r, Nit^r	This work
CORUS 16 [pOPS0508]	LB-ID CORUS 16[pOPS0508], Kan^r/Nm^r, Nit^r	This work
E2-1A [pOPS0582]	LB-ID E2-1A[pOPS0582], Kan^r/Nm^r, Nit^r	This work
E2-1B [pOPS0540]	LB-ID E2-1B[pOPS0540], Kan^r/Nm^r, Nit^r	This work
G004 [pOPS0519]	LB-ID G004 pOPS0519], Kan^r/Nm^r, Nit^r	This work
G007 [pOPS0529]	LB-ID G007[pOPS0529], Kan^r/Nm^r, Nit^r	This work
G008 [pOPS0532]	LB-ID G008[pOPS0532], Kan^r/Nm^r, Nit^r	This work
G011 [pOPS0501]	LB-ID G011[pOPS0501], Kan^r/Nm^r, Nit^r	This work
G016 [pOPS0571]	LB-ID G016[pOPS0571], Kan^r/Nm^r, Nit^r	This work
G028 [pOPS0566]	LB-ID G028[pOPS0566], Kan^r/Nm^r, Nit^r	This work
G051 [pOPS0530]	LB-ID G051[pOPS0530], Kan^r/Nm^r, Nit^r	This work
G067 [pOPS0567]	LB-ID G067[pOPS0567], Kan^r/Nm^r, Nit^r	This work
G073 [pOPS0521]	LB-ID G073[pOPS0521], Kan^r/Nm^r, Nit^r	This work
G077 [pOPS0520]	LB-ID G077[pOPS0520], Kan^r/Nm^r, Nit^r	This work
G083 [pOPS0553]	LB-ID G083[pOPS0553], Kan^r/Nm^r, Nit^r	This work
G088 [pOPS0568]	LB-ID G088[pOPS0568], Kan^r/Nm^r, Nit^r	This work
G093 [pOPS0551]	LB-ID G093[pOPS0551], Kan^r/Nm^r, Nit^r	This work
G094 [pOPS0556]	LB-ID G094[pOPS0556], Kan^r/Nm^r, Nit^r	This work
G099 [pOPS0552]	LB-ID G099[pOPS0552], Kan^r/Nm^r, Nit^r	This work
G108 [pOPS0555]	LB-ID G108[pOPS0555], Kan^r/Nm^r, Nit^r	This work
G109 [pOPS0554]	LB-ID G109[pOPS0554], Kan^r/Nm^r, Nit^r	This work

H009 [pOPS0536]	LB-ID H009[pOPS0536], Kan^r/Nm^r, Nit^r	This work
H012 [pOPS0524]	LB-ID H012[pOPS0524], Kan^r/Nm^r, Nit^r	This work
H031 [pOPS0584]	LB-ID H031[pOPS0584], Kan^r/Nm^r, Nit^r	This work
H082 [pOPS0564]	LB-ID H082[pOPS0564], Kan^r/Nm^r, Nit^r	This work
H127 [pOPS0548]	LB-ID H127[pOPS0548], Kan^r/Nm^r, Nit^r	This work
H130 [pOPS0560]	LB-ID H130[pOPS0560], Kan^r/Nm^r, Nit^r	This work
H174 [pOPS0565]	LB-ID H174[pOPS0565], Kan^r/Nm^r, Nit^r	This work
H178 [pOPS0572]	LB-ID H178[pOPS0572], Kan^r/Nm^r, Nit^r	This work
JED2 [pOPS0504]	LB-ID JED2[pOPS0504], Kan^r/Nm^r, Nit^r	This work
L008 [pOPS0557]	LB-ID L008[pOPS0557], Kan^r/Nm^r, Nit^r	This work
L010 [pOPS0575]	LB-ID L010[pOPS0575], Kan^r/Nm^r, Nit^r	This work
L018 [pOPS0576]	LB-ID L018[pOPS0576], Kan^r/Nm^r, Nit^r	This work
L019 [pOPS0558]	LB-ID L019[pOPS0558], Kan^r/Nm^r, Nit^r	This work
L070 [pOPS0577]	LB-ID L070[pOPS0577], Kan^r/Nm^r, Nit^r	This work
L074 [pOPS0578]	LB-ID L074[pOPS0578], Kan^r/Nm^r, Nit^r	This work
L079 [pOPS0579]	LB-ID L079[pOPS0579], Kan^r/Nm^r, Nit^r	This work
L082 [pOPS0580]	LB-ID L082[pOPS0580], Kan^r/Nm^r, Nit^r	This work
L102 [pOPS0527]	LB-ID L102[pOPS0527], Kan^r/Nm^r, Nit^r	This work
L104 [pOPS0528]	LB-ID L104[pOPS0528], Kan^r/Nm^r, Nit^r	This work
LegTech [pOPS0518]	LB-ID LegTech[pOPS0518], Kan^r/Nm^r, Nit^r	This work
PB3-3 [pOPS0541]	LB-ID PB3-3[pOPS0541], Kan^r/Nm^r, Nit^r	This work
PB6-3B [pOPS0506]	LB-ID PB6-3B[pOPS0506], Kan^r/Nm^r, Nit^r	This work
PeanNW [pOPS0544]	LB-ID PeanNW[pOPS0544], Kan^r/Nm^r, Nit^r	This work
Rlv3841 [pOPS0492]	LB-ID Rlv3841[pOPS0492], Kan^r/Nm^r, Nit^r	This work
SU303 [pOPS0517]	LB-ID SU303[pOPS0517], Kan^r/Nm^r, Nit^r	This work
UPM791 [pOPS0491]	LB-ID UPM791[pOPS0491], Kan^r/Nm^r, Str^r	This work
V002 [pOPS0513]	LB-ID V002[pOPS0513], Kan^r/Nm^r, Nit^r	This work
V004 [pOPS0525]	LB-ID V004[pOPS0525], Kan^r/Nm^r, Nit^r	This work
V006 [pOPS0533]	LB-ID V006[pOPS0533], Kan^r/Nm^r, Nit^r	This work
V008 [pOPS0537]	LB-ID V008[pOPS0537], Kan^r/Nm^r, Nit^r	This work
V010 [pOPS0549]	LB-ID V010[pOPS0549], Kan^r/Nm^r, Nit^r	This work
V014 [pOPS0500]	LB-ID V014[pOPS0500], Kan^r/Nm^r, Nit^r	This work
V030 [pOPS0561]	LB-ID V030[pOPS0561], Kan^r/Nm^r, Nit^r	This work
V050 [pOPS0573]	LB-ID V050[pOPS0573], Kan^r/Nm^r, Nit^r	This work
V057 [pOPS0585]	LB-ID V057[pOPS0585], Kan^r/Nm^r, Nit^r	This work
V060 [pOPS0499]	LB-ID V060[pOPS0499], Kan^r/Nm^r, Nit^r	This work
V067 [pOPS0523]	LB-ID V067[pOPS0523], Kan^r/Nm^r, Nit^r	This work
V068 [pOPS0511]	LB-ID V068[pOPS0511], Kan^r/Nm^r, Nit^r	This work
V069 [pOPS0535]	LB-ID V069[pOPS0535], Kan^r/Nm^r, Nit^r	This work

V074 [pOPS0547]	LB-ID V074[pOPS0547], Kan^r/Nm^r, Nit^r	This work
V100 [pOPS0559]	LB-ID V100[pOPS0559], Kan^r/Nm^r, Nit^r	This work
V116 [pOPS0531]	LB-ID V116[pOPS0531], Kan^r/Nm^r, Nit^r	This work
VF2 [pOPS0507]	LB-ID VF2[pOPS0507], Kan^r/Nm^r, Nit^r	This work
VF5 [pOPS0516]	LB-ID VF5[pOPS0516], Kan^r/Nm^r, Nit^r	This work
VSX1 [pOPS0502]	LB-ID VSX1[pOPS0502], Kan^r/Nm^r, Nit^r	This work
VSX10 [pOPS0526]	LB-ID VSX10[pOPS0526], Kan^r/Nm^r, Nit^r	This work
VSX11 [pOPS0538]	LB-ID VSX11[pOPS0538], Kan^r/Nm^r, Nit^r	This work
VSX16 [pOPS0550]	LB-ID VSX16[pOPS0550], Kan^r/Nm^r, Nit^r	This work
VSX28 [pOPS0574]	LB-ID VSX28[pOPS0574], Kan^r/Nm^r, Nit^r	This work
VSX32 [pOPS0586]	LB-ID VSX32[pOPS0586], Kan^r/Nm^r, Nit^r	This work
VSX7 [pOPS0514]	LB-ID VSX7[pOPS0514], Kan^r/Nm^r, Nit^r	This work
WSM1475 [pOPS0493]	LB-ID WSM1475[pOPS0493], Kan^r/Nm^r, Nit^r	This work
WSM1481 [pOPS0494]	LB-ID WSM1481[pOPS0494], Kan^r/Nm^r, Nit^r	This work
WSM1488 [pOPS0496]	LB-ID WSM1488[pOPS0496], Kan^r/Nm^r, Nit^r	This work

Table 2. 3 Plasmids used in this study

Plasmid	Details	Source
EC15071	pL0M-SC-mCherry. Level 0 SC module cloned in pMS, Sp^r	ENSA, supplied by Invitrogen
pIJ11268	Promoterless reporter vector containing *luxCDABE* operon. Tet^r	Geddes et al. (2018)
pIJ11282	pIJ11268 with *luxCDABE* under control of nptII (pNeo) promoter. Tet^r	Geddes et al. (2018)
pLMB640	KpnI/BamHI fragment from pLMB636 (pNifH) was cloned into pIJ11268. Tet^r	Geddes et al. (2018)
pOGG001	pL0M-PU-pNeo, promoter 379 bp upstream of start of *luxC* from pIJ11268 of nptII (pNeo) for constitutive gene expression. Level 0 PU module cloned in pMS. Spc^r	Geddes et al. (2018)
pOGG002	pL0M-SC-luxCDABE operon. Level 0 SC cloned in pMS, Sp^r, Spc^r	This work
pOGG003	pL0M-T-pharma. Level 0 T module cloned in pMS, Sp^r	Geddes et al. (2018)
pOGG004	pLVC-P1-Lv1 (Golden Gate Level 1 cloning site with cloned *lacZ*) position 1 module for vector construction cloned in pMS, Sp^r	Geddes et al. (2018)
pOGG008	pLVC-P2-neo, neomycin-resistance gene, position 2 module for vector construction cloned in pMS, Sp^r, Nm^r/Kan^r	Geddes et al. (2018)

pOGG010	pLVC-P3-RK2, RK2 origin of replication and *oriT* from *E. coli*, position 3 module for vector construction cloned in pMS, Sp^r	Geddes et al. (2018)
pOGG012	pLVC-P4-par, partition genes (*parABCDE* from pMS) for plasmid stability, position 4 module for vector construction cloned in pMS, Sp^r	Geddes et al. (2018)
pOGG014	pLVC-ELT-4, connecting position 4 to position 1 to circularise vector, endlinker module for vector construction cloned in pMS, Sp^r	Geddes et al. (2018)
pOGG026	pL1V-Lv1-neo-RK2-par-ELT4, 6.6 kb, low copy, environmentally-stable, broad-host range Level 1 cloning vector, Nm^r/Kan^r	Geddes et al. (2018)
pOGG037	pL0M-SC-sfGFP, pMS Level 0 SC module cloned in pMS, Sp^r	Geddes et al. (2018)
pOGG043	pL0M-PU-psNifH. Synthetic consensus promoter to drive nodule-specific expression in biovars and strains of *R. leguminosarum*. Level 0 PU module synthesised as a fragment, Sp^r	This work
pOGG050	pL0M-SC-celB. Level 0 SC module cloned in pMS, Sp^r, Spc^r	This work
pOGG054	Destination vector for pL1V-F2, Amp^r	Weber et al. (2011)
pOGG068	Destination vector for pL0V-PU, Sp^r	Weber et al. (2011)
pOGG072	Destination vector for pL0V-SC, Sp^r	Weber et al. (2011)
pOGG082	pL0M-PU-pNifH, promoter 714 bp upstream of the ATG of *nifH* for nodule-specific gene expression in *R. leguminosarum*. Made by PCR using oxp0474 and oxp0475 primers and cloned in destination vector pOGG068, Sp^r, Spc^r	This work
pOGG083	pL0M-SC-gusA, *gusA* from pJP2. Made by PCR using oxp0376 and oxp0377 primers and cloned in destination vector pOGG072, Sp^r, Spc^r	This work
pOGG113	pL0M-PU-NoP, promotorless region from pIJJ11268 for PU module. Made by using oxp1334 and oxp1335 and cloned in destination vector pOGG068, Sp^r	This work
pOPS0252	Reporter plasmid for *R. leguminosarum* constructed with Rlv3841pnifH (pOGG082), *luxCDABE (*pOGG002) and T-pharma (pOGG003) assembled in pOGG026, Nm^r/Kan^r	This work
pOPS0253	Reporter plasmid for *R. leguminosarum* constructed with Rlv3841pNifH (pOGG082), *gusA* (pOGG083) and T-pharma (pOGG003) assembled in pOGG026, Nm^r/Kan^r	This work
pOPS0254	Reporter plasmid for *R. leguminosarum* constructed with Rlv3841pNifH (pOGG082), *celB* (pOGG050) and T-pharma (pOGG003) assembled in pOGG026, Nm^r/Kan^r	This work
pOPS0261	Reporter plasmid for *R. leguminosarum* constructed with psNifH (pOGG043), *luxCDABE* (pOGG002) and T-pharma	This work

	(pOGG003) assembled in pOGG026, Nm^r/Kan^r	
pOPS0262	Reporter plasmid for *R. leguminosarum* constructed with psNifH (pOGG043), *celB* (pOGG050) and T-pharma (pOGG003) assembled in pOGG026, Nm^r/Kan^r	This work
pOPS0263	Reporter plasmid for *R. leguminosarum* constructed with psNifH (pOGG043), *gusA* (pOGG083) and T-pharma (pOGG003) assembled in pOGG026, Nm^r/Kan^r	This work
pOPS0314	Reporter plasmid for *R. leguminosarum* constructed with constitutive promoter pNeo (pOGG001), *celB* (pOGG050) and T-pharma (pOGG003) assembled in pOGG026, Nm^r/Kan^r	This work
pOPS0377	Reporter plasmid for *R. leguminosarum* constructed with constitutive promoter pNeo (pOGG001), sfGFP (pOGG037) and T-pharma (pOGG003) assembled in pOGG026, Nm^r/Kan^r	This work
pOPS0378	Reporter plasmid for *R. leguminosarum* constructed with constitutive promoter pNeo (pOGG001), mCherry (EC15071) and T-pharma (pOGG003) assembled in pOGG026, Nm^r/Kan^r	This work
pOPS0379	Reporter plasmid for *R. leguminosarum* constructed with constitutive promoter Rlv3841pNifH (pOGG082), sfGFP (pOGG037) and T-pharma (pOGG003) assembled in pOGG026, Nm^r/Kan^r	This work
pOPS0380	Reporter plasmid for *R. leguminosarum* constructed with constitutive promoter Rlv3841pnifH (pOGG082), mCherry (EC15071) and T-pharma (pOGG003) assembled in pOGG026, Nm^r/Kan^r	This work
pOPS0381	Reporter plasmid for *R. leguminosarum* constructed with constitutive promoter psNifH (pOGG043), sfGFP (pOGG037) and T-pharma (pOGG003) assembled in pOGG026, Nm^r/Kan^r	This work
pOPS0382	Reporter plasmid for *R. leguminosarum* constructed with constitutive promoter psNifH (pOGG043), mCherry (EC15071) and T-pharma (pOGG003) assembled in pOGG026, Nm^r/Kan^r	This work
pOPS0383	Reporter plasmid for *R. leguminosarum* constructed with NoP (pOGG113), sfGFP (pOGG037) and T-pharma (pOGG003) assembled in pOGG026, Nm^r/Kan^r	This work
pOPS0384	Reporter plasmid for *R. leguminosarum* constructed with NoP (pOGG113), mCherry (EC15071) and T-pharma (pOGG003) assembled in pOGG026, Nm^r/Kan^r	This work
pOPS0385	Reporter plasmid for *R. leguminosarum* constructed with NoP (pOGG113), *gusA* (pOGG083) and T-pharma (pOGG003) assembled in pOGG026, Nm^r/Kan^r	This work

pOPS0386	Reporter plasmid for *R. leguminosarum* constructed with NoP (pOGG113), *celB* (pOGG050) and T-pharma (pOGG003) assembled in pOGG026, Nm^r/Kan^r	This work
pOPS0387	Reporter plasmid for *R. leguminosarum* constructed with constitutive promoter NoP (pOGG113), luxCDABE (pOGG002) and T-pharma (pOGG003) assembled in pOGG026, Nm^r/Kan^r	This work
pOPS0491	psNifH-sfGFP-A1 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: TCCCTTGTCTCC Golay reference name:806rcbc0. Nm^r/Kan^r	This work
pOPS0492	psNifH-sfGFP-A2 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: ACGAGACTGATT Golay reference name:806rcbc1. Nm^r/Kan^r	This work
pOPS0493	psNifH-sfGFP-A3 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: GCTGTACGGATT Golay reference name:806rcbc2. Nm^r/Kan^r	This work
pOPS0494	psNifH-sfGFP-A4 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: ATCACCAGGTGT Golay reference name:806rcbc3. Nm^r/Kan^r	This work
pOPS0495	psNifH-sfGFP-A5 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: TGGTCAACGATA Golay reference name:806rcbc4. Nm^r/Kan^r	This work
pOPS0496	psNifH-sfGFP-A6 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: ATCGCACAGTAA Golay reference name:806rcbc5. Nm^r/Kan^r	This work
pOPS0497	psNifH-sfGFP-A7 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: GTCGTGTAGCCT Golay reference name:806rcbc6. Nm^r/Kan^r	This work
pOPS0498	psNifH-sfGFP-A8 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: AGCGGAGGTTAG Golay reference name:806rcbc7. Nm^r/Kan^r	This work
pOPS0499	psNifH-sfGFP-A9 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: ATCCTTTGGTTC Golay reference name:806rcbc8. Nm^r/Kan^r	This work

pOPS0500	psNifH-sfGFP-A10 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: TACAGCGCATAC Golay reference name:806rcbc9. Nm^r/Kan^r	This work
pOPS0501	psNifH-sfGFP-A11 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: ACCGGTATGTAC Golay reference name:806rcbc10. Nm^r/Kan^r	This work
pOPS0502	psNifH-sfGFP-A12 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: AATTGTGTCGGA Golay reference name:806rcbc11. Plasmid also in *E. coli* ST-ALA, Nm^r/Kan^r	This work
pOPS0503	psNifH-sfGFP-B1 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: TGCATACACTGG Golay reference name:806rcbc12. Nm^r/Kan^r	This work
pOPS0504	psNifH-sfGFP-B2 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: AGTCGAACGAGG Golay reference name:806rcbc13. Nm^r/Kan^r	This work
pOPS0505	psNifH-sfGFP-B3 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: ACCAGTGACTCA Golay reference name:806rcbc14. Nm^r/Kan^r	This work
pOPS0506	psNifH-sfGFP-B4 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: GAATACCAAGTC Golay reference name:806rcbc15. Nm^r/Kan^r	This work
pOPS0507	psNifH-sfGFP-B5 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: GTAGATCGTGTA Golay reference name:806rcbc16. Nm^r/Kan^r	This work
pOPS0508	psNifH-sfGFP-B6 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: TAACGTGTGTGC Golay reference name:806rcbc17. Nm^r/Kan^r	This work
pOPS0509	psNifH-sfGFP-B7 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: CATTATGGCGTG Golay reference name:806rcbc18. Nm^r/Kan^r	This work
pOPS0510	psNifH-sfGFP-B8 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: CCAATACGCCTG Golay reference name:806rcbc19. Nm^r/Kan^r	This work

pOPS0511	psNifH-sfGFP-B9 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: GATCTGCGATCC Golay reference name:806rcbc20. Nm^r/Kan^r	This work
pOPS0512	psNifH-sfGFP-B10 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: CAGCTCATCAGC Golay reference name:806rcbc21. Nm^r/Kan^r	This work
pOPS0513	psNifH-sfGFP-B11 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: CAAACAACAGCT Golay reference name:806rcbc22. Nm^r/Kan^r	This work
pOPS0514	psNifH-sfGFP-B12 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: GCAACACCATCC Golay reference name:806rcbc23. Nm^r/Kan^r	This work
pOPS0515	psNifH-sfGFP-C1 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: GCGATATATCGC Golay reference name:806rcbc24. Nm^r/Kan^r	This work
pOPS0516	psNifH-sfGFP-C2 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: CGAGCAATCCTA Golay reference name:806rcbc25. Nm^r/Kan^r	This work
pOPS0517	psNifH-sfGFP-C3 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: AGTCGTGCACAT Golay reference name:806rcbc26. Nm^r/Kan^r	This work
pOPS0518	psNifH-sfGFP-C4 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: GTATCTGCGCGT Golay reference name:806rcbc27. Nm^r/Kan^r	This work
pOPS0519	psNifH-sfGFP-C5 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: CGAGGGAAAGTC Golay reference name:806rcbc28. Nm^r/Kan^r	This work
pOPS0520	psNifH-sfGFP-C6 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: CAAATTCGGGAT Golay reference name:806rcbc29. Nm^r/Kan^r	This work
pOPS0521	psNifH-sfGFP-C7 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: AGATTGACCAAC Golay reference name:806rcbc30. Nm^r/Kan^r	This work

pOPS0522	psNifH-sfGFP-C8 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: AGTTACGAGCTA Golay reference name:806rcbc31. Nm^r/Kan^r	This work
pOPS0523	psNifH-sfGFP-C9 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: GCATATGCACTG Golay reference name:806rcbc32. Nm^r/Kan^r	This work
pOPS0524	psNifH-sfGFP-C10 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: CAACTCCCGTGA Golay reference name:806rcbc33. Nm^r/Kan^r	This work
pOPS0525	psNifH-sfGFP-C11 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: TTGCGTTAGCAG Golay reference name:806rcbc34. Nm^r/Kan^r	This work
pOPS0526	psNifH-sfGFP-C12 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: TACGAGCCCTAA Golay reference name:806rcbc35. Nm^r/Kan^r	This work
pOPS0527	psNifH-sfGFP-D1 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: CACTACGCTAGA Golay reference name:806rcbc36, Nm^r/Kan^r	This work
pOPS0528	psNifH-sfGFP-D2 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: TGCAGTCCTCGA Golay reference name:806rcbc37, Nm^r/Kan^r	This work
pOPS0529	psNifH-sfGFP-D3 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: ACCATAGCTCCG Golay reference name:806rcbc38. Nm^r/Kan^r	This work
pOPS0530	psNifH-sfGFP-D4 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: TCGACATCTCTT Golay reference name:806rcbc39. Nm^r/Kan^r	This work
pOPS0531	psNifH-sfGFP-D5 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: GAACACTTTGGA Golay reference name:806rcbc40. Nm^r/Kan^r	This work
pOPS0532	psNifH-sfGFP-D6 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: GAGCCATCTGTA Golay reference name:806rcbc41. Nm^r/Kan^r	This work

pOPS0533	psNifH-sfGFP-D7 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: TTGGGTACACGT Golay reference name:806rcbc42. Nm^r/Kan^r	This work
pOPS0534	psNifH-sfGFP-D8 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: AAGGCGCTCCTT Golay reference name:806rcbc43. Nm^r/Kan^r	This work
pOPS0535	psNifH-sfGFP-D9 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: TAATACGGATCG Golay reference name:806rcbc44. Nm^r/Kan^r	This work
pOPS0536	psNifH-sfGFP-D10 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: TCGGAATTAGAC Golay reference name:806rcbc45. Nm^r/Kan^r	This work
pOPS0537	psNifH-sfGFP-D11 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: TGTGAATTCGGA Golay reference name:806rcbc46. Nm^r/Kan^r	This work
pOPS0538	psNifH-sfGFP-D12 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: CATTCGTGGCGT Golay reference name:806rcbc47. Nm^r/Kan^r	This work
pOPS0539	psNifH-sfGFP-E1 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: TACTACGTGGCC Golay reference name:806rcbc48. Nm^r/Kan^r	This work
pOPS0540	psNifH-sfGFP-E2 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: GGCCAGTTCCTA Golay reference name:806rcbc49. Nm^r/Kan^r	This work
pOPS0541	psNifH-sfGFP-E3 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: GATGTTCGCTAG Golay reference name:806rcbc50. Nm^r/Kan^r	This work
pOPS0542	psNifH-sfGFP-E4 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: CTATCTCCTGTC Golay reference name:806rcbc51. Nm^r/Kan^r	This work
pOPS0543	psNifH-sfGFP-E5 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: ACTCACAGGAAT Golay reference name:806rcbc52. Nm^r/Kan^r	This work

pOPS0544	psNifH-sfGFP-E6 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: ATGATGAGCCTC Golay reference name:806rcbc53. Nm^r/Kan^r	This work
pOPS0545	psNifH-sfGFP-E7 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: GTCGACAGAGGA Golay reference name:806rcbc54. Nm^r/Kan^r	This work
pOPS0546	psNifH-sfGFP-E8 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: TGTCGCAAATAG Golay reference name:806rcbc55. Nm^r/Kan^r	This work
pOPS0547	psNifH-sfGFP-E9 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: CATCCCTCTACT Golay reference name:806rcbc56. Nm^r/Kan^r	This work
pOPS0548	psNifH-sfGFP-E10 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: TATACCGCTGCG Golay reference name:806rcbc57. Nm^r/Kan^r	This work
pOPS0549	psNifH-sfGFP-E11 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: AGTTGAGGCATT Golay reference name:806rcbc58. Nm^r/Kan^r	This work
pOPS0550	psNifH-sfGFP-E12 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: ACAATAGACACC Golay reference name:806rcbc59. Nm^r/Kan^r	This work
pOPS0551	psNifH-sfGFP-F1 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: CGGTCAATTGAC Golay reference name:806rcbc60. Nm^r/Kan^r	This work
pOPS0552	psNifH-sfGFP-F2 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: TGGAGTAGGTGG Golay reference name:806rcbc61. Nm^r/Kan^r	This work
pOPS0553	psNifH-sfGFP-F3 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: GCTCGAAGATTC Golay reference name:806rcbc62. Nm^r/Kan^r	This work
pOPS0554	psNifH-sfGFP-F4 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: AGGCTTACGTGT Golay reference name:806rcbc63. Nm^r/Kan^r	This work

pOPS0555	psNifH-sfGFP-F5 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: TCTCTACCACTC Golay reference name:806rcbc64. Nm^r/Kan^r	This work
pOPS0556	psNifH-sfGFP-F6 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: ACTTCCAACTTC Golay reference name:806rcbc65. Nm^r/Kan^r	This work
pOPS0557	psNifH-sfGFP-F7 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: CTCACCTAGGAA Golay reference name:806rcbc66. Nm^r/Kan^r	This work
pOPS0558	psNifH-sfGFP-F8 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: GTGTTGTCGTGC Golay reference name:806rcbc67. Nm^r/Kan^r	This work
pOPS0559	psNifH-sfGFP-F9 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: CCACAGATCGAT Golay reference name:806rcbc68. Nm^r/Kan^r	This work
pOPS0560	psNifH-sfGFP-F10 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: TATCGACACAAG Golay reference name:806rcbc69. Nm^r/Kan^r	This work
pOPS0561	psNifH-sfGFP-F11 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: GATTCCGGCTCA Golay reference name:806rcbc70. Nm^r/Kan^r	This work
pOPS0562	psNifH-sfGFP-F12 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: CGTAATTGCCGC Golay reference name:806rcbc71. Nm^r/Kan^r	This work
pOPS0563	psNifH-sfGFP-G1 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: GGTGACTAGTTC Golay reference name:806rcbc72. Nm^r/Kan^r	This work
pOPS0564	psNifH-sfGFP-G2 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: GTGGAGTCTCAT Golay reference name:806rcbc73. Nm^r/Kan^r	This work
pOPS0565	psNifH-sfGFP-G3 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: TAGGCATGCTTG Golay reference name:806rcbc74. Nm^r/Kan^r	This work

pOPS0566	psNifH-sfGFP-G4 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: AACTAGTTCAGG Golay reference name:806rcbc75. Nm^r/Kan^r	This work
pOPS0567	psNifH-sfGFP-G5 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: ATTCTGCCGAAG Golay reference name:806rcbc76. Nm^r/Kan^r	This work
pOPS0568	psNifH-sfGFP-G6 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: AGCATGTCCCGT Golay reference name:806rcbc77. Nm^r/Kan^r	This work
pOPS0569	psNifH-sfGFP-G7 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: GTACGATATGAC Golay reference name:806rcbc78. Nm^r/Kan^r	This work
pOPS0570	psNifH-sfGFP-G8 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: GTGGTGGTTTCC Golay reference name:806rcbc79. Nm^r/Kan^r	This work
pOPS0571	psNifH-sfGFP-G9 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: TAGTATGCGCAA Golay reference name:806rcbc80. Nm^r/Kan^r	This work
pOPS0572	psNifH-sfGFP-G10 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: TGCGCTGAATGT Golay reference name:806rcbc81. Nm^r/Kan^r	This work
pOPS0573	psNifH-sfGFP-G11 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: ATGGCTGTCAGT Golay reference name:806rcbc82. Nm^r/Kan^r	This work
pOPS0574	psNifH-sfGFP-G12 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: GTTCTCTTCTCG Golay reference name:806rcbc83. Nm^r/Kan^r	This work
pOPS0575	psNifH-sfGFP-H1 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: CGTAAGATGCCT Golay reference name:806rcbc84. Nm^r/Kan^r	This work
pOPS0576	psNifH-sfGFP-H2 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: GCGTTCTAGCTG Golay reference name:806rcbc85. Nm^r/Kan^r	This work

pOPS0577	psNifH-sfGFP-H3 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: GTTGTTCTGGGA Golay reference name:806rcbc86. Nm^r/Kan^r	This work
pOPS0578	psNifH-sfGFP-H4 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: GGACTTCCAGCT Golay reference name:806rcbc87. Nm^r/Kan^r	This work
pOPS0579	psNifH-sfGFP-H5 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: CTCACAACCGTG Golay reference name:806rcbc88. Nm^r/Kan^r	This work
pOPS0580	psNifH-sfGFP-H6 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: CTGCTATTCCTC Golay reference name:806rcbc89. Nm^r/Kan^r	This work
pOPS0581	psNifH-sfGFP-H7 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: ATGTCACCGCTG Golay reference name:806rcbc90. Nm^r/Kan^r	This work
pOPS0582	psNifH-sfGFP-H8 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: TGTAACGCCGAT Golay reference name:806rcbc91. Nm^r/Kan^r	This work
pOPS0584	psNifH-sfGFP-H10 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: ATGGGTTCCGTC Golay reference name:806rcbc93. Nm^r/Kan^r	This work
pOPS0585	psNifH-sfGFP-H11 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: TTGGCTCTATTC Golay reference name:806rcbc94. Nm^r/Kan^r	This work
pOPS0586	psNifH-sfGFP-H12 in pOGG026 Plasmid constructed by GG with universal primer binding site. Confirmed BC sequence: GATCCCACGTAC Golay reference name:806rcbc95. Nm^r/Kan^r	This work
pOPS0750	Reporter Plasmid constructed with constitutive promoter pNeo (pOGG001), *gusA* (pOGG083) and T-pharma (pOGG003) assembled in pOGG026, Nm^r/Kan^r	This work

Amp^r, ampicillin resistance; Chl^r, Chloramphenicol resistance; Gm^r, gentamicin resistance; Kan^r, kanamycin resistance; Nm^r, neomycin resistance; Rf^r, rifampicin resistance, Sp^r, spectinomycin resistance; Str^r streptomycin resistance; Tet^r, tetracycline resistance.

All strains were preserved at -80°C, frozen in media and supplemented with 20% glycerol.

Primers

Table 2. 4 Primers used in this study

Primer	Sequence (5'-3')	Description	Source
oxp0282	CCTACCTAGATCCTCGC	Forward primer. Used for screening plasmid constructions in pOGG026 backbone vector	This work
oxp0283	AGCGTTCTGAACAAATCC	Reverse primer. Used for screening plasmid constructions in pOGG026 backbone vector	This work
oxp1334	CACTCTGTGGTCTCAGGAG CATCCTCCCTAATGCGCC	Forward primer for amplification of promotorless region from pIJJ11268 for PU module	This work
oxp1335	CACTTCGTGGTCTCACATTC GTATCCTCCAAGCCTGAA	Reverse primer for amplification of promotorless region from pIJJ11268 for PU module	This work
oxp0474	CACTCTGTGGTCTCAGGAG TCGATGCTGACCGCCT	Forward primer for amplification of pNifH region from Rlv3841 plasmid pRL10 for PU module	Geddes et al. (2018)
oxp0475	CACTTCGTGGTCTCACATTT TTGGCGTTCCTTCATGTGT	Reverse primer for amplification of pNifH region from Rlv3841 plasmid pRL10 for PU module	Geddes et al. (2018)
oxp0376	CACTCTGTGGTCTCAAATG GTCCGTCCTGTAG	Forward primer for amplification of *gusA* gene for SC module	Geddes et al. (2018)
oxp0377	CACTTCGTGGTCTCAAAGC TCATTGTTTGCCTCCC	Reverse primer for amplification of *gusA* gene for SC module	Geddes et al. (2018)
1_IT	GCCCAGTCTACTCGAGGGG CTACATCACGCATGGTATG GA	First-step PCR forward primer supplemented with specific barcodes used in IT sequencing	This work
2_IT	GCCCAGTCTACTCGAGGGT GTGTCATCACGCATGGTAT GGA	First-step PCR forward primer supplemented with specific barcodes used in IT sequencing	This work
3_IT	GCCCAGTCTACTCGAGGGA GTCTGCATCACGCATGGTA TGGA	First-step PCR forward primer supplemented with specific barcodes used in IT sequencing	This work

4_IT	GCCCAGTCTACTCGAGGGATCACATCACGCATGGTATGGA	First-step PCR forward primer supplemented with specific barcodes used in IT sequencing	This work
5_IT	GCCCAGTCTACTCGAGGGGACGACATCACGCATGGTATGGA	First-step PCR forward primer supplemented with specific barcodes used in IT sequencing	This work
6_IT	GCCCAGTCTACTCGAGGGTCGTCGCATCACGCATGGTATGGA	First-step PCR forward primer supplemented with specific barcodes used in IT sequencing	This work
7_IT	GCCCAGTCTACTCGAGGGTGCTCATCACGCATGGTATGGA	First-step PCR forward primer supplemented with specific barcodes used in IT sequencing	This work
8_IT	GCCCAGTCTACTCGAGGGCAGTTCATCACGCATGGTATGGA	First-step PCR forward primer supplemented with specific barcodes used in IT sequencing	This work
9_IT	GCCCAGTCTACTCGAGGGACATGTCATCACGCATGGTATGGA	First-step PCR forward primer supplemented with specific barcodes used in IT sequencing	This work
10_IT	GCCCAGTCTACTCGAGGGGCGGCATCACGCATGGTATGGA	First-step PCR forward primer supplemented with specific barcodes used in IT sequencing	This work
11_IT	GCCCAGTCTACTCGAGGGGTTGACATCACGCATGGTATGGA	First-step PCR forward primer supplemented with specific barcodes used in IT sequencing	This work
12_IT	GCCCAGTCTACTCGAGGGGTGGCTCATCACGCATGGTATGGA	First-step PCR forward primer supplemented with specific barcodes used in IT sequencing	This work
A_IT	ATCTCGGTGGTCGCCGTAGCTCCCCAGTCACGACGTTGTAAAACG	First-step PCR reverse primer supplemented with specific barcodes used in IT sequencing	This work
B_IT	ATCTCGGTGGTCGCCGTACTAGTCCCAGTCACGACGTTGTAAAACG	First-step PCR reverse primer supplemented with specific barcodes used in IT sequencing	This work
C_IT	ATCTCGGTGGTCGCCGTATAGATCCCCAGTCACGACGTTGTAAAACG	First-step PCR reverse primer supplemented with specific barcodes used in IT sequencing	This work
D_IT	ATCTCGGTGGTCGCCGTATCGCCCCAGTCACGACGTTGTAAAACG	First-step PCR reverse primer supplemented with specific barcodes used in IT sequencing	This work
E_IT	ATCTCGGTGGTCGCCGTACCTTACCCAGTCACGACGTTGTAAAACG	First-step PCR reverse primer supplemented with specific barcodes used in IT sequencing	This work
F_IT	ATCTCGGTGGTCGCCGTACATAACCCCAGTCACGACGTTGTAAAACG	First-step PCR reverse primer supplemented with specific barcodes used in IT sequencing	This work

G_IT	ATCTCGGTGGTCGCCGTAC AGACCCAGTCACGACGTTG TAAAACG	First-step PCR reverse primer supplemented with specific barcodes used in IT sequencing	This work
H_IT	ATCTCGGTGGTCGCCGTA TGTTCCCCAGTCACGACGT TGTAAAACG	First-step PCR reverse primer supplemented with specific barcodes used in IT sequencing	This work
IT_A_FP_1	CCATCTCATCCCTGCGTGTC TCCGACTCAGCTAAGGTAA CGATATAAAACCGCCCAGT CTACTCGAGGG	Second-step PCR Forward primer with Ion Xpress™ Barcodes for library identification in IT sequencing	Wheatley et al. (2017)
IT_A_FP_2	CCATCTCATCCCTGCGTGTC TCCGACTCAGTAAGGAGAA CGATATAAAACCGCCCAGT CTACTCGAGGG	Second-step PCR Forward primer with Ion Xpress™ Barcodes for library identification in IT sequencing	Wheatley et al. (2017)
IT_A_FP_3	CCATCTCATCCCTGCGTGTC TCCGACTCAGAAGAGGATT CGATATAAAACCGCCCAGT CTACTCGAGGG	Second-step PCR Forward primer with Ion Xpress™ Barcodes for library identification in IT sequencing	Wheatley et al. (2017)
IT_A_FP_4	CCATCTCATCCCTGCGTGTC TCCGACTCAGTACCAAGAT CGATATAAAACCGCCCAGT CTACTCGAGGG	Second-step PCR Forward primer with Ion Xpress™ Barcodes for library identification in IT sequencing	Wheatley et al. (2017)
IT_A_FP_5	CCATCTCATCCCTGCGTGTC TCCGACTCAGCAGAAGGAA CGATATAAAACCGCCCAGT CTACTCGAGGG	Second-step PCR Forward primer with Ion Xpress™ Barcodes for library identification in IT sequencing	Wheatley et al. (2017)
IT_A_FP_6	CCATCTCATCCCTGCGTGTC TCCGACTCAGCTGCAAGTT CGATATAAAACCGCCCAGT CTACTCGAGGG	Second-step PCR Forward primer with Ion Xpress™ Barcodes for library identification in IT sequencing	Wheatley et al. (2017)
IT_A_FP_7	CCATCTCATCCCTGCGTGTC TCCGACTCAGTTCGTGATTC GATATAAAACCGCCCAGTC TACTCGAGGG	Second-step PCR Forward primer with Ion Xpress™ Barcodes for library identification in IT sequencing	Wheatley et al. (2017)
IT_A_FP_8	CCATCTCATCCCTGCGTGTC TCCGACTCAGTTCCGATAA CGATATAAAACCGCCCAGT CTACTCGAGGG	Second-step PCR Forward primer with Ion Xpress™ Barcodes for library identification in IT sequencing	Wheatley et al. (2017)
IT_A_FP_9	CCATCTCATCCCTGCGTGTC TCCGACTCAGTGAGCGGAA CGATATAAAACCGCCCAGT CTACTCGAGGG	Second-step PCR Forward primer with Ion Xpress™ Barcodes for library identification in IT sequencing	Wheatley et al. (2017)
IT_A_FP_10	CCATCTCATCCCTGCGTGTC TCCGACTCAGCTGACCGAA CGATATAAAACCGCCCAGT	Second-step PCR Forward primer with Ion Xpress™ Barcodes for library	Wheatley et al. (2017)

	CTACTCGAGGG	identification in IT sequencing	
IT_trP1_FP_ Synthetic	CCTCTCTATGGGCAGTCGG TGATATCTCGGTGGTCGCC GTA	Second-step PCR reverse primer with truncated version of the Illumina Nextera adapter to introduce trP1 adapter	This work
1_MiSeq	GCTAGTTGAGGATGGCAGC GTG	Forward primer supplemented with specific barcodes used in MiSeq sequencing	This work
2_MiSeq	TGTGT GTTGAGGATGGCAGCGTG	Forward primer supplemented with specific barcodes used in MiSeq sequencing	This work
3_MiSeq	AGTCTGGTTGAGGATGGCA GCGTG	Forward primer supplemented with specific barcodes used in MiSeq sequencing	This work
4_MiSeq	ATCAGTTGAGGATGGCAGC GTG	Forward primer supplemented with specific barcodes used in MiSeq sequencing	This work
5_MiSeq	GACGAGTTGAGGATGGCA GCGTG	Forward primer supplemented with specific barcodes used in MiSeq sequencing	This work
6_MiSeq	TCGTCG GTTGAGGATGGCAGCGTG	Forward primer supplemented with specific barcodes used in MiSeq sequencing	This work
7_MiSeq	TGCTGTTGAGGATGGCAGC GTG	Forward primer supplemented with specific barcodes used in MiSeq sequencing	This work
8_MiSeq	CAGTT GTTGAGGATGGCAGCGTG	Forward primer supplemented with specific barcodes used in MiSeq sequencing	This work
9_MiSeq	ACATGTGTTGAGGATGGCA GCGTG	Forward primer supplemented with specific barcodes used in MiSeq sequencing	This work
10_MiSeq	GCGGGTTGAGGATGGCAG CGTG	Forward primer supplemented with specific barcodes used in MiSeq sequencing	This work
11_MiSeq	GTTGAGTTGAGGATGGCA GCGTG	Forward primer supplemented with specific barcodes used in MiSeq sequencing	This work
12_MiSeq	GTGGCTGTTGAGGATGGC AGCGTG	Forward primer supplemented with specific barcodes used in MiSeq sequencing	This work
A_MiSeq	GCTCCCCAGTCACGACGTT GTAAAACG	Reverse primer supplemented with specific barcodes used in MiSeq sequencing	This work

B_MiSeq	CTAGTCCCAGTCACGACGT TGTAAAACG	Reverse primer supplemented with specific barcodes used in MiSeq sequencing	This work
C_MiSeq	TAGATC CCCAGTCACGACGTTGTAA AACG	Reverse primer supplemented with specific barcodes used in MiSeq sequencing	This work
D_MiSeq	TCGCCCCAGTCACGACGTT GTAAAACG	Reverse primer supplemented with specific barcodes used in MiSeq sequencing	This work
E_MiSeq	CCTTACCCAGTCACGACGTT GTAAAACG	Reverse primer supplemented with specific barcodes used in MiSeq sequencing	This work
F_MiSeq	CATAACCCCAGTCACGACG TTGTAAAACG	Reverse primer supplemented with specific barcodes used in MiSeq sequencing	This work
G_MiSeq	CAGACCCAGTCACGACGTT GTAAAACG	Reverse primer supplemented with specific barcodes used in MiSeq sequencing	This work
H_MiSeq	TGTTCCCCAGTCACGACGT TGTAAAACG	Reverse primer supplemented with specific barcodes used in MiSeq sequencing	This work

Molecular techniques

DNA isolation

Plasmid DNA was extracted from *E. coli* using GeneJET Plasmid Minirep Kit (Thermo Scientific) according to the manufacturer's instructions.

Genomic DNA was isolated rhizobial cultures using DNeasy Blood and Tissue Kit (Qiagen) according to the manufacturer's instructions.

Extraction of DNA from nodules samples with Alkaline PEG 200 modified from Chomczynski and Rymaszewski (2006). Nodules were crushed in Eppendorf tubes and alkaline PEG reagent was added at 10 times higher the sample volume. Samples were incubated for 15 minutes at 60°C and vortexed. Plant tissue was precipitated by a slow

spin of 1000 rpm for 15 minutes. The supernatant was transfer to sterile PCR tubes. From 1 to 5 µL product can be used directly in a 20 - 50 µL PCR reaction. Samples were diluted with MilliQ H_2O 1/10 when the DNA concentration was very high.

Primers design

PCR primers were designed in Geneious 10.0.9, with no mismatches within 5bp of 3'-end. Oligonucleotides were synthesised by Eurofins MWG Operon.

DNA amplification by PCR

Mapping/screening PCR reactions were carried out in 10 µl reactions using GoTaq® Green Master Mix (Promega). Amplification PCR reactions were carried out using in 25 µl or 50 µl reactions using Phusion® High-fidelity PCR Master Mix (Thermo Scientific) using GC buffer with DMSO included in the reaction. Thermocycler conditions were set according to the manufacturer's instructions for polymerases, with T_m determined by Geneious 10.0.9. Colony PCR was used as an alternative method for quick screen for plasmid inserts directly from *E. coli or Rhizobium* colonies in 50 µl total volume, adding a small amount of a colony from a plate. Where large-scale screening of rhizobial colonies was required colonies were transferred into 10 µl sterile H_2O and heated at 60°C for five minutes. The resulting suspension (1 µl) was used in the PCR reaction.

DNA Gel Electrophoresis

Electrophoresis was carried out in 0.9% agarose (Sigma Aldrich) in TAE buffer (400 mM Tris acetate, 1 mM EDTA). DNA was stained adding Sybr® Safe (1:10000) (Invitrogen)

and GeneRuler 1 kb (Thermo Scientific) was used as a marker ladder. Gels were run at 120 mV for 40 minutes. For fragments smaller than 500 bp, electrophoresis was carried out in 1.7% agarose and Sybr® Safe concentration of 1:5000 was used. GeneRuler 100 bp (Thermo Scientific) was used as a marker ladder and gels were run at 100 mV for 60 minutes. Gels were visualized using the GelDoc EZ System (BioRad). DNA concentration in each band was calculated using Image Lab™ Software for Mac Version 6.0.1.

Restriction digest and DNA ligation

Restriction enzyme digestions were carried out by standard protocols (Roche, Thermo Scientific or NEB). When necessary, reactions were stopped by an incubation of 10 min at 65°C. DNA ligations were carried out using T4 DNA Ligase according to the manufacturer's instructions (Thermo Scientific). Ligations were carried out at room temperature overnight or at 37°C when necessary.

Golden Gate cloning

The protocol was modified from that described by Engler et al., (2008). Most parts were constructed by DNA synthesis by Invitrogen and supplied in the GeneArt cloning vector pMS (Invitrogen, Thermo Fisher Scientific Inc.). As an alternative, some vector modules were constructed by PCR amplification from commonly used vectors. For Level 1 modules synthesized by Invitrogen were domesticated prior to DNA synthesis by removal of BsaI, BpiI, DraIII and Esp3I (Golden Gate cloning restriction enzyme sites) by altering a single nucleotide in the recognition site. In an open reading frame (ORF) a

neutral change was made in the wobble base of a codon, such that the same amino acid was encoded. Outside of an ORF, a transition mutation was introduced. The steps followed are summarised in Figure 2. 1.

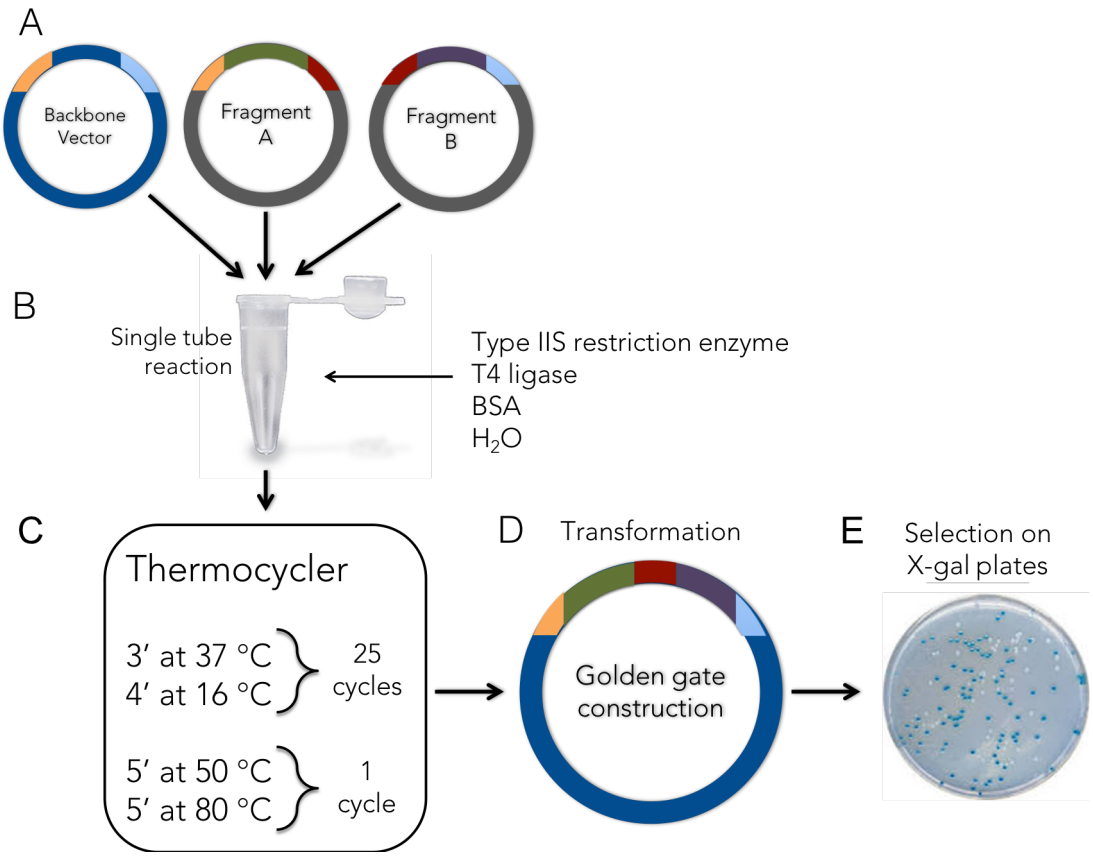


Figure 2. 1 Golden Gate cloning steps. A) Backbone vector and fragments of interest were cut with Type IIS restriction enzymes (Type-IIS). B) All the components (equal to 40 fmol) were placed in a one-tube one-step reaction with 1 µl of corresponding Type-IIS enzyme, 1.5 µl of T4 ligase buffer, 1 µl of T4 DNA ligase, 1.5 µl BSA (2 mg/ml) and H_2O to a total volume of 15 µl. C) The Golden Gate reaction mix was placed in a PCR Thermal Cycler with the following parameters: 25 cycles of digestion (3 minutes at 37°C), and ligation (4 minutes at 16°C) and one cycle with a final digestion (5 minutes at 50°C) and inactivation (5 minutes at 80°C). D) Golden Gate reaction product was transformed into *E. coli* strains to obtain the final plasmid constructions. E) A great advantage of this technique is that, in addition to the antibiotic selection chosen during vector construction, a second selection (by colour) can be employed, making it possible to identify the colonies with the fragments of interest by spreading colonies on X-gal plates and correct antibiotics. The change of colour goes from blue to white for level 1, and from orange to blue or white for level 2. Colonies thought to be correct were verified by restriction enzyme digest.

Standard transformation.

Chemically competent *E. coli* DH5α or *E. coli* ST18, a *hemA* deletion mutant (Thoma & Schobert, 2009) cells were used for transformations. Commercial Alpha-Select Gold Efficiency competent cells (Bioline) were used when high efficiency was needed.

An aliquot of *E. coli* (approx. 50 µl) of prepared competent cells was thawed on ice and 5 µl of plasmid or ligation reaction added. Cells were incubated on ice for 30 minutes before being heat- shocked at 42°C for 45 seconds. Cells were incubated on ice for a further 2 minutes. 400 µl of SOC media (Melford) was added and for *E. coli* ST18 strains (ALA) was also added. Cells were shaken at 37°C for 1 hour at 200 rpm. Cells were spread-plated on LB agar containing appropriate antibiotics and plates incubated overnight at 37°C.

High-throughput transformation (custom protocol)

Chemically competent *E. coli* DH5α or *E. coli* ST18, a *hemA* deletion mutant cells were used for transformations. An aliquot of *E. coli* was thawed on ice and transferred to a 96-well PCR plate by pipetting 30 µl into each well. 4 µl of plasmid or ligation reaction to be transformed was added to the well. One well was left empty as a control. The plate was covered with a sterile hydrophobic porous sealing film to reduce contamination. Cells were incubated on ice for 30 minutes before being heat-shocked at 42 °C for 45 seconds. 100 µl of SOC media (Melford) was added to all wells, and for *E. coli* ST18 strains 5-aminolevulinic acid (ALA) was also added. The plate was placed on a shaking incubator, 150-200 rpm at 37 °C for 1.5 hour. A 96-deep-well plate with

1.5 ml of LB, relevant antibiotics and ALA (as required), was inoculated with 50 µl of each transformation mixture, including a negative control. The plate was covered with a sterile hydrophobic porous sealing film and incubated in a shaking incubator, 150-200 rpm at 37 °C over-night. The OD_{600} of a 200 µl sample from each well was measured to confirm bacterial growth using a FLUOStar Omega Microplate Reader (BMG). Colony PCR was run to confirm the presence of the plasmid of interest. These strains were stocked at -80 °C in microtiter plates by mixing 50 µl of sterile glycerol (50% v/v) with 150 µl of each strain.

Standard conjugation from *E. coli* into Rhizobial strains

This protocol is modified from Figurski and Helinski (1979). The strains of rhizobia to be conjugated were grown on a slope for 3 days before conjugation. On the day of conjugation, 10 ml of *E. coli* strains were grown at 200 rpm for 5 hours. The rhizobial slope was washed with 3-5 ml of fresh media. All cultures were placed in conical tubes and centrifuged at 4000 rpm for 10 minutes, the pellets were re-suspended in fresh media and this process was carried out three times to remove any traces of antibiotics. Finally, the pellets were re-suspended in 3-5 ml of fresh media. A mix of 400 µl of donor, 400 µl of recipient, and 200 µl of the helper plasmid pRK2013 was centrifuged for 5 min at 7000 rpm. The supernatant was eliminated. The pellet was mixed, placed onto a nitrocellulose filter on a TY plate and incubated overnight at 27 °C. On the following day, bacteria growing on the filter were streaked onto TY plates with corresponding antibiotics.

High-throughput conjugation from *E. coli* into Rhizobial strains in 96-well microtiter plates (custom protocol)

Rhizobial strains to be conjugated (recipients) were inoculated into a 96-deep-well plate with 1.2 ml of media and relevant antibiotics. The plate was covered with a sterile hydrophobic porous sealing film to reduce contamination. Strains were grown in a shaking incubator, 150 rpm at 28 °C for 3 days. This plate was used for re-inoculation of 5 plates to allow concentration of a high number of cells.

The day before the conjugation, *E. coli* ST18, a *hemA* deletion mutant (Thoma & Schobert, 2009) (donor strain) was inoculated into 96-deep-well plate with 1.2 ml of media, relevant antibiotics and ALA, before sealing with sterile hydrophobic porous sealing film. Strains were grown in a shaking incubator, 150 rpm at 37 °C.

On the day of the conjugation experiment, all plates the were centrifuged at 4000 rpm for 15 minutes. The 5 rhizobial pellets were combined into a single plate to concentrate the bacterial cells. The final plate with rhizobial strains and that with *E. coli* ST18 were washed with fresh media three times to remove any traces of antibiotics. The final pellets were resuspended in 1.2 ml of TY media. In a PCR plate, 100 µl donor was mixed with 100 µl recipient and spun down at 4000 rpm for 15 minutes. The supernatant was discarded, and pellet was re-suspended in 30 µl of TY-ALA media. The re-suspended mix was transferred to a microtiter plate with 150 µl of TY-ALA agar and sealed with sterile hydrophobic porous sealing film. The plate was incubated for 48 hours at 28 °C. 100 µl TY was added to the bacteria microfilm growing on top of the agar and transferred to a 96-deep-well plate with 1.5 ml of TY media and

relevant antibiotics. To have better aeriation resulting in an increase of strain growth, we prepared each 96-deep-well plate with TY agar slopes in each of the well. The 96-deep-well plate was incubated in a shaking incubator, 150 rpm at 28 °C for 3 days. Serial dilutions were done and 100µl of each conjugation was plated onto TY agar (with relevant antibiotics) to get single colonies. Single colonies were re-grown in a 96-deep-well plate with TY liquid media with relevant antibiotics in a shaking incubator, 150 rpm at 28 °C for 3 days. The OD_{600} of a 200 µl sample from each well was measured to confirm bacterial growth (where OD_{600} 1.0 = $1x10^9$ cfu) using a FLUOStar Omega Microplate Reader (BMG). To eliminate possible spontaneous resistance to relevant antibiotics, positive plasmid acquisition was verified by colony PCR. Microtiter plates for -80 °C stocks were done by mixing 50µl of sterile glycerol 50% v/v + 150µl of each strain.

Plant Experiments

General plant growth

P. sativum cv. Aveola (Thompson and Morgan, UK), common bean (Tendergreen) (Thompson and Morgan, UK) white clover and soybean seeds were surface sterilised by immersing in 95% ethanol for 30 seconds, followed by washing with sterile water. Seeds were then immersed in 2% sodium hypochlorite for 5 minutes. After washing the seeds five times with sterile water, they were placed onto 1% DWA plates. The plates were placed in the dark at room temperature for 4-5 days. Alfalfa seeds were sterilized with 75% ethanol for 45 minutes with continuous mixing, followed by washing with sterile water, and 2% sodium hypochlorite for 15 minutes with

continuous mixing. After washing the seeds several times with sterile water, these were placed onto DWA plates and placed in the dark at room temperature for 2 days. Seedlings were placed in sterile 1 Lt pots with medium vermiculite and supplied with 400 ml nitrogen-free rooting solution (Poole *et al.*, 1994). Common bean and soybean were grown in fine vermiculite and supplied with 400ml of bean nitrogen-free rooting solution (Broughton & Dilworth, 1971). At least five biological replicates were grown, with the exception of soybean plants, where only three replicates were grown. The inoculation was approximately $1x10^7$ cfu of rhizobial strain per pot. Plants without inoculation were grown as a negative control. (water control - WC). Plants were grown (random positioning of pots) in a controlled growth chamber at 21°C, 16-h/8-h day/night cycle. Alfalfa and white clover plants were harvested 21 dpi. Pea plants were harvested 21 or 28 dpi. Common bean and soybean plants were harvested 42 dpi.

Nodulation test

Following the methodology described by Howieson and Dilworth (2016). Boiling tubes were filled with sterile fine vermiculite and 20 ml nitrogen-free rooting solution (Poole *et al.*, 1994). Surface-sterilised pea seeds were placed in boiling tubes and inoculated with a rhizobial strain. Plants were harvested 21 dpi and roots were washed and assessed for nodulation.

Soil chemical analysis from Yatesbury soil

Agricultural soil was collected in mid-August 2017 from Yatesbury House Farm (Yatesbury, Calne, Wiltshire, UK). This farm was selected because we have direct

contact with the owner. Soil was dried it at room temperature for two weeks, then mixed thoroughly and removed stones with a 5 mm pore-diameter sieve. Soil samples were sent to James Hutton Limited (Aberdeen, Scotland) where they were stored at 4 °C on receipt. A sub-sample was obtained and air dried (30 °C) prior to sieving to pass a 2mm aperture. Determination of soil pH and elemental concentrations extracted with 0.43M acetic acid were obtained from the air-dry soil. Elemental concentrations were determined using inductively coupled plasma–optical emission spectroscopy (ICP-OES). Nitrate and ammonium concentrations were obtained using colourimetry after extraction of the field moist soil with 1 M KCl. The moisture content of both the air-dried (30 °C) and field moist soil were determined at 105 °C.

Table 2. Chemical analysis performed by James Hutton Limited

Methods	**Accreditation[3]**
pH Determination	DM007
Soil Extraction – 0.43M Acetic Acid – Ca, Mg, K & P	DM005
Inductively coupled plasma – Optical Emission Spectroscopy (ICP-OES)	BM005
Soil Extraction – 1M KCl – NO_3 & NH_4	N/A
Analysis 1M KCl Extractions	N/A
Moisture Determination	DM007

Most probable number (MPN) of indigenous rhizobia from Yatesbury soil

Following the methodology described by Somasegaran and Hoben (1994) and Howieson and Dilworth (2016). Four biological replicates of an inoculum were prepared by mixing 1 g of soil in 10 ml of sterile D-H_2O. Serial dilutions from 10^{-1} to 10^{-9} were done for each replicate. Boiling tubes were filled with sterile fine vermiculite and

[3] United Kingdom Accreditation Service (UKAS) approved methods

20 ml nitrogen-free rooting solution (Poole *et al.*, 1994). Surface-sterilised pea seeds were placed in the boiling tubes and inoculated with 1 ml of the dilutions 10^{-2} to 10^{-9}. A Water Control (WC) for each of the dilutions was placed randomly amongst the inoculated plants. Pea plants were harvested 21 dpi and assessed for the presence (nodulation-positive) or absence (nodulation-negative) of nodules. Subsequently, ten 0.5 L pots were prepared with Yatesbury soil and then weighed to obtain total cells of indigenous rhizobia per pot.

Competition assay: co-inoculation with two strains

This protocol was modified from that described by Sánchez-Cañizares & Palacios (2013)

Rhizobium cultures to be used as inoculum were grown to early stationary phase (OD_{600} < 0.6 - 0.8). Cultures were adjusted to 1000 cfu/ ml. An inoculum mix was prepared with approx. 1000 cfu in total of three different inoculation ratios: 10:1, 1:1 and 1:10 strain A vs strain B, This mix was added to 75 ml 2.6-times concentrated nitrogen-free rooting solution (Poole *et al.*, 1994) and distributed homogenously into 500 ml sterile pots with a mix of 50% silver sand/fine vermiculite. Seeds of pea (Avolar) were surface sterilised and germinated on 1% agar plates. Five day-old and similar sized seedlings were transferred to pre-inoculated pots. Plants without inoculum were grown as a negative control (Water Control or WC). Plants were harvested at 21 dpi and were sequentially stained first with Magenta-glcA (5-bromo-6-chloro-3-indolyl-β-D-glucuronide acid) and then with X-gal after thermal treatment of 1 h at 70°C (Sessitsch *et al.*, 1996).

Competition assay: co-inoculation with multiple rhizobial strains

Rhizobium cultures to be used as inoculum were grown to early stationary phase (OD_{600} < 0.6 - 0.8). The OD_{600} of a 200 µl sample from each strain was measured in 96-well plate using a FLUOStar Omega Microplate Reader (BMG). Cultures were adjusted to the desired OD and mixed in a flask. Inoculum was prepared by adding 1ml of the Tagged-strain mix in 74 ml 2.5-times concentrated nitrogen-free rooting solution (Poole *et al.*, 1994) and then distributed homogenously into 500 ml pots. Seeds of pea (Avolar) were surface sterilised and germinated on 1% agar plates. Five-day-old similar sized seedlings were transferred to pre-inoculated pots. Pea plants were grown without inoculum as negative controls (WC).

Measurement of shoot dry weight

Pea plant shoots were placed in paper envelopes and incubated at 70 °C for 48 hours. Dried shoots were weighed using an Acculab Sartorius Group scale model ALC-80-40.

Acetylene reduction assay

Acetylene reduction assays (ARA) were used to determine the activity of nitrogenase (Hardy *et al.*, 1968). Plants were placed in 250 ml Duran bottles (total volume 320 ml) with a wet piece of paper to prevent drying out during the process. Neoprene lids were used on the Duran bottles to allow for injection whilst not breaking the gas seal. Air was removed from each bottle using a syringe and needle (8ml) and substituted with 6.4 ml of acetylene, to give 2% acetylene per bottle. Bottles were incubated for 1 hour and subsequently three samples of 1 ml from each bottle were removed using a

syringe and needle. Needles were left in rubber plugs to prevent gas escaping from the syringes. Each sample was analysed using a Gas Chromatograph System Clarus® 480 (Perkin-Elmer) in order to determine the reduction of acetylene to ethylene in the bottles, and then was converted to a rate of nitrogenase activity $[plant]^{-1}$ $[hour]^{-1}$. 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 261.59 µmol (based on 6.4 ml of gas added and 24.465 $Lmol^{-1}$). Multiplying the fraction of acetylene converted by the amount of acetylene initially added, yields the amount of ethylene produced.

Nodule collection and re-isolation of nodule bacteria

Following the methodology described by Johnston & Beringer, (1975) and Poole et al. (1994) at least three nodules from each plant were removed and surface-sterilised by immersion in 2% sodium hypochlorite for 5 minutes, washed three times in sterile water and crushed into 100µl sterile distilled water. Aliquots of each nodule extract were spotted onto TY agar plates. Eight independent colonies were then replica-plated and screened for appropriate plasmid antibiotic marker.

Detection of enzyme activity of β-glucuronidase (GusA) and β-galactosidase (CelB) in nodules

The protocol was modified from that described by Sessitsch et al., (1996). Pea plants were harvested at 21 dpi and bean plants at 42dpi. Roots were cut where the last nodules are observed.

β-glucuronidase activity from cloned *gusA*

Roots were submerged upside down in a 50 ml Falcon tube containing 40 ml of phosphate buffer supplemented with 400 µl of Magenta-glcA (200 µg/ml) or X-glcA (250 µg/ml). Tubes were covered with aluminium foil and incubated over night at 28 °C.

Thermostable β-galactosidase activity from cloned *celB*

Roots were submerged upside down in a 50-ml Falcon tube containing 40 ml of phosphate buffer and kept at 70°C for 1 hour in order to destroy endogenous β-galactosidases. After cooling, 500 µl of X-gal solution (250 µg/ml) was added. Tubes were covered with aluminium foil and incubated over night at 28 °C.

Sequential double staining to detect both β-glucuronidase and thermostable β-galactosidase activity

The gusA activity assay was run first with Magenta-glcA as substrate and, after overnight incubation, the *celB* activity assay was done.

Image acquisition

Image acquisition for luminescence expression

Bacteria on agar petri dishes or full legume roots systems placed in square plates were photographed using a NightOWL LB-983 camera (Berthold Technologies). Following the methodology described by Pini *et al.* (2017) and the *in vivo* Imaging System protocol provided by manufacturer, luminescence CCD images output were exposed for 120 seconds. Each CCD image consisted of an array 1,024 by 1,024 pixels, and after acquisition, images were post-processed for cosmic suppression and back- ground correction. All images were analysed with the software IndiGO (Berthold Technologies).

Image acquisition for green and red fluorescence expression (using a ring-light)

Fluorescence CCD images were acquired with a ring-light epi illumination accessory for NightOWL LB-983 camera. Samples were exposed for one second with special filters for quantification of GFP (excitation wavelength 475/20 nm and emission wavelength 520/10 nm) and mCherry (excitation wavelength 550/10 nm and emission wavelength 620/10 nm). All images were analysed with the software IndiGO (Berthold Technologies).

Image acquisition for green fluorescence (blue-light transilluminator)

To increase the area of green fluorescence quantification by the NightOWL camera to cover a full root system, a blue-light transilluminator (VWR) together with a Kodak Wratten Gelatine Filter no 49 were used. In this case the exposure was 0.1 second and

only the emission filter was needed. All images were analysed with the software IndiGO (Berthold Technologies).

Dissecting microscope images

Photographs were taken using a dissecting microscope (Leica M165 FC) with a Leica DFC310 FX digital camera accompanying software (LAS v4.5). Visible light filter, GFP and red fluorescent filters were used.

Confocal microscopy for bacteroids analysis

A 42-dpi pea nodule was crushed and resuspended in 200 µl of water. 20 µl of sample was placed on a microscope slide in order to perform analysis of individual bacteroids. Confocal microscope images were acquired using an HCX PL APO CS 20x/0.7 IMM UV lens on a Leica SP 5. GFP was excited at 488 nm. 3D-segmentation was performed in MorphographX following Barbier de Reuille *et al.*, (2015). A GFP TIF image stack was loaded into MorphoGraphX, filtered using a Gaussian Blur Filter with a radius of 0.1 µm, and converted into a binary stack (threshold: 800). Meshes were generated using the Marching Cubes Algorithm at a cube spacing of 0.1 µm (3x smoothing) to obtain the volume of individual bacteroids. Individual bacteroids were colour identified according to their volume.

Computational Methods

In silico cloning

In silico cloning reactions were tested in Geneious 10.0.9. Nucleotide alignments were carried out using BLASTn in National Centre for Biotechnology Information (NCBI) (Coordinators, 2013) or within Geneious 10.0.9.

Bioinformatic analysis

Paired-end merging of amplicon reads was done with USEARCH 10.0.240 (Edgar, 2010). Quality filtering of amplicon reads was carried out with Trimmomatic-0.36 (Bolger *et al.*, 2014) and we developed a script in Biopython (Cock *et al.*, 2009)

Data manipulation and statistical analysis

Data manipulation of PivotTables and conditional tables were done in Microsoft Excel, and data was processed in the statistics program Prism (GraphPad Software) 7.

Chapter 3: Development and validation of a vector for high-throughput cloning in rhizobia

Introduction

One of the key aspects of applied microbiology is the monitoring of bacteria in nature, ideally with as little perturbation as possible. In the case of the rhizobal-legume system, different strategies have been used to label and follow bacteria. These include chromosomally-encoded marker genes (Wilson *et al.*, 1995b; Sessitsch *et al.*, 1997, 1998) and naturally occurring or induced antibiotic-resistance genes (Oresnik *et al.*, 1999; Poysti *et al.*, 2007; Geddes & Oresnik, 2012), but these methods are highly time-consuming since one involves a complex procedure of chromosomal integration of genes and in the other it is necessary to manually remove root nodules, crush them to extract bacteria and then conduct a strain differentiation test on antibiotic plates.

Plasmid-based marker systems to monitor rhizobia in symbiosis have often suffered poor plasmid maintenance hampered by the inability to apply antibiotics and select for plasmid-borne resistance (Schmidhauser & Helinski, 1985; Weinstein *et al.*, 1992; Corich *et al.*, 2001). The stable promoter-probe vector pJP2 (13.1 kb) described by Prell et al. (2002), has been a useful addition for work in this area as it is maintained stably in rhizobial strains without the need for antibiotic selection (Karunakaran *et al.*, 2005; Tett *et al.*, 2012; Pini *et al.*, 2017). Unfortunately, pJP2 is a standard cloning vector which still presents an obstacle (time-wise) for construction of the large numbers of reporter-plasmids needed in this project.

Golden Gate cloning is a DNA assembly technology that utilises type IIS restriction enzymes to assemble multiple fragments of DNA in a linear order (Weber *et al.*, 2011). Golden Gate cloning platforms have been developed for plants, fungi and *E. coli* (Terfrüchte *et al.*, 2014; Engler *et al.*, 2014; Agmon *et al.*, 2015; Moore *et al.*, 2016). We have recently used the principles of Golden Gate cloning for the rapid construction of bacterial cloning vectors for use in a wide-range of bacteria (Geddes *et al.*, 2018). Work reported in Geddes et al. (2018) also forms part of this thesis and is described in this chapter. In Geddes et al., (2018), we report the construction of pOGG026, a broad host range, low copy-number vector, based on pJP2, and, in addition, compatible with high-throughput Golden Gate cloning. As it has a broad host range, constructs can be made and confirmed in *E. coli* before transfer to rhizobia to investigate an environment of interest through further experimentation.

This chapter describes the design and construction pOGG026 and demonstrates its expression in free-living rhizobia and bacteroids even in the absence of antibiotics. Experiments show how the high-throughput cloning capacity of pOGG026 can be used to construct strains to follow bacterial gene expression in nitrogen-fixing nodules during symbiosis and can be applied to studying competition between rhizobial strains to form nodules on legume roots.

Results and discussion

Design and construction of pOGG026

We have designed a bacterial vector assembly system, based on Golden Gate cloning and the MoClo system (Weber *et al.*, 2011), with a series of discrete key modules

based on the Standard European Vector Archive (SEVA) (Silva-Rocha *et al.*, 2013; Martínez-Garćía *et al.*, 2015) (Geddes et al., 2018). In the SEVA system, modules are arranged in a predefined order and assembled using standard restriction enzymes and standard restriction/ligation reactions. We designed our bacterial vector with a similar module system but assembled using Golden Gate cloning. Modules were assigned a universal position for linear assembly based on their function; position 1 - cloning site(s), position 2 - antibiotic resistance, position 3 - origins of replication and transfer, with positions 4 - 6 reserved for accessory modules (Fig. 3. 1). Modules at each position (1 to 6 and Endlinkers) have inward-facing Esp3I sites which on digestion give 4 bp-overhangs. These overhangs are able to anneal with those from neighbouring modules in a linear fashion and, on vector assembly, give 4 bp motifs which remain at the join between modules.

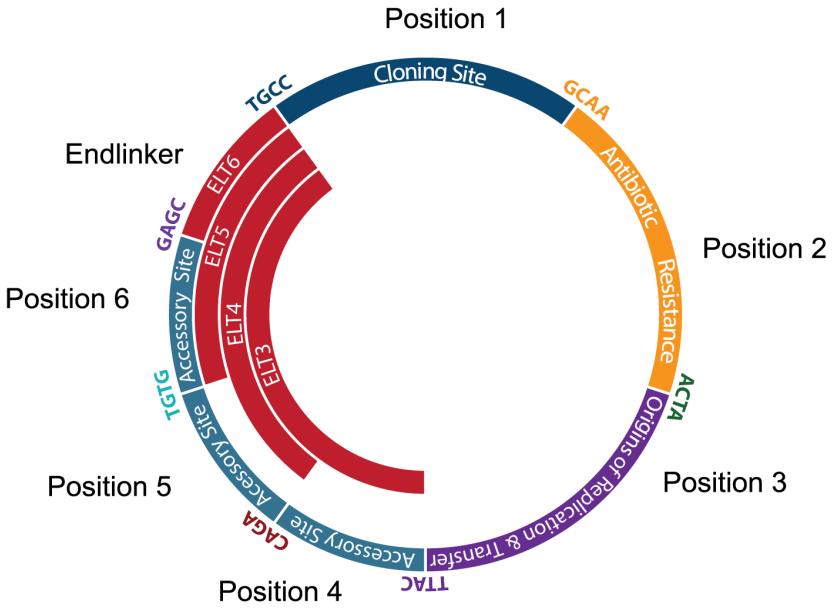


Figure 3. 1 Standardised system for vector assembly of bacterial vectors. Vector modules are designated a 'position' and combined in the vector construction as shown. Sequences (5'→ 3') used to anneal the fragments in order are shown (colour-coded) at junctions between modules and remain in the final construct as a scar.

Modules described in Figure 3. 2A and Table 3. 1 were used to construct vector pOGG026 (6.6 kb, Fig. 3. 2B) in a one-pot reaction (digestion with Esp3I and ligation). The modules are: position 1; level 1 cloning site from pOGG004, position 2; neomycin resistance gene from pOGG008, position 3; RK2-based origin of replication including *oriT* from pOGG010, position 4; *parABCDE* genes which confer stability in the absence of antibiotics from pOGG012 and the appropriate Endlinker (ELT4) from pOGG014. Endlinker modules contain the *E. coli* T_1 *rrnBT1* terminator to help buffer any transcription effects from the cloning site (Lutz & Bujard, 1997). No position 5 or 6 modules were needed to construct pOGG026.

Table 3. 1 Modules used to construct pOGG026

Module	Description of module	Module size (kb)	Cloned module	5'-side (5'- 3')	3'-side (5'- 3')
Position 1: Cloning site				TGCC	GCAA
pLVC-P1-Lv1	Golden Gate Level 1 cloning site with cloned *lacZ* and T_0 lambda phage transcription terminator	2601	pOGG004		
Position 2: Antibiotic resistance				GCAA	ACTA
pLVC-P2-neo	Neomycin/kanamycin-resistance	936	pOGG008		
Position 3: Origin of replication and transfer				ACTA	TTAC
pLVC-P3-RK2	RK2 (IncP) origin of replication for low copy number plasmid, *oriT* from *E. coli* for plasmid transfer	2485	pOGG010		
Position 4: Accessory				TTAC	CAGA
pLVC-P4-par	Partition genes (*parABCDE*) for plasmid stability in absence of antibiotic selection	2408	pOGG012		
Endlinker				CAGA	TGCC
pLVC-ELT-4	Endlinker to circularize plasmid after position 4	113	pOGG014		

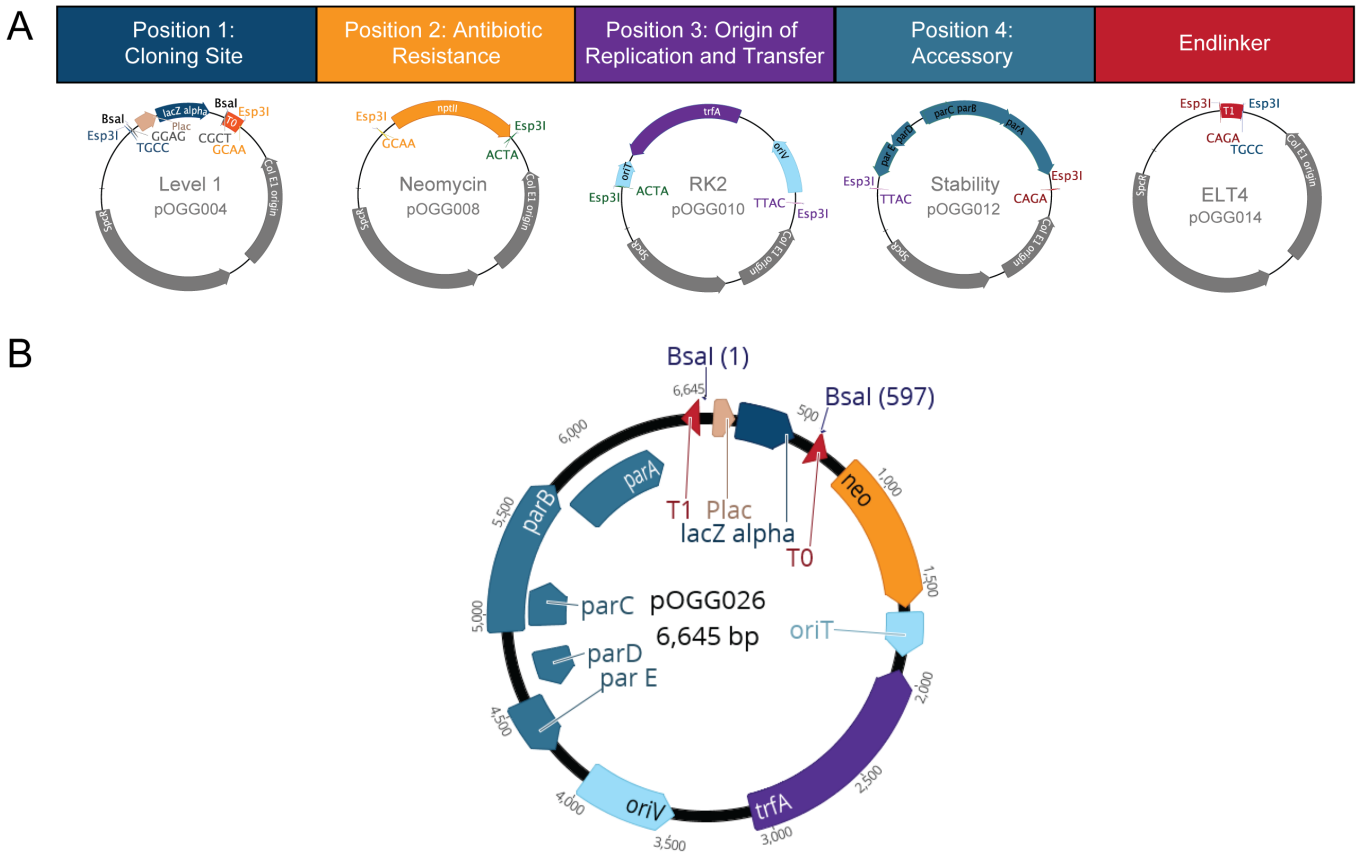


Figure 3. 2 A) Modules used to construct pOGG026. Position 1: Cloning site - Golden Gate Level 1 cloning site; position 2: Antibiotic resistance – neomycin resistance; position 3: Origin of replication and transfer - RK2 with *oriT*; position 4: Accessory - *par* locus for stable plasmid maintenance in the absence of antibiotic selection; Endlinker: ELT4 to circularise the plasmid. B) Schematic map of pOGG026 (6.6 kb) Level 1 broad host range, low copy number vector constructed for stable environmental gene expression in the absence of antibiotic selection.

An advantage of this modular assembly is the ability to utilise minimal parts by excluding any non-functional DNA. This allows construction of vectors with the same key features as those currently widely adopted, but the resulting plasmid is significantly smaller in size. This can be observed in the difference of size pOGG026 (6.6 kb) compared to that of pJP2 (13.1 kb), the plasmid on which pOGG026 is based. The reduced size enhances the ease of DNA manipulation and allows accommodation of larger cloned inserts.

High-throughput cloning in pOGG026 using Golden Gate assembly

Golden Gate cloning can be used to combine Level 0 parts (genetic features controlling a gene's transcription (P module), translation (U module) (often together as a PU module) the open reading frame (ORF) (SC module) and termination of transcription (T module) in a single one-pot reaction (referred to as Level 1 assembly) (Fig. 3. 3). Subsequently, in hierarchical assembly, multiple Level 1 modules can be combined together in a vector to form a Level 2 multigene construct (Weber *et al.*, 2011).

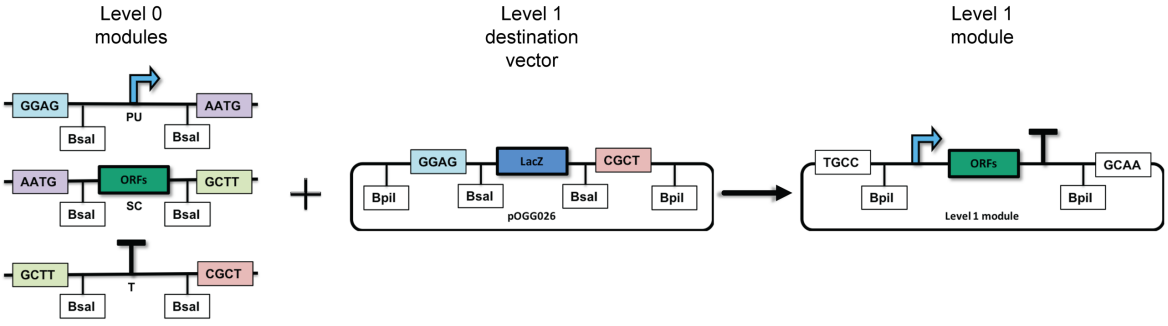


Figure 3. 3 Flexible cloning system Level 0 modules used in Golden Gate assembly to make a Level 1 transcriptional unit. Each Level 0 module (PU, SC, T) has unique overhangs (colour coded) which are used to assemble multiple fragments of DNA in a linear order. Plasmid pOGG026 is a Level 1 destination vector.

In order to test pOGG026 as a vector in Level 1 cloning (outlined in Fig. 3. 3), two promoters routinely used for gene expression in rhizobia were adapted to form PU modules; the constitutive neomycin cassette promoter (pNptII) (Frederix *et al.*, 2014) and the symbiosis-induced *nifH* promoter from *R.leguminosarum* biovar *vicia* 3841 (Rlv3841pNifH) (Johnston & Beringer, 1975) (Table 3. 2). For these PU modules, the nucleotide immediately upstream of the ATG was domesticated to an A to construct the 5'-AATG-3' cut-site, such that the ATG encodes the start codon of the ORF to be expressed.

Three reporter genes were adapted to form SC modules; superfolder green fluorescent protein (sfGFP) (Pédelacq *et al.*, 2006), *E. coli* β-glucoronidase (*gusA*) and *Pyrococcus furiosus* thermostable β-glucosidase/ β-galactosidase (*celB*) (Wilson *et al.*, 1995b; Sessitsch *et al.*, 1996, 1997; Sánchez-Cañizares & Palacios, 2013) (Table 3. 2). All SC modules were constructed flanked by an inward-facing BsaI restriction site cutting 5'-AATG-3' at the 5'-end and 5'-GCTT-3' at the 3'-end (Fig. 3. 3). In these SC modules, the final three nucleotides of the scar 5'-AATG-3' (complementary to the PU overhang, see above) encode the ORF start codon and, at the other end of the module, the 5'-GCTT-3' scar immediately follows the ORF stop codon.

The only T module used here was T-pharma (pOGG003) (Table 3. 2) which has 5'-GCTT-3' and 5'-CGCT-3' at the 5'- and 3'- ends of the T module, respectively (Fig. 3. 3).

Table 3. 2 Level 0 modules cloned in pOGG026 Level 1 vector.

Module	Description of module	Module size (bp)	Cloned module	5'-side (5' - 3')	3'-side (5' - 3')
PU modules: promoters with ribosome-binding sites				GGAG	AATG
pL0M-PU-pNeo	Promoter upstream of *nptII* for constitutive gene expression	386	pOGG001		
pL0M-PU-pNifH	Promoter upstream of Rlv3841 *nifH* for nodule-specific gene expression in *R. leguminosarum*. Made by PCR using oxp0474 and oxp0475 primers and cloned in destination vector pOGG068	721	pOGG082		
pL0M-PU-psNifH	Synthetic consensus promoter to drive nodule-specific expression in biovars and strains of *Rhizobium leguminosarum.*	258	pOGG043		
SC modules: ORFs				AATG	GCTT
pL0M-SC-sfGFP	Superfolder GFP, stable and bright reporter of gene expression	722	pOGG037		
pL0M-SC-celB	Thermostable beta-galactosidase, reporter of gene expression	1424	pOGG050		
pL0M-SC-gusA	Beta-glucosidase gene, reporter of gene expression. Made by PCR using oxp0376 and oxp0377 primers and cloned in destination vector pOGG072	1817	pOGG083		
T modules: terminators				GCTT	CGCT
pL0M-T-pharma	Pharmacia terminator (Invitrogen)	188	pOGG003		

All modules were supplied cloned in Sp^r plasmid pMS except for pOGG082 and pOGG083. Following BsaI digestion of cloned modules and a Level 1 vector, the resulting overhangs are used to anneal the fragments in a linear fashion and the sequences shown remain between modules upon ligation.

Table 3. 3 shows plasmids constructed by Golden Gate cloning which are used to validate pOGG026 as a suitably vector (see below). Although these plasmids all have the same vector and T modules, the PU and SC modules are different, illustrating the ease by which plasmids can be designed and constructed with an interchangeable modular system. A simple one-pot reaction, mixing preformed plasmids for BsaI digestion and ligation (approx. 3 hrs), followed by transformation of *E. coli* cells to give colonies after overnight growth on agar plates, is a simple, straight-forward and time-saving method to construct these eight plasmids. Conjugation of the plasmids into rhizobial cells is also a reliable technique used routinely. Taken together these techniques have shown how plasmids, based on pOGG026, can be quickly assembled and transferred to rhizobia.

Table 3. 3 Plasmids made with vector pOGG026 using high-throughput Golden Gate assembly.

Module/ plasmid	PU (clone)[a]	SC (clone)[a]	T (clone)[a]	Vector
pOPS0253	pNifH (pOGG082)	*gusA* (pOGG083)	T-pharma (pOGG003)	pOGG026
pOPS0254	pNifH (pOGG082)	*celB* (pOGG050)	T-pharma (pOGG003)	pOGG026
pOPS0262	psNifH (pOGG043)	*celB* (pOGG050)	T-pharma (pOGG003)	pOGG026
pOPS0263	psNifH (pOGG043)	*gusA* (pOGG083)	T-pharma (pOGG003)	pOGG026
pOPS0314	pNeo (pOGG001)	*celB* (pOGG050)	T-pharma (pOGG003)	pOGG026
pOPS0377	pNeo (pOGG001)	sfGFP (pOGG037)	T-pharma (pOGG003)	pOGG026
pOPS0379	pNifH (pOGG082)	sfGFP (pOGG037)	T-pharma (pOGG003)	pOGG026
pOPS0750	pNeo (pOGG001)	*gusA* (pOGG083)	T-pharma (pOGG003)	pOGG026

[a] Plasmid of cloned module is shown in brackets.

To illustrate the use of high-throughput construction, plasmids were made to act as positive controls in experiments to test nodule-specific expression. Plasmids

pOPS0750, pOPS0314 and pOPS0377 have the pNeo promoter (PU module), giving constitutive expression in rhizobia, upstream of *gusA, celB* and sfGFP, respectively (Fig. 3. 4A-C, Table 3. 3). These plasmids show reporter gene expression in free-living *R. leguminosarum* 3841 (Rlv3841) (Fig. 3. 4D-F) and were an important positive control in experiments examining nodule-specific gene expression (see below).

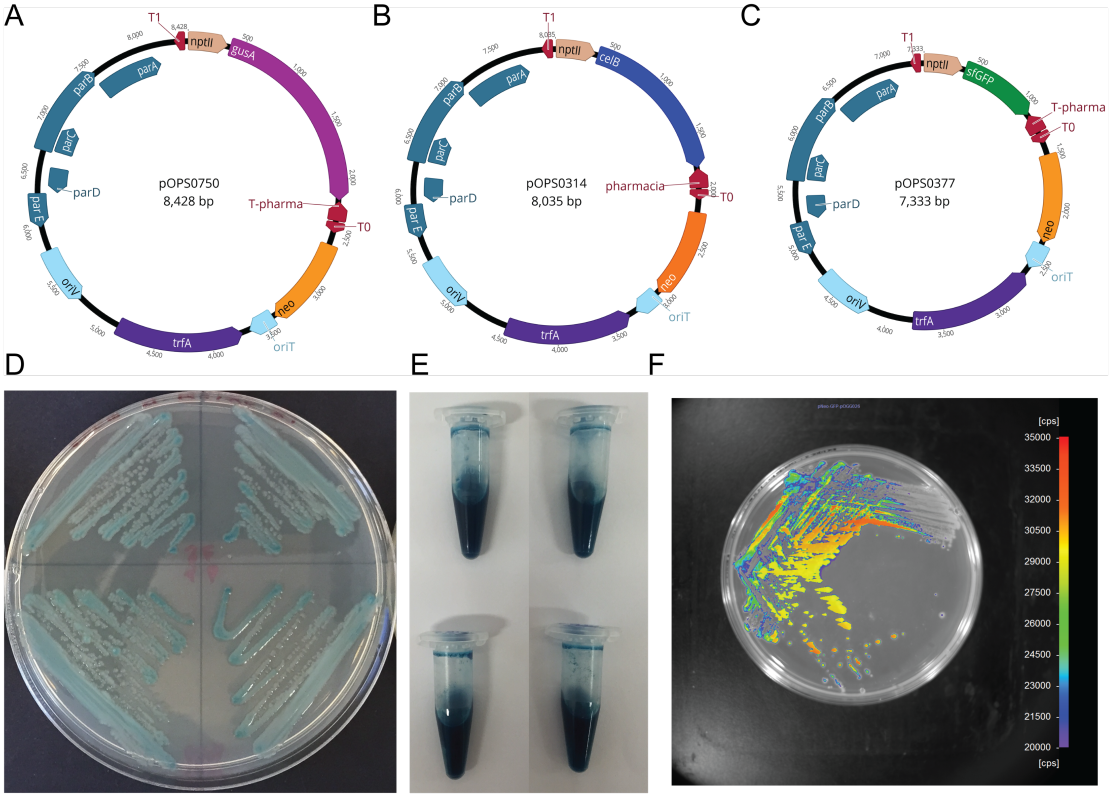


Figure 3. 4 Schematic maps of plasmids A) pOPS750 containing cloned gusA, B) pOPS0314 containing cloned celB and C) pOPS0377 containing cloned sfGFP. Each cloned gene is under the control of a constitutive promoter (pNeo). Free living cells of D) Rlv3841[pOPS750] stained with TY X-glucA, E) Rlv3841[pOPS314] stained with X-gal after thermal treatment and F) green fluorescence expression from Rlv3841[pOPS0377] (scale, 2,000–35,000 cps).

Assessing marker gene detection in symbiosis using vector pOGG026

To investigate marker gene detection derived from pOGG026, and therefore its suitability for use in experiments carried out in different environments it was used to construct plasmid pOPS0253 (Fig. 3. 5A) by Golden Gate cloning. Using a Level 1 one-pot reaction (BsaI digestion and ligation) (Engler *et al.*, 2008), pOPS0253 was constructed from the following components: vector; pOGG026, PU module; pNifH from pOGG082, SC module; *gusA* from pOGG083; T module; T-pharma terminator from pOGG003 (Table 3. 2, Table 3. 3). Plasmids pOPS0254 (Fig. 3. 5B) and pOPS0379 (Fig. 3. 5C) were constructed in a similar way to pOPS0253, except that their SC modules are *celB* (from pOGG050) and sfGFP (from pOGG037) (Table 3. 2, Table 3. 3).

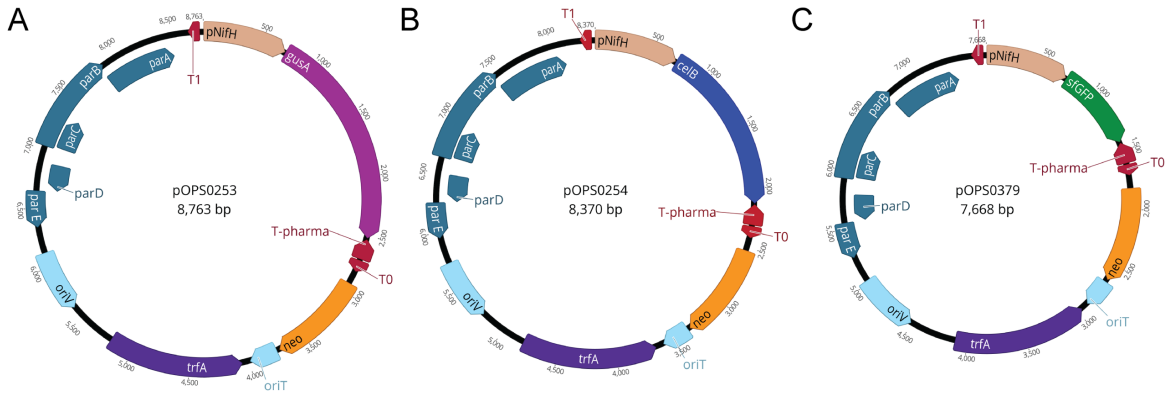


Figure 3. 5 Schematic maps of plasmids A) pOPS0253 containing cloned *gusA*, B) pOPS0254 containing cloned *celB* and C) pOPS0379 containing cloned sfGFP. Each cloned gene is under the control of *nifH* promoter (pNifH).

Plasmids pOPS0253, pOPS0254 and pOPS0379 were conjugated into Rlv3841 and used to inoculate peas plants. Plasmid pOPS0254 was also conjugated into *R. tropici* CIAT 899 (Graham *et al.*, 1982) and used to inoculate common bean plants.

Expression of the GusA and CelB chromogenic markers from plasmids pOPS0253 and pOPS0254, respectively, can be followed by staining either pink (*gusA*) or blue (*celB*) when the appropriate substrate is supplied (Sessitsch *et al.*, 1996). Examples of stained pea nodules are shown in Figure 3. 6A-B and stained bean nodules are shown in Figure 3. 6C. In each case all the nodules were stained, and none were visible that did not take up the stain.

The best examination of plasmid loss in symbiosis is by visualisation of stained or fluorescent nodules since, in indeterminate nodules, rhizobia differentiation from free-living to bacteroids is irreversible and bacteroids are not cells which can re-grow (Prell *et al.*, 2002).

In our assays, all mature nodules showed fluorescence or corresponding staining after enzymatic reaction with Magenta-glcA or X-gal, but to further verify plasmid stability, at least three nodules from each plant were removed and surface-sterilised and crushed into 100µl sterile distilled water. Aliquots of each nodule extract were spotted onto TY agar plates. Afterwards, eight independent colonies from each crushed nodule were screened for antibiotic resistance and all colonies showed neomycin resistance, corresponding to plasmid vector pOGG026. The inducible expression of sfGFP can be observed when nodules reach a mature stage, with expression being confined to the nitrogen-fixing zone (Fig. 3. 6D).

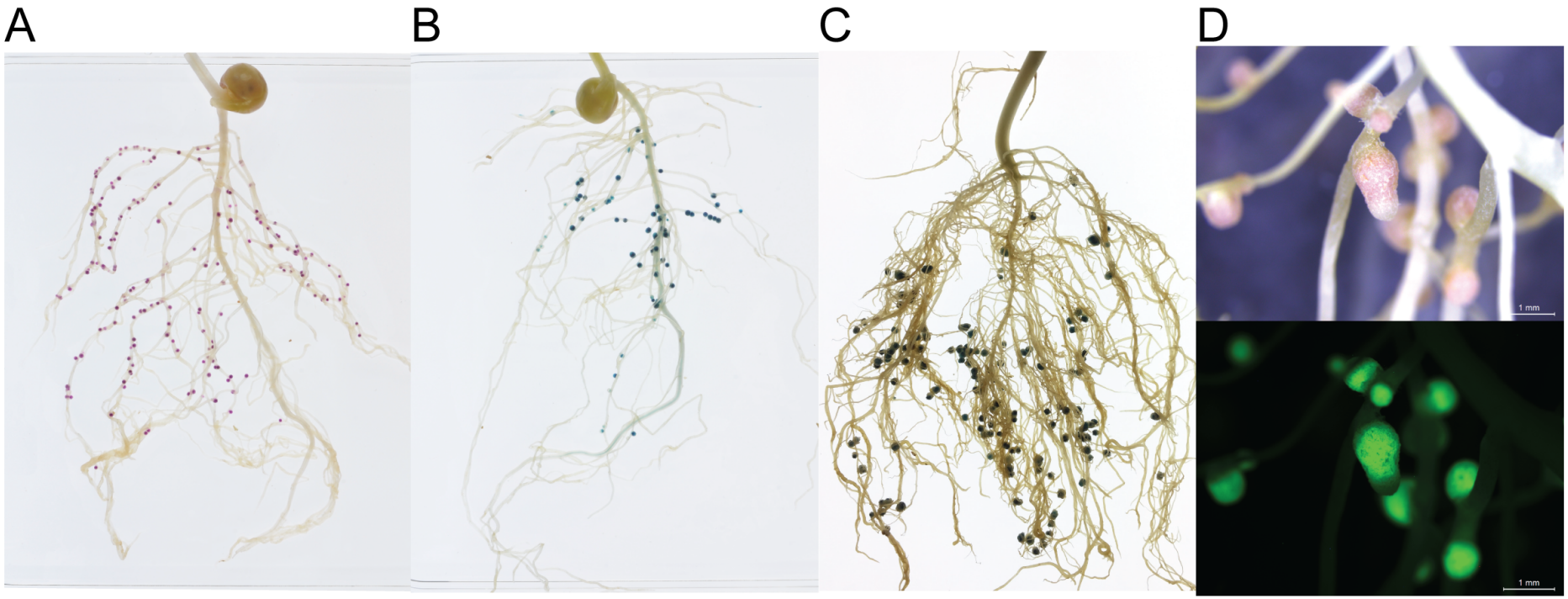


Figure 3. 6 A) Pea nodules formed by Rlv3841[pOPS253] stained with Magenta-glcA. Marker gene *gusA* is only expressed in nodules. B) Pea nodules formed by Rlv3841[pOPS254] and C) bean nodules formed by CIAT899[pOPS253] stained with X-gal after thermal treatment. Marker gene *celB* is only expressed in nodules. D) Pea nodules formed by Rlv384[pOPS0379], bright-field (top) and fluorescence (GFP filter) (bottom) images. Fluorescence detection is exclusively found in the nitrogen-fixing zone. Pea root systems were harvested at 21 dpi and those of common bean were harvested at 42 dpi. Images shown are representative of three biological replicates.

In nodules plasmids are often lost and there is an inability to apply antibiotic selection to maintain those which have a plasmid-borne resistance (Schmidhauser & Helinski, 1985; Weinstein *et al.*, 1992; Corich *et al.*, 2001). Here we have validated pOGG026 for monitoring strains during symbiosis. Marker gene detection in nodules is due to the RK2-based origin of replication (Figurski & Helinski, 1979) and the presence of the *parABCDE* genes (Prell *et al.*, 2002) in pOGG026.

While the first step of plasmid stability has been demonstrated in short term plant assays, a more rigours analysis of plasmid stability, as it is described in great detail in Schmidhauser & Helinski, (1985), Blatny et al. (1997) and Corich et al. (2001), are recommended.

Due to the fact that marker genes have been detected in all mature nodules in assays, as well as the ease and speed at which plasmids are constructed, we propose that this plasmid-based marker system can be used as an alternative to the chromosomally-encoded strategies previously presented (Wilson *et al.*, 1995b; Sessitsch *et al.*, 1997, 1998).

Using the plasmid-based marker system to assess rhizobial competitiveness

As explained in Chapter 1, if two strains are equality competitive an equal relationship between the percentage of formed nodules from each strain and their percentage of cells in the inoculum mix is expected. In Sánchez-Cañizares and Palacios (2013) they applied this idea by performing competition assays with known inoculation ratios: 10, 50 and 90% of *R. leguminosarum bv. viciae* UPM791 (Sánchez-Cañizares *et al.*, 2018) versus Rlv3841 and scored nodule formation 21 dpi from each

strain to evaluate their competitiveness in pea plants. Their results showed that UPM791 was significantly more competitive than Rlv3841.

This marker system has been used to great effect in the evaluation of rhizobial competitiveness, but since *gusA* and *celB* need to be incorporated in the chromosome of the strains to be analysed, it is still a time-consuming system. Therefore, we decided to compare the chromosomally-encoded system with our plasmid-based system by repeating as closely as possible the same competition experiment presented by Sánchez-Cañizares and Palacios (2013).

To do this, we labelled UPM791 and Rlv3841 with a plasmid containing *gusA* and *celB* respectively, under the control of a synthetic consensus promoter (psNifH). Like pNifH, psNifH also drives gene expression only in nodules. This consensus promoter and the assessment of reporter plasmids against non-labelled isogenic strain will be explained in detail in Chapter 4.

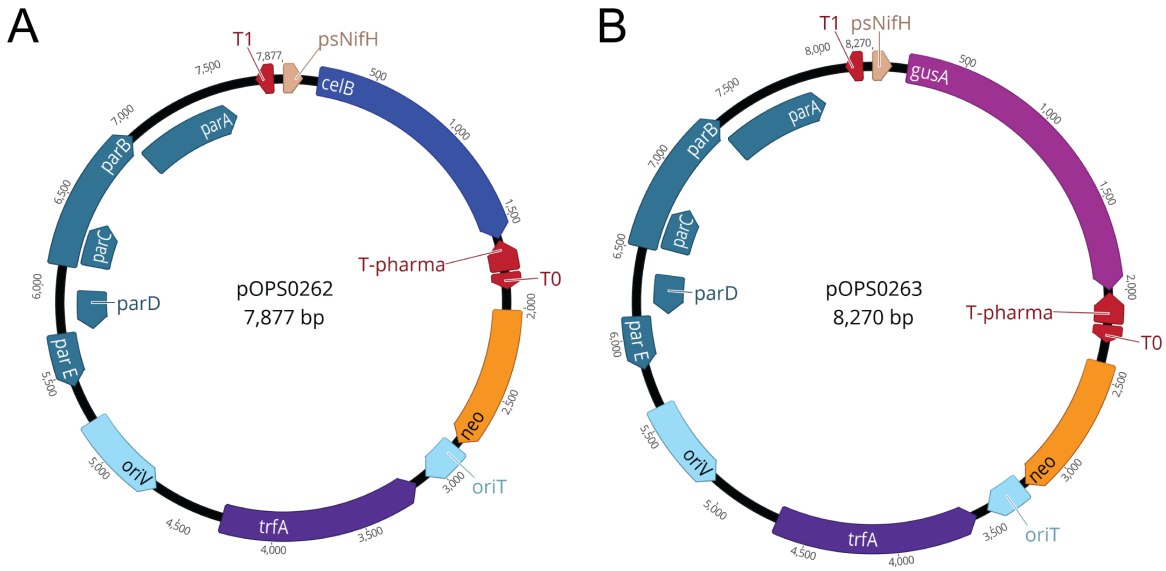


Figure 3. 7 Schematic maps of A) pOPS0262 (cloned *celB*) and B) pOPS0263 (cloned *gusA*). Promoter psNifH drives expression of these genes in nodules.

Rlv3841 was labelled with a plasmid containing *celB* (pOPS0262, Table 3. 3, Fig. 3. 7A) and UPM791 was labelled with a plasmid containing *gusA* (pOPS0263, Table 3. 3, Fig. 3. 7B). These plasmids are almost identical to pOPS254 and pOPS253, respectively (Fig. 3. 5), the only difference is in their PU modules; a synthetic consensus promoter (psNifH) replaces pNifH (Table 3. 3).

A competition assay in pea plants between UPM791[pOPS0263] (*gusA* under the control of psNifH) and Rlv3841[pOPS0262] (*celB* under the control of psNifH) was performed using three different co-inoculation ratios; 10, 50 and 90% of UPM791[pOPS0263] versus Rlv3841[pOPS0262]. At 21 dpi, pea root systems underwent a sequential double staining with Magenta-glcA, and X-gal after thermal treatment (Fig. 3.8 A-C). We scored nodules from UPM791[pOPS0263] and Rlv3841[pOPS0262] from co-inoculated plants. In this case we have decided to show results in terms of UPM791 (Fig. 3.8 D) but directly proportional results correspond to Rlv3841 competitiveness.

If the UPM791 was as competitive as Rlv3841, figure 3.8D should have a slope of 1 and the line should pass through the origin. If the slope does not pass through the origin then the Y-intercept can be used as a competitiveness index, where a positive value will indicate that UPM791 is more competitive and a negative value will indicate that it is less competitive (Amarger & Lobreau, 1982; Sessitsch *et al.*, 1997).

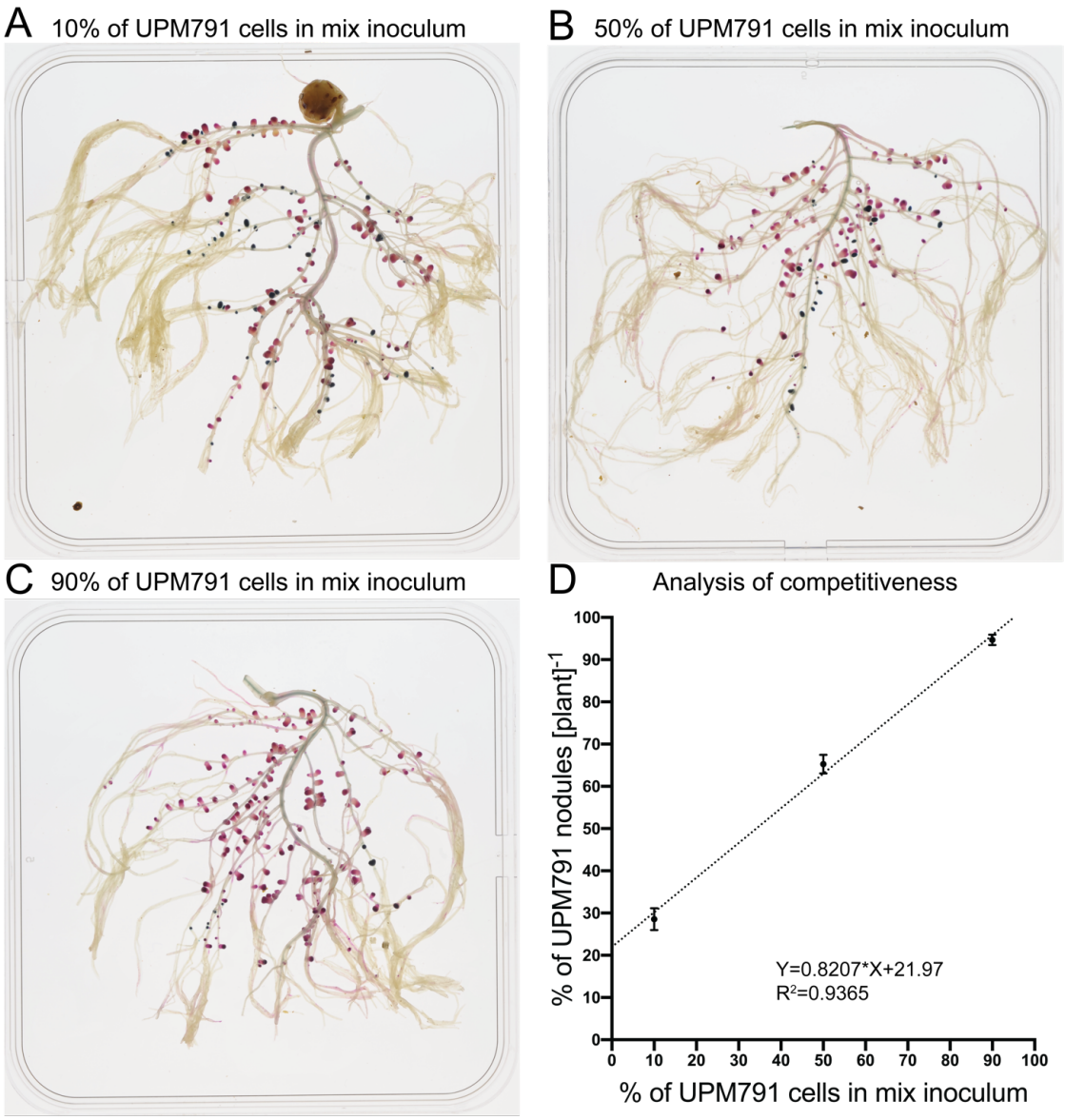


Figure 3. 8 Competition assays for nodule formation between UPM791[pOPS0263] (*gusA* under the control of psNifH) and Rlv3841[pOPS0262] (*celB* under the control of psNifH) on pea roots. Three different co-inoculation ratios; A) 10%, B) 50% and C) 90% of UPM791[pOPS0263] versus Rlv3841[pOPS0262] were used. After sequential double-staining pink nodules were formed by UPM791[pOPS0263] and blue nodules were formed by Rlv3841[pOPS0262]. The analysis of competitiveness was performed by comparing % of UPM791 nodules and % of UPM791 cells in mix inoculum. D) Results from ten independent biological replicates, where bars are the SEM, showed that UPM791 was significantly more competitive than Rlv3841.

Our results showed that the slope of the regression line was 0.8207 ± 0.03719 (95% CI [0.7450 to 0.8964]), significantly different from 1 (p<0.0001) and with a Y-intercept = 21.97 (p<0.0001 different from 0). Therefore, we can conclude that UPM791 was significantly more competitive than Rlv3841. Due to the lack of a competitiveness index in Sánchez-Cañizares and Palacios (2013), is not possible to do a quantitative comparison between their results and our results, but we found the same tendency, UPM791 significantly more competitive than Rlv3841.

Hence, this system can be used for assessing rhizobial competitiveness, but with the added advantage that it is only necessary to perform the simple conjugation of a plasmid from *E. coli* to rhizobia instead of the more complicated and time-consuming (PCR and double recombination) steps needed in the method of Sánchez-Cañizares and Palacios (2013).

Conclusions

In this chapter our aim was the development and validation of a potentially stable bacterial vector for high-throughput cloning in rhizobia. We have designed and constructed a small, broad host range vector, pOGG026, suitable for high-throughput cloning by adapting Golden Gate methodology (Geddes et al., 2018). Vector pOGG026 is crucial for this thesis since it can be used in the flexible and cost-effective multi-fragment assembly through Golden Gate cloning for building the large number of plasmids required.

We have shown that plasmid clones constructed by Golden Gate cloning into vector pOGG026 behave consistently well in free living *R. leguminosarum* and in nitrogen-fixing pea and bean root nodules. Expression of reporter proteins GusA, CelB and GFP florescence detection from these clones confirms that vector pOGG026 can be used in experiments in environments where antibiotic selection is not possible or desirable. Application of pOGG026 has permitted development of the plasmid-based system to mark bacteria with GuSA/CelB and been used to evaluate rhizobial competitiveness in forming nodules on legume roots. In addition, GFP florescence detection in the nitrogen-fixing zone of nodules using pOPS0379, demonstrates a non-invasive, symbiosis-induced screening system which can be used to recover live rhizobia from specific nodules for further analysis. This technique will be developed further in Chapter 4.

Now that a suitable vector, pOGG026, has been developed, the next step is the design and validation of a suitable promoter, specifically expressed only in nodules, but able to drive that expression effectively across a wide variety of different rhizobial strains.

Chapter 4: Design, validation and application of a synthetic consensus promoter

Introduction

The region upstream of the gene encoding nitrogenase (*nifH*), which forms the *nifH* promoter (pNifH), is well-studied and has the characteristics of a type -24/-12 promoter recognised by RNA polymerase sigma factor RpoN (σ54) (see Thöny and Hennecke, 1989 for discussion). This promoter has been used for many years in studying *Rhizobium* competition (Wilson *et al.*, 1995b; Sessitsch *et al.*, 1996; Denton *et al.*, 2003) and more recently it has been used to express the GFP reporter gene as an indicator for nitrogenase gene expression in *Sinorhizobium meliloti* by Lang et al., (2018).

As the experiments we want to perform will include many different rhizobial strains, it essential to standardise, as far as possible, nodule-specific expression in a broad range of *Rhizobium leguminosarum* strains. As substantial sequence data is now available in public databases it can be combined with synthetic biology to identify consensus sequence patterns for particular biological functions. This approach can be used to identify a consensus promoter able to drive nodule-specific expression in biovars and strains of *Rhizobium leguminosarum.* Using this information, a synthetic consensus promoter can be made which will then be introduced into different host rhizobial strains.

The aims of this chapter are: 1) Design a synthetic consensus promoter to drive nodule-specific expression (psNifH) for use in a broad-range of *Rhizobium*

leguminosarum strains, 2) Prove that plasmid-borne psNifH driving expression of a reporter protein does not reduce the competitiveness of strains to form nodules, 3) Show that this promoter can be used as a proxy for nitrogenase activity in individual nodules on pea roots.

Results and discussion

Design of a synthetic consensus promoter to drive nodule-specific expression across a variety of Rhizobium *leguminosarum* biovars and strains

To determine a consensus sequence for a nodule-specific promoter expressed in a wide variety of biovars and strains of *R. leguminosarum* we followed the process described in Figure 4. 1. The starting point was to identify the sequence of the gene *nifH* (pRL100162) and its upstream Inter Genic Region (IGR) from *R. leguminosarum* biovar *viciae* 3841 (Rlv3841*nifH*+IGR).

The Rlv3841*nifH*+IGR sequence was used as a template to find all the genomes of the genus *Rhizobium* in the Integrated Microbial Genomes (IMG) system (Markowitz *et al.*, 2012). As a result, 118 genomes (some of them as draft genomes) were obtained.

The next step was to run a *BLASTn* (DNA vs. DNA E-value of $1e^{-5}$) on the selected 118 genomes. Scaffolds were generated of each strain and those with an *E-value* of 0 were extracted. Each extraction was confirmed in the *National Center for Biotechnology Information* (*NCBI)* (Coordinators, 2013) to verify that they contained *nifH* (although some strains contained more than one copy of *nifH*). For example, *Rhizobium etli* strains contained three copies of *nifH* (a, b and c) but just two copies (a and b) contain

the two complete *nifHDK* operons (Quinto *et al.*, 1985; Peralta *et al.*, 2004). In these cases, only the copies with the complete *nifHDK* operons were considered for design of the consensus *nifH* promoter. Following initial analysis, a new extraction was performed, this time only of the IGR-*nifH* and then aligned by ClustalW in the bioinformatics software Geneious®.

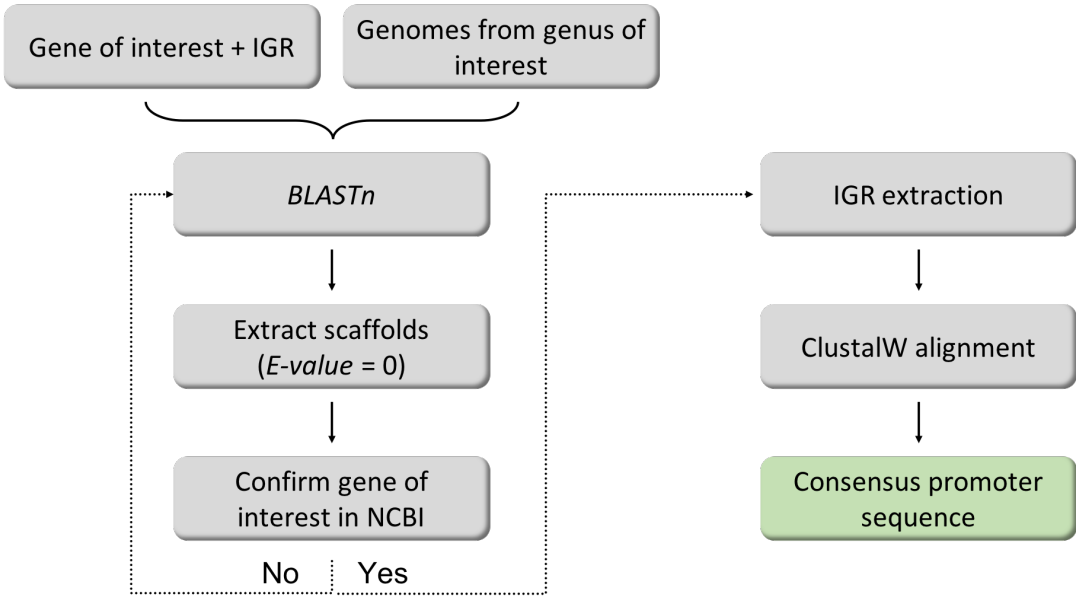


Figure 4. 1 Process followed to determine a consensus sequence for a nodule-specific promoter expressed in a wide range of biovars and strains of *R. leguminosarum.*

The final alignment of the consensus *nifH* IGR from 48 strains of the genus *Rhizobium* is 227 bp compared to 717 bp from the full Rlv3841 *nifH* IGR. This consensus *nifH* IGR (Fig. 4. 2) allowed the identification of three characteristic consensus sequences: an upstream activator sequence (UAS) (TGT-N_{10}-ACA), a binding sequence for RpoN σ^{54} (sigma factor for expression all *nif* genes in nodules) (TGGCACG-N_4-TTG) and a ribosome-binding site (RBS) (GAAGGAA), with the respective spaces in between.

Reports in the literature confirmed the conserved sequences of UAS and the RpoN-binding site (Thöny & Hennecke, 1989; Roelvink & Harmsen, 1990; Brito *et al.*, 1997; Sullivan *et al.*, 2013), and the RBS identified has the characteristic of being rich in purine bases (Oliveira *et al.*, 2004).

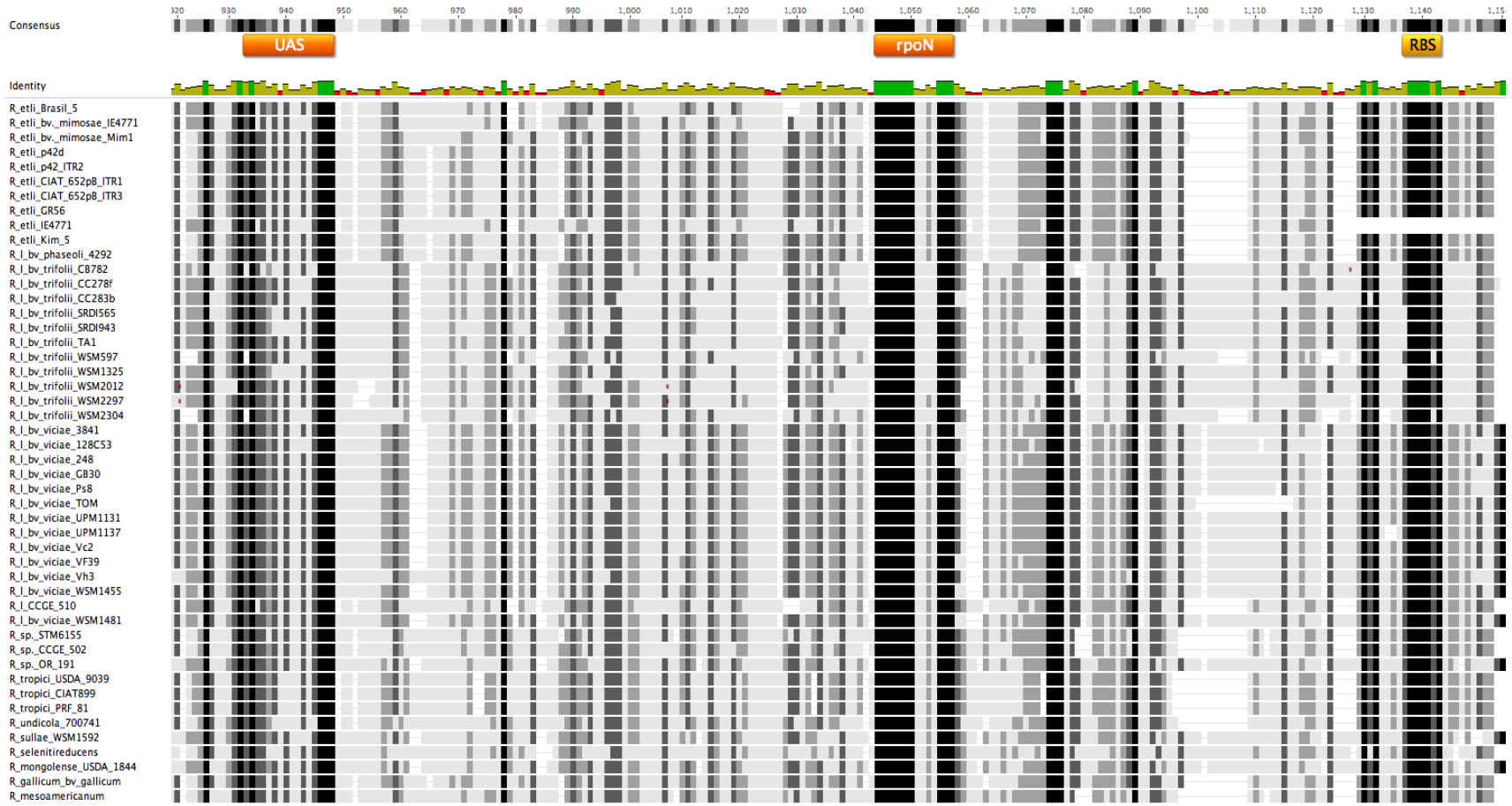


Figure 4. 2 Alignment of the IGR upstream of *nifH* from 48 rhizobial strains. The mean pairwise identity map over all pairs is colour coded: green = 100% identity, green-brown = 99-30% identity and red < 30% identity.

In order that the synthetic consensus promoter would be a PU module for assembly using Golden Gate Cloning, the consensus *nifH* IGR (Fig. 4. 2) was completed with BsaI enzyme sites at either end. The final sequence of 258bp was synthesised as a fragment (Fig. 4. 3). The features of the consensus promoter are: BsaI enzymes are at nt 6 through 11 and nt 247 through 242, with GGAC and AATG overhangs for Golden Gate assembly as a PU module. A UAS (TGT-N_{10}-ACA) is at nt 42 through 57, an RpoN binding site (TGGCACG-N_4-TTG) is at nt 144 through 157 and an RBS (GAAGGAA) is at nt 225 through 231 with a start codon ATG at nt 238 (Fig. 4. 3).

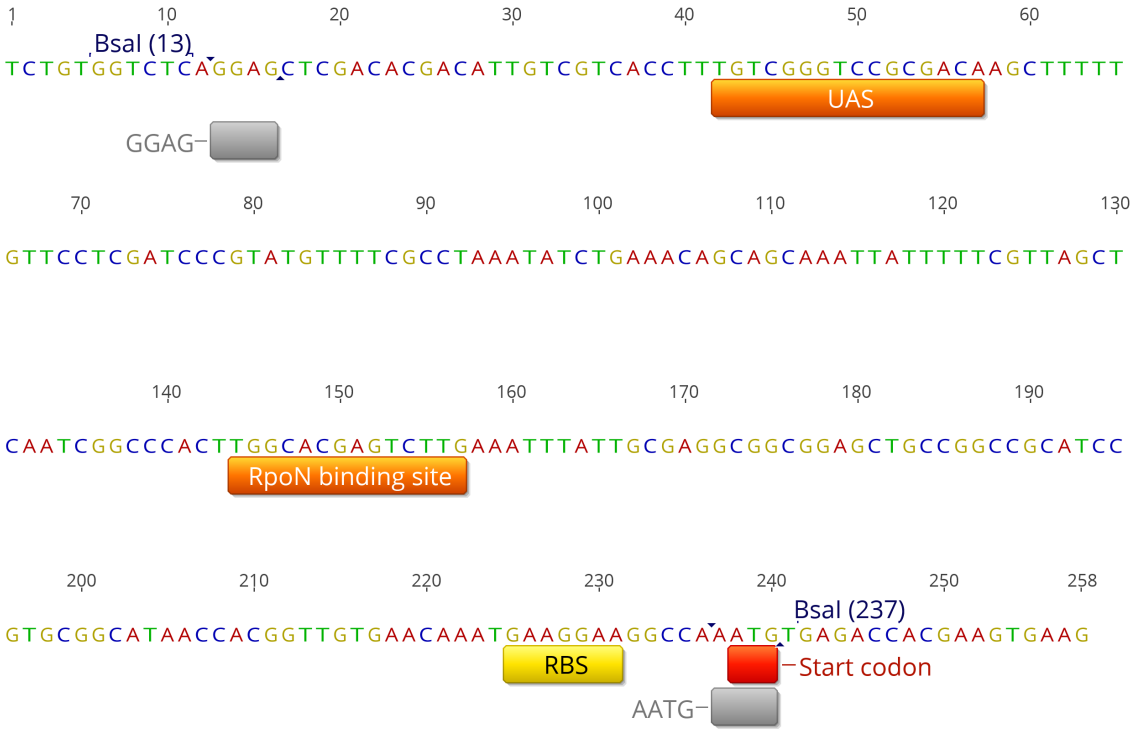


Figure 4. 3 psNifH synthesised as a fragment with terminal BsaI restriction sites for Golden Gate assembly as a PU module. The following features are indicated UAS (orange), RpoN binding site (orange) and RBS (yellow), with a start codon ATG (red).

Assessment of the performance of psNifH to drive strong, nodule-specific expression across Rhizobial strains

A reporter plasmid pOPS0261 was constructed to demonstrate the performance of psNifH as the promoter of the *luxCDABE* operon (Fan *et al.*, 2008). Vector pOGG026 was used as it has been shown in Chapter 3 that it is suitable for experiments in environments where there is no antibiotic selection present.

Reporter plasmid pOPS0261 (Fig. 4. 4A) was constructed as a Golden Gate level 1 cloning reaction with T4 DNA ligase and the restriction enzyme BsaI in a one-tube one-step reaction (Engler *et al.*, 2008) with the following modules: vector; pOGG026, PU module; psNifH from pOGG043, SC module; *luxCDABE* operon from pOGG002; T module; T-pharma terminator from pOGG003. Full details for each Golden Gate Cloning module can be found in Table 2. 1 (Chapter 2). The final plasmid was confirmed by restriction digests (using HindIII) and sequenced to confirm the junctions between vector and insert (primers oxp0282 and oxp0283).

A plasmid with no promoter (No P) upstream of *luxCDABE* pJI11268 (Fig. 4. 4B) (Frederix *et al.*, 2014) was used as a negative (-ve) control. Two plasmids, similarly to pJI11268 but with promoters upstream of *luxCDABE* were also used as controls: plasmid pIJ11282 contains a constitutive promoter pNptII for expression of Lux genes in free-living bacteria (Frederix *et al.*, 2014) and plasmid pLMB640 has the *nifH* promoter from Rlv3841 (Rlv3841pNifH) cloned in front of the Lux genes (expression is nodule-specific). Plasmids pJI11268, pIJ11282 and pLMB640 are stable during symbiosis as they are based on plasmid pJP2 (Prell *et al.*, 2002).

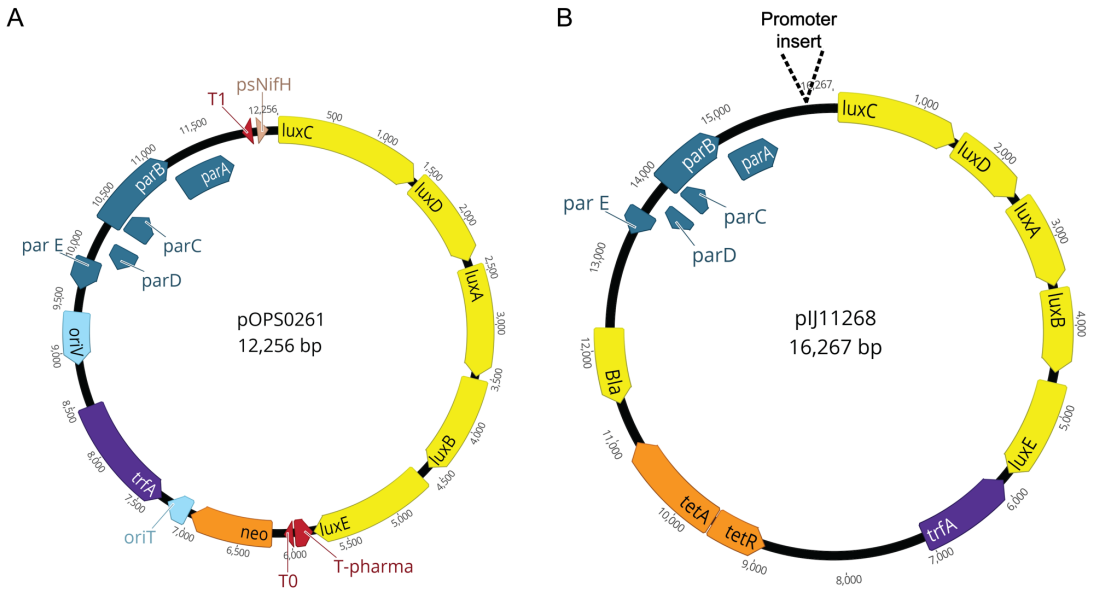


Figure 4. 4 Schematic maps of A) plasmid pOPS0261 and B) -ve control pIJ11268. pOPS0261 was constructed using a modular system of minimal parts. For pIJ11268 is lacking a promoter upstream of the reporter genes, the area where a promoter for the expression of Lux genes can be inserted in pIJ11268 is indicated. pIJ11282 contains a promoter pNptII and pLMB640 contains the nodule-specific promoters Rlv3841pNifH.

Plasmids pOPS0261, pJI11268, pIJ11282 and pLMB640 were conjugated into eleven rhizobial strains (Table 4. 1) and used to inoculate their corresponding legume host. One typical example of the nodules formed on each legume host is shown in Figure 4. 5.

Our hypothesis is that strains will not express Lux when they contain the plasmid lacking a promoter upstream of the reporter (-ve control), there will be low expression with the plasmid containing the promoter pNptII (+ve control) but that they will show medium or high Lux expression (the extent of expression depending on the rhizobial

host strain) with plasmids containing the nodule-specific promoters Rlv3841pNifH (+ve control) and psNifH. To examine *lux* expression in the nodules formed, alfalfa, pea and white clover plants were harvested 21 dpi while those of common bean and soybean were harvested at 42 dpi. NightOWL images were used to assess expression of the *luxCDABE* operon in nodules on plant roots (Fig. 4. 5). For all plants inoculated by bacteria with the -ve control (No P), no lux expression was detected (Fig 4. 5 (i)). Constitutive expression from the pNptII promoter gave low levels of Lux compared to those from the Rlv3841pNifH promoter, with the exception of plants inoculated with Rl bv. *trifolii* WSM1325 (WSM1325) (Fig 4. 5C) and *S. meliloti* strains (data not shown) which showed no expression. All plants inoculated with strains containing psNifH driving lux expression showed no Lux signal and their nodules were small and white across all hosts.

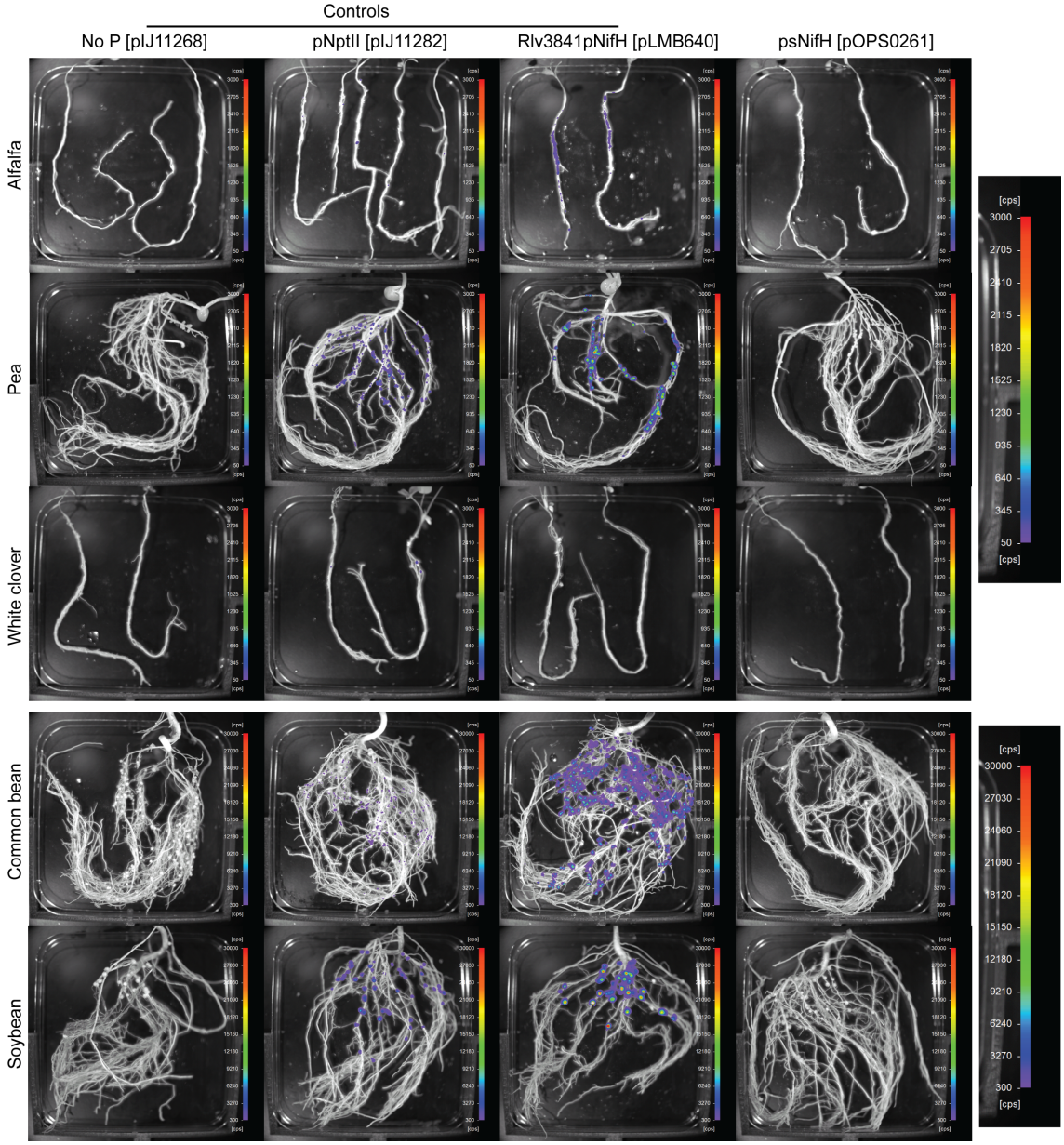


Figure 4. 5 Rhizobial strains expressing *luxCDABE* operon under the control of (i) No P [pJI11268], (ii) constitutive pNptII [pIJ11282], (iii) nodule-inducible Rlv3841pNifH [pLMB640] (iv) synthetic consensus nodule-inducible psNifH [pOPS0261] inoculated onto host plants: (A) alfalfa inoculated with *S. medicae* WSM419, (B) pea inoculated with *Rl* bv. *vicia* A34 RlvA34, (C) white clover inoculated with *Rl* bv. trifolii WSM1325, all harvested at 21 dpi (scale, 50-3,000 cps), (D) common bean inoculated with *Rl* bv. *phaseoli* 4292 and (E) soybean inoculated with *S. fredii* HH103, both harvested after 42 dpi (scale, 300-30,000 cps). The images shown are representative of all legume hosts inoculated with rhizobial strains described in Table 4. 1.

With the aim of analysing how well psNifH was working as a promoter of the *nifH* gene across rhizobial strains, a qualitative assessment of psNifH performance was obtained using reporter plasmids in eleven rhizobial strains inoculated onto five legume hosts. Since it is not possible to make a valid comparison of reporter plasmid expression between hosts, because there are indeterminate and determinate nodules and also different plant sizes, promoter performance was classified within strains sharing the same plant host. For practical proposes four classifications of reporter plasmid expression were used: (-) none observed, (+) low, (++) medium and (+++) high. The entire qualitative assessment of reporter performance is found in Table 4. 1

Table 4. 1 Summary of the qualitative assessment of promoter performance from observing Lux expression in nodules formed on plants inoculated with rhizobial strains. Classification of reporter plasmid expression: (-) none observed, (+) low, (++) medium and (+++) high. *Images of nodules formed by these strains are shown in Figure 4. 5.

Legume host	Strain	No P	pNptII	Rlv3841pNifH	psNifH
		pIJ11268	pIJ11282	pLMB640	pOPS0261
Alfalfa	*S. medicae* WSM419*	-	+	++	-
	S. meliloti WSM1022	-	+	-	-
	S. meliloti 1021	-	+	-	-
Pea	*Rl* A34*	-	+	++	-
	Rlv UPM791	-	++	+++	-
	Rlv 3841	-	++	+++	-
White clover	*Rl* bv. *trifolii* WSM1325*	-	+	-	-
Common bean	*Rl* bv. *phaseoli* 4292*	-	+	+++	-
	R. tropici CIAT 899	-	++	+++	-
	R. etli CFN42	-	+	++	-
Soybean	*S. fredii* HH103*	-	++	+++	-

While the majority of plants inoculated with strains containing promoters pNptII and Rlv3841pNifH driving *lux* expression showed luminescent nodules, all strains with psNifH failed to express luminescence and indeed nodules failed to reach a mature state in all legume hosts. To verify that these plants were not contaminated with a non-symbiotic strain, at least three randomly selected nodules were crushed and the isolated bacteria were screened for antibiotic resistance. All colonies showed neomycin resistance, consistent with pOGG026.

Since the psNifH only consists of a minimal 258 bp region of the *nifH* promoter compared to 717 bp in the full Rlv3841pNifH region it could be that some part was missed in psNifH that is required for a functional promoter. To eliminate this option, plasmids pOPS0262 and pOPS0263 (maps shown in Fig. 4. 6) were constructed as a Golden Gate level 1 cloning reaction with T4 DNA ligase and the restriction enzyme BsaI in a one-tube one-step reaction (Engler *et al.*, 2008) with the following modules: vector; pOGG026, PU module; psNifH from pOGG043; T module; rrnBT1 terminator (T-pharma) from pOGG003, just changing SC module; *celB* from pOGG050 to build pOPS0262 and *gusA* from pOGG083 to build pOPS0263. The final plasmids were confirmed by restriction digests (using EcoRI) and sequenced to confirm the junctions between vector and insert (primers oxp0282 and oxp0283). Plasmids pOPS0262 and pOPS0263 were conjugated into the same eleven rhizobial strains (Table 4. 1) tested before and used to inoculate their corresponding legume host. To assess psNifH driving expression of *celB* and *gusA* the root systems of plants inoculated with relevant stains were stained with X-gal or Magenta-glcA (Fig. 4. 7).

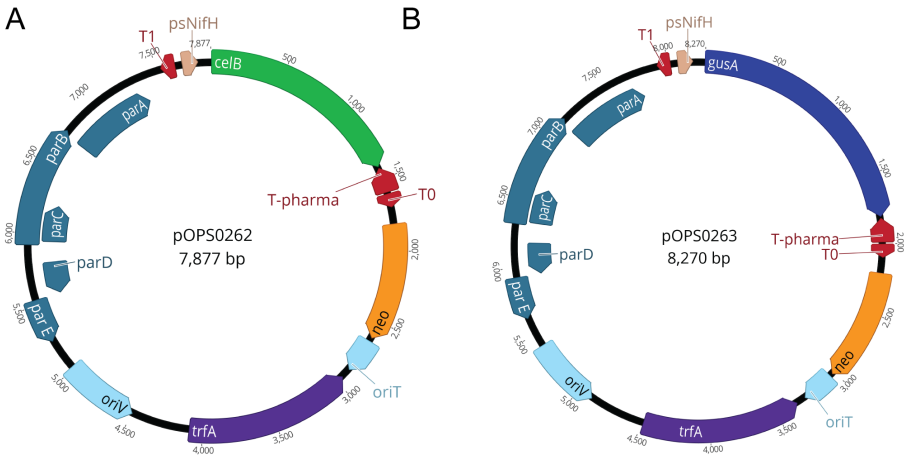


Figure 4. 6 Schematic maps of pOPS0262 and pOPS0263. Promoter psNifH drives expression of *celB* in pOPS0262 and of *gusA* in pOPS0263.

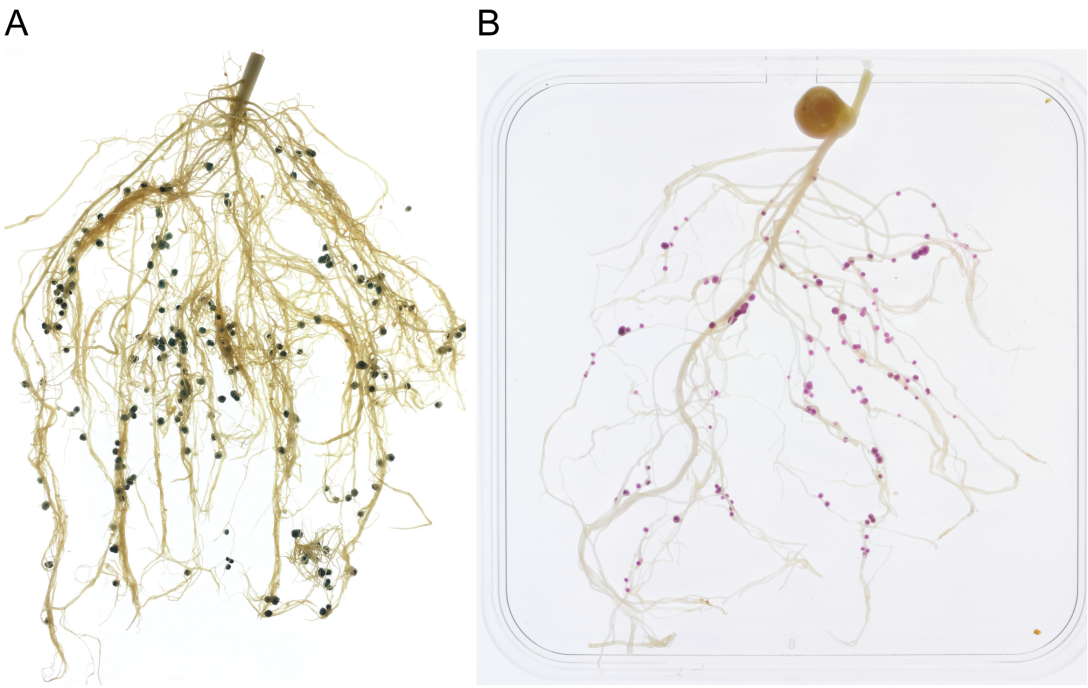


Figure 4. 7 Assessment of the ability of synthetic consensus promoter psNifH to drive *celB* and *gusA* reporter gene expression in nodules formed by Rhizobial strains. A) Common bean root system inoculated with CIAT899[pOPS0262] (*celB* under the control of psNifH) harvested 42 dpi, stained with X-gal after thermal treatment. B) Pea root system inoculated with Rlv3841[pOPS0263] (*gusA* under control of psNifH) harvested 21 dpi, stained with Magenta-glcA. The images shown are representative of legume hosts inoculated with rhizobial strains described in Table 4. 1.

The entire qualitative assessment of reporter performance is found in Table 4. 2. In this experiment, the psNifH promoter driving reporter expression gave strong staining in nodules on all of the plant hosts, with the exception of those inoculated with WSM1325 and *S. meliloti* strains where there was no expression of either reporter. Similar results obtained with the control pLMB640 (Table 4. 1).

Table 4. 2. Summary of the qualitative assessment of promoter performance from observing CelB and GusA expression in nodules formed on plants inoculated with rhizobial strains. Classification of reporter plasmid expression: (-) none observed, (+) low, (++) medium and (+++) high. *Images of nodules formed by these strains are shown in Figure 4. 7.

Legume host	Strain	psNifH	
		celB	***gusA***
		pOPS0262	**pOPS0263**
Alfalfa	*S. medicae* WSM419	++	++
	S. meliloti WSM1022	-	-
	S. meliloti 1021	-	-
Pea	*Rl* A34	++	++
	Rlv UPM791	+++	+++
	Rlv 3841*	+++	+++
White clover	*Rl* bv. *trifolii* WSM1325	-	-
Common bean	*Rl* bv. *phaseoli* 4292	+++	+++
	*R. tropici CIAT 899**	+++	+++
	R. etli CFN42	++	++
Soybean	*S. fredii* HH103	+++	+++

The *S. meliloti* strains did not show expression since they have different pNifH promoter than *R. leguminosarum* and to the other strains used, this is not unexpected. Expression of both these reporters was observed in all *R. leguminosarum* strains, with the exception of WSM1325 (Table 4. 2). WSM1325 is not a typical *R. leguminosarum*

trifolii strain and it has previously been shown by Melino et al., (2012) that the intergenic region of *nifH* in WSM1325 is different to other *R. leguminosarum* strains (i.e. *R. l. trifolii* WSM2304).

The results of psNifH in reporter plasmids expressing *celB* and *gusA* were sufficient to conclude that this promoter can be used in *Rhizobium leguminosarum* strains.

The advantages of using marker genes *celB* and *gusA* for monitoring strains in symbiosis are that they are very reliable, simple and inexpensive, making them easy for use in almost any laboratory (Sessitsch *et al.*, 1997, 1998; Sánchez-Cañizares & Palacios, 2013). The disadvantages are that with these marker genes it is not possible to recover living bacteria from inside the nodule, since a harsh hypochlorite fixation and enzymatic reaction is necessary to distinguish the nature of nodule occupancy, and with the *gusA* and *celB* system it is only possible to assess two strains at the same time.

Since in this project, is very important to have a high-throughput screening, we decided to test an additional control using the promoter of Rlv3841pNifH that has been used previously and worked perfectly in pOGG026 (Chapter 3). Plasmid pOPS0252, was assembled as a Golden Gate level 1 cloning reaction with T4 DNA ligase and the restriction enzyme BsaI in a one-tube one-step reaction (Engler *et al.*, 2008) with the following modules: vector; pOGG026, PU module; Rlv3841pNifH from pOGG082, SC module; *luxCDABE* operon from pOGG002; T module; T-pharma terminator from pOGG003. The final plasmid was confirmed by restriction digests (using HindIII) and sequenced to confirm the junctions between vector and insert (primers oxp0282 and oxp0283).

A new plant assay was performed inoculating pea plants with Rlv3841[pOPS0252] and harvested at 28 dpi. In the case of the expression of the *luxCDABE* under control of Rlv3841pNifH, resulting nodules were white and appeared to be still immature at 28 dpi (Fig. 4. 8). This result suggests that it is the use of the *luxCDABE* operon cloned in pOGG026 and high expression of Lux in the nodule environment that leads to poor nodule development and lack of luminescence, and not the psNifH promoter, *per se,* as psNifH has been shown to work with CelB and GusA reporters.

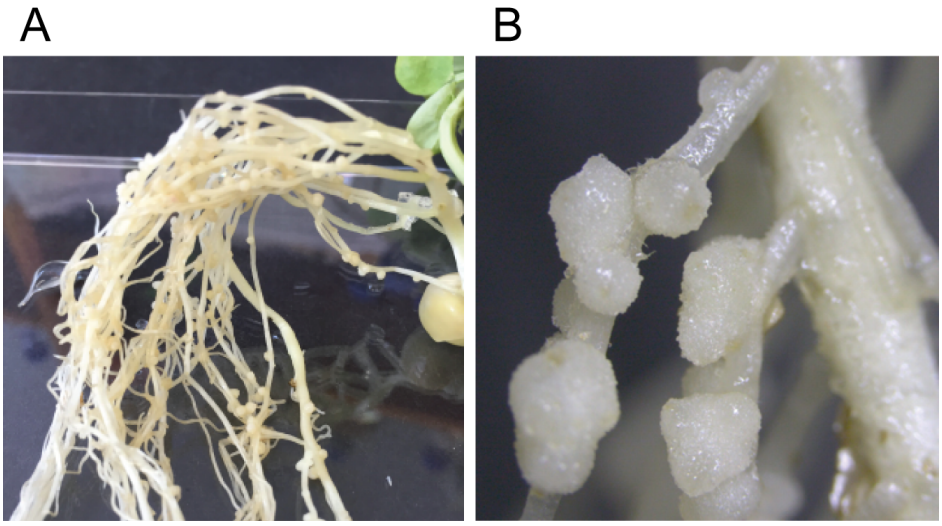


Figure 4. 8 Pea root system harvested at 28 dpi showing nodules formed by Rlv3841[pOPS0252], cloned *luxCDABE* operon under control of Rlv3841pNifH. A) Full root system shows immature nodules B) Image with dissecting microscope shows small, white nodules, normally characteristic of ineffective nodules.

Since bioluminescence reactions require energy from cell metabolism (Hakkila *et al.*, 2002), it is possible that the *lux* operon could not be expressed, even when using the previously successful promoter Rlv3841pNifH, because the energy required by it was too high, and the rhizobia started to die in the nodules. As *lux* is only expressed in

plasmid pOPS0252 at the stage when nitrogen fixation is carried out (i.e. in mature nodules), it is expected that the *lux* operon would not be a problem in the first stages of rhizobial establishment inside the legume. Strains are therefore able to enter the roots, infect the plant and make nodules without any problem. It is only when *nifH* starts to be transcribed that the *lux* operon starts to be expressed that the bacteria die. This did not happen when the *lux* operon is expressed from plasmid pLMB640, possibly because the plasmid copy number, and consequently expression of *lux* was lower in pLMB640 due to the greater size of vector pJP2 (13.1 kb) compared to the vector pOGG026 (6.6 kb) which was used in the construction of pOPS0261 and pOPS0252.

Based on these results, we decided to test other marker genes that can be used in high-throughput screenings. In this case, superfolder green (sfGFP) (Pédelacq *et al.*, 2006) and red (mCherry) (Shaner *et al.*, 2004) fluorescent proteins were chosen since the size of the inserts were considerably smaller than the *luxCDABE* operon and because, unlike with bioluminescence, fluorescent proteins do not require energy from cell metabolism nor do they require additional gene products in order to be expressed (Chalfie *et al.*, 1994; Hakkila *et al.*, 2002). We can therefore assume they are less energy demanding.

New reporter plasmids placing these reporters under the control of psNifH were built as a Golden Gate level 1 cloning reaction with T4 DNA ligase and the restriction enzyme BsaI in a one-tube one-step reaction (Engler *et al.*, 2008) with the following modules: vector; pOGG026, PU module; psNifH from pOGG043; T module; T-pharma

terminator from pOGG003, SC module; sfGFP from pOGG037 to build pOPS0381 (Fig.4.9 A) and SC module; mCherry from EC15071 to build pOPS0382 (Fig. 4.9 B).

Following the same strategy additionally plasmids were constructed to be used as controls. Plasmids pOPS0379 (sfGFP) and pOPS0380 (mCherry) placing reporters under the control of Rlv3841pNifH were used as positive controls for nodule-specific expression. Plasmids pOPS0383 (sfGFP) and pOPS0384 (mCherry), which lacked a promoter upstream of the reporter (No P), were used as negative controls.

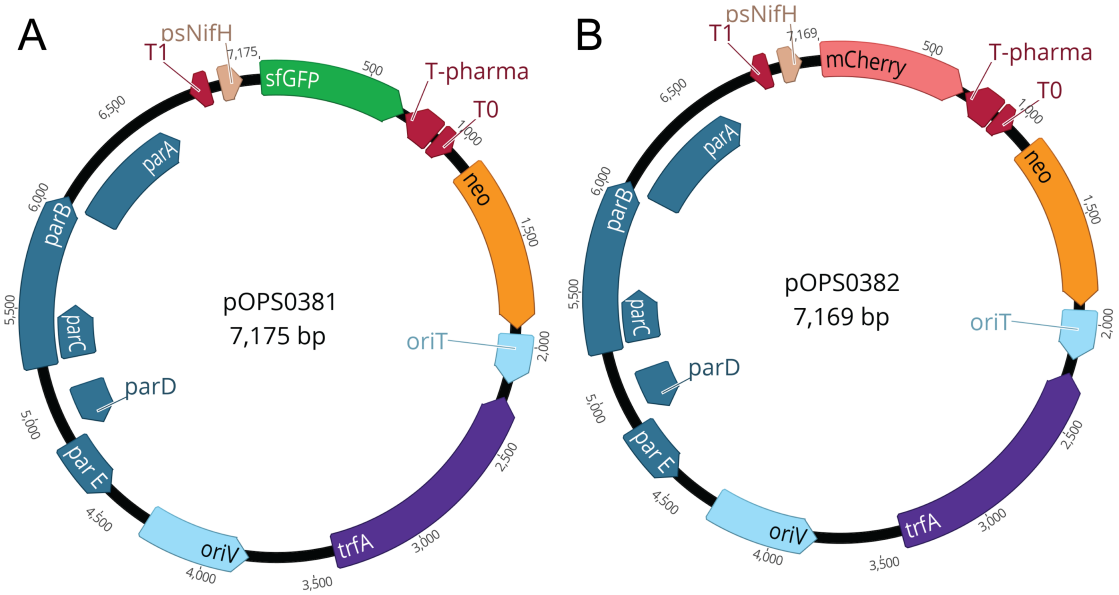


Figure 4. 9 Schematic maps of pOPS0381 and pOPS0382 plasmids. Promoter psNifH drives expression of sfGFP in pOPS0381, and of mCherry in pOPS0382.

The reporter plasmids described above were conjugated into Rlv3841 and inoculated onto pea plants. Their root systems (harvested at 21 dpi) were analysed with the NightOWL CCD camera to quantify green and red fluorescence. Expression of both sfGFP and mCherry was observed in the +ve control and with psNifH (shown in Fig. 4.

10). Expression was nodule-specific and without an apparent fitness penalty. There was no expression of florescence proteins in either -ve control.

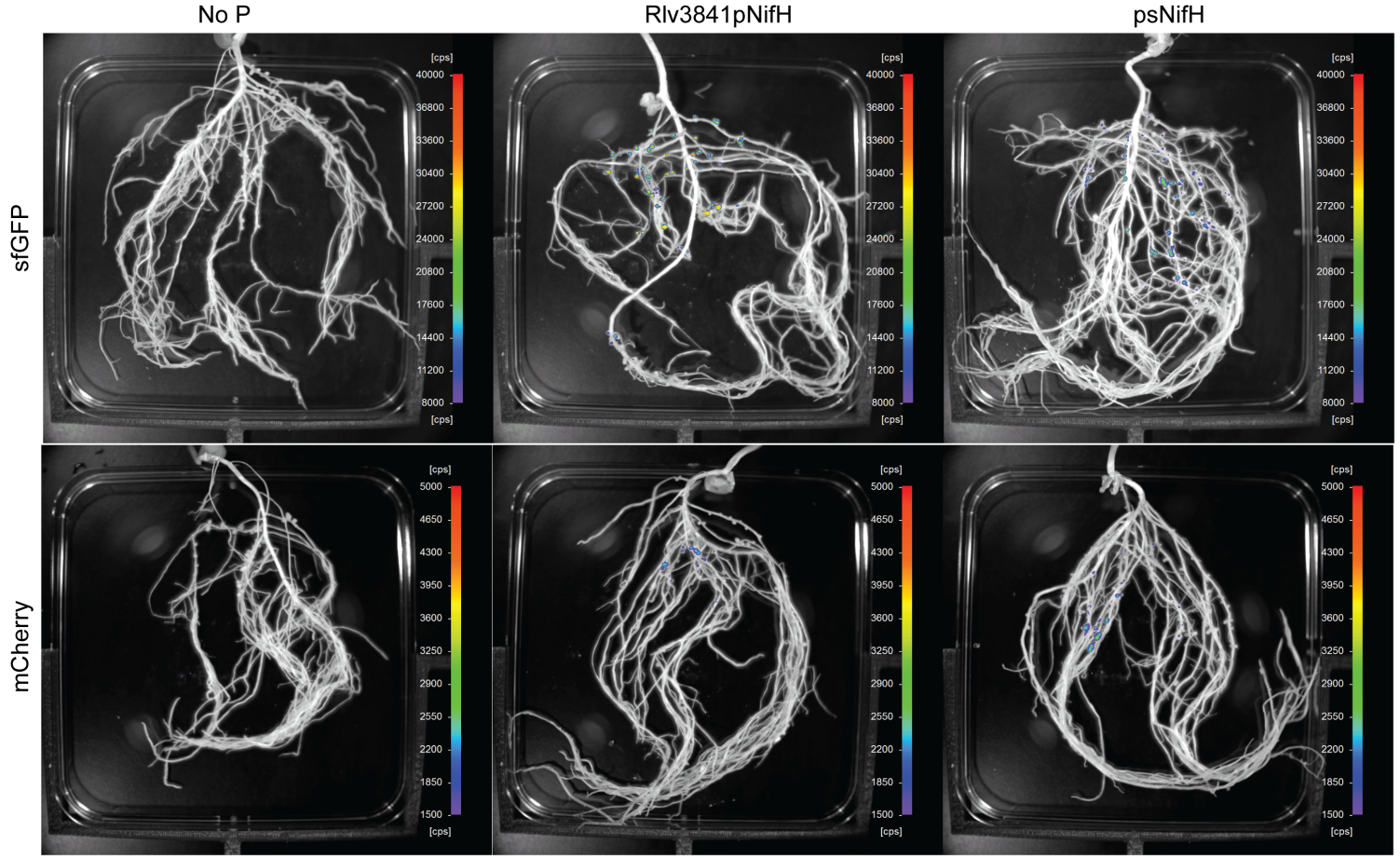


Figure 4. 10 Pea root system from plants harvested at 21 dpi. Green fluorescence from Rlv3841 expressing GFP under the control of i) No P Rlv3841[pOPS0383] (-ve control), ii) Rlv3841[pOPS0379] (+ve control) and iii) Rlv3841[pOPS0381] (psNifH) (scale, 8,000-40,000 cps). Rlv3841 expressing mCherry under the control of i) No P Rlv3841[pOPS0384] (-ve control), ii) Rlv3841[pOPS0380] (+ve control) and iii) Rlv3841[pOPS0382] (psNifH) (scale, 1,500-5,000 cps). Images shown are representative of four biological replicates.

Since the expression of sfGFP was almost ten times greater than mCherry, it was decided to continue working only with sfGFP for future experiments.

Using psNifH as promoter of sfGFP not only allowed us to use high-throughput screening but was also shown to be an excellent reporter proxy of nitrogenase activity. Figure 4. 11 shows pea nodules formed by Rlv3841[pOPS0381] (sfGFP under control of psNifH). Fluorescence is detected only when these pea nodules reach a mature stage (i.e. no expression in immature nodules Fig. 4. 11A), with strong expression in mature nodules (Fig 4. 11B). In addition, sfGFP expression is solely found in the nitrogen-fixing zone (Fig. 4. 11B-C). Fluorescent proteins do not need cofactors or substrates and can also be fluorescent in non-viable cells (Chalfie *et al.*, 1994). Additionally sfGFP has a greater resistance to chemical denaturants (Pédelacq *et al.*, 2006). This could have been problematic since we might have had GFP fluorescence in senescent areas of nodules via the accumulation of the fluorescent protein. However, senescent areas of nodules do not contain GFP fluorescence (Fig. 4. 11C).

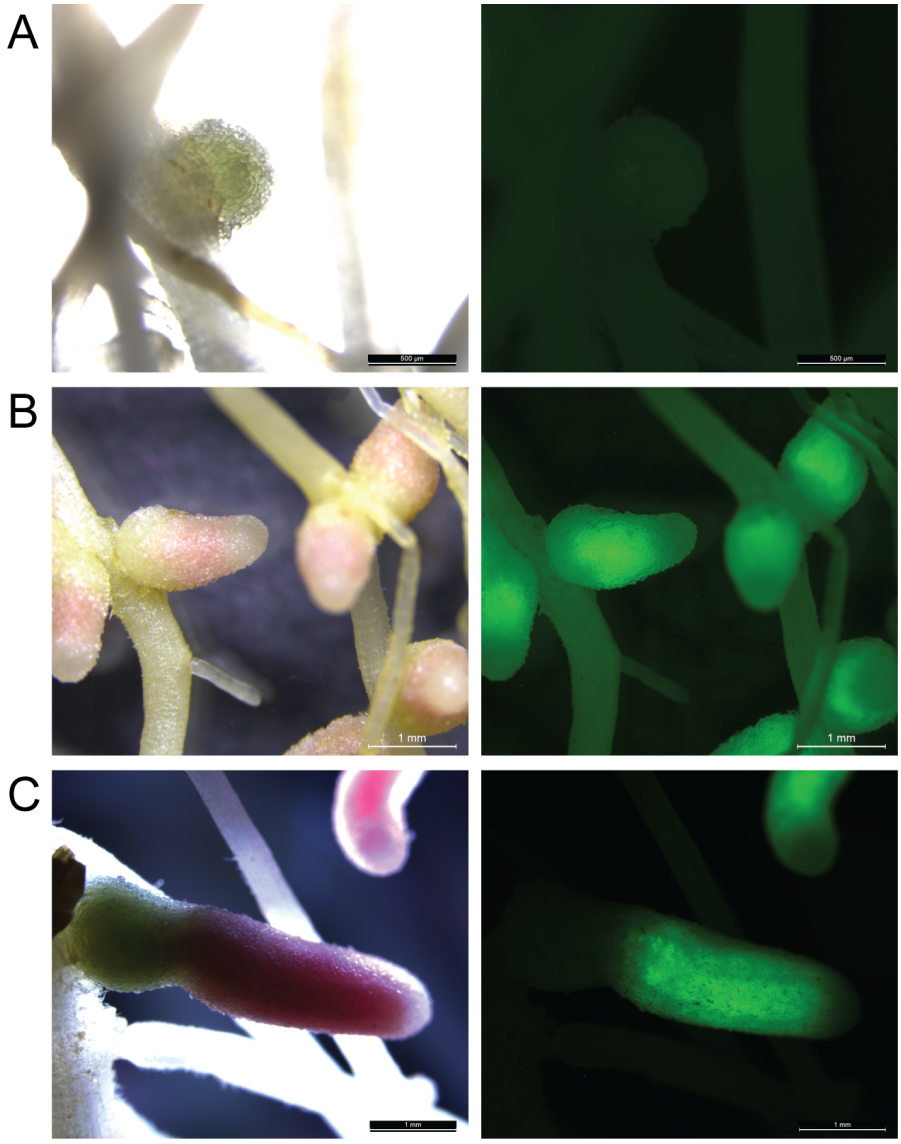


Figure 4. 11 Dissecting microscope images of pea nodules formed by Rlv3841[pOPS0381] which has a plasmid containing sfGFP under control of psNifH. A) immature nodule, 21 dpi. B) mature nodule, 28 dpi. C) senescent nodule, 42 dpi with the senescent area furthest from the tip. Bright-field (left) and fluorescence (GFP filter) (right) images.

This same GFP-expressing strain, Rlv3841[pOPS0381], was used to visualise and identify individual bacteroids from a crushed pea root nodule (harvested at 42 dpi)

using confocal microscopy (Fig. 4. 12A-D). The 3D shape of bacteroids was visualised using their GFP expression (Fig. 4. 12E-F) following the methodology of Barbier de Reuille et al., (2015).

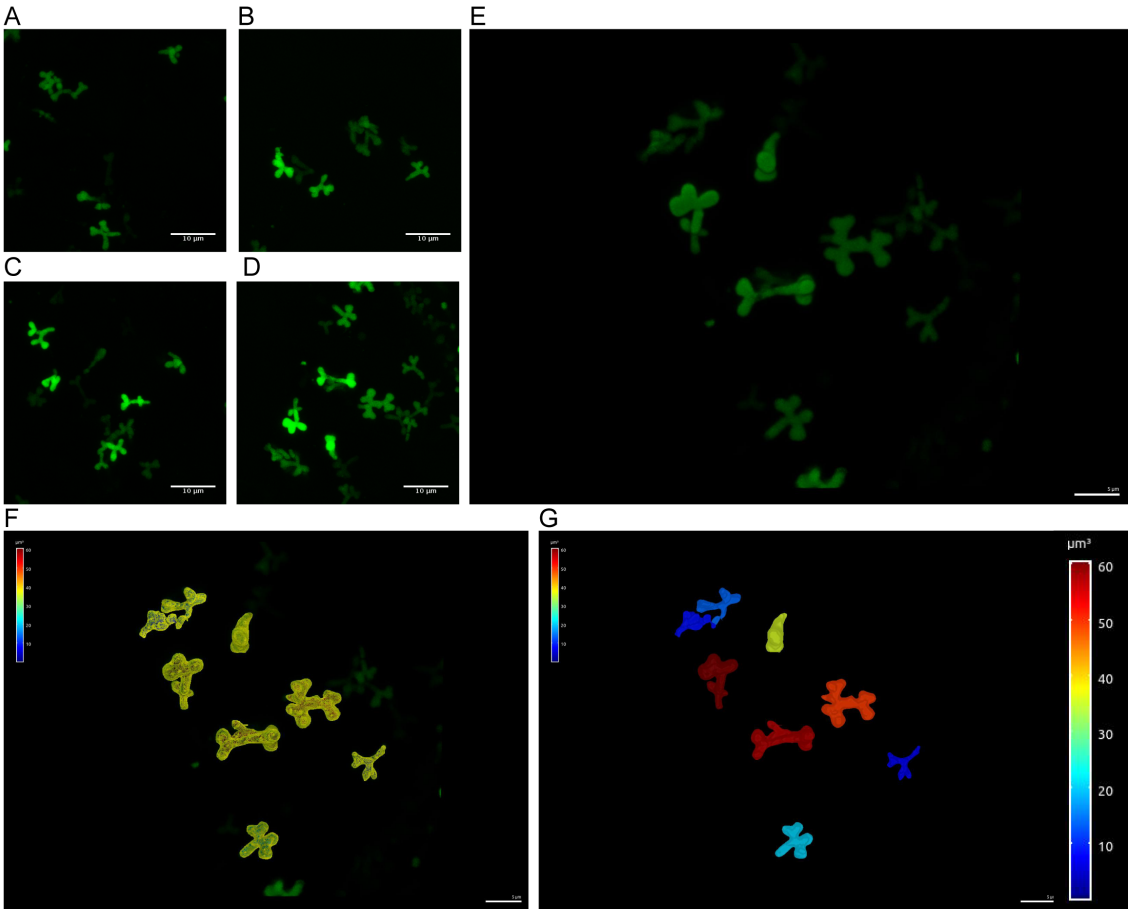


Figure 4. 12 A-D) Confocal images of bacteroids formed by Rlv3841[pOPS0381] (sfGFP under control of psNifH) from crushed pea root nodules harvested at 42 dpi. GFP excited at 488 nm. E-G) 3D segmentation. E) GFP TIF image stack loaded into MorphoGraphX. F) Image converted into a binary stack and meshes generated at a cube spacing of 0.1 µm (3x smoothing). G) Volume colour codes; scale from 0 to 60 µm^3. In collaboration with: Dr Charlotte Kirchhelle.

Bacteroids in temperate legumes are very commonly described in the literature as having an elongated shape that can change into a characteristic Y shape (Oke & Long, 1999; Gage, 2004; Fan *et al.*, 2014; Farkas *et al.*, 2014), but this assumption is a consequence of 2D observation. The combination of this technique with this reporter plasmid under the control of psNifH gave us the opportunity to observe very irregular shapes in individual bacteroids (Fig. 4. 12F) and also to assess and classify individual bacteroids by their volume. In Figure 4. 12G we can observe individual bacteroids, colour-coded according to their volume, which ranged from approx. 10 to 60 μm^3.

With these results, it is clear that pea bacteroids should be described as coralloid rather than T or Y shaped.

Future research could use this technique and this class of reporter plasmids to answer additional biological questions, such as a comparison of bacteriod packing in determinate and indeterminate nodules or the assessment of sanctioning in mixed nodules at the plant cell and bacteriod level.

Assessment of the consensus synthetic promoter psNifH in competition for nodule formation assays using isogenic marked and unmarked rhizobial strains

To test whether the presence of a cloned reporter gene under the control of the consensus synthetic promoter psNifH affects the competitiveness of strains with respect to forming nodules on their host plant, we performed competition assays. To do this we used three different co-inoculation ratios; 10, 50 and 90% of marked versus unmarked strains. In the initial experiment, pea plants were co-inoculated with three biological replicates of two isogenic strains: strain Rlv3841[pOPS0263] marked with

gusA under the control of psNifH and unmarked Rlv3841 (WT). At 21 dpi, pea root systems were stained with Magenta-glcA (Fig. 4. 13 A-C) and the number of stained (formed by the marked strain) and unstained (from the unmarked strain) nodules were counted. In a second experiment, this time analysing five biological replicates, pea plants were co-inoculated with two isogenic strains: Rlv3841[pOPS0381] marked with sfGFP under control of psNifH, and unmarked Rlv3841 (WT). At 21 dpi, pea root systems were analysed with the NightOWL CCD camera accompanied by blue-light transilluminator to detect green fluorescence in the nodules (Fig. 4. 13E-G). Those expressing GFP were detected and counted automatically using Indigo® software.

If these reporter plasmids do not reduce their host strain's competitiveness, then we should observe an equal relationship between the percentage of marked nodules (pink or GFP fluorescent) and the percentage of cells from marked strains (*gusA* and sfGFP under the control of psNifH) in the mix inoculum, where a linear regression of the slope = 1 and the line should pass through the origin.

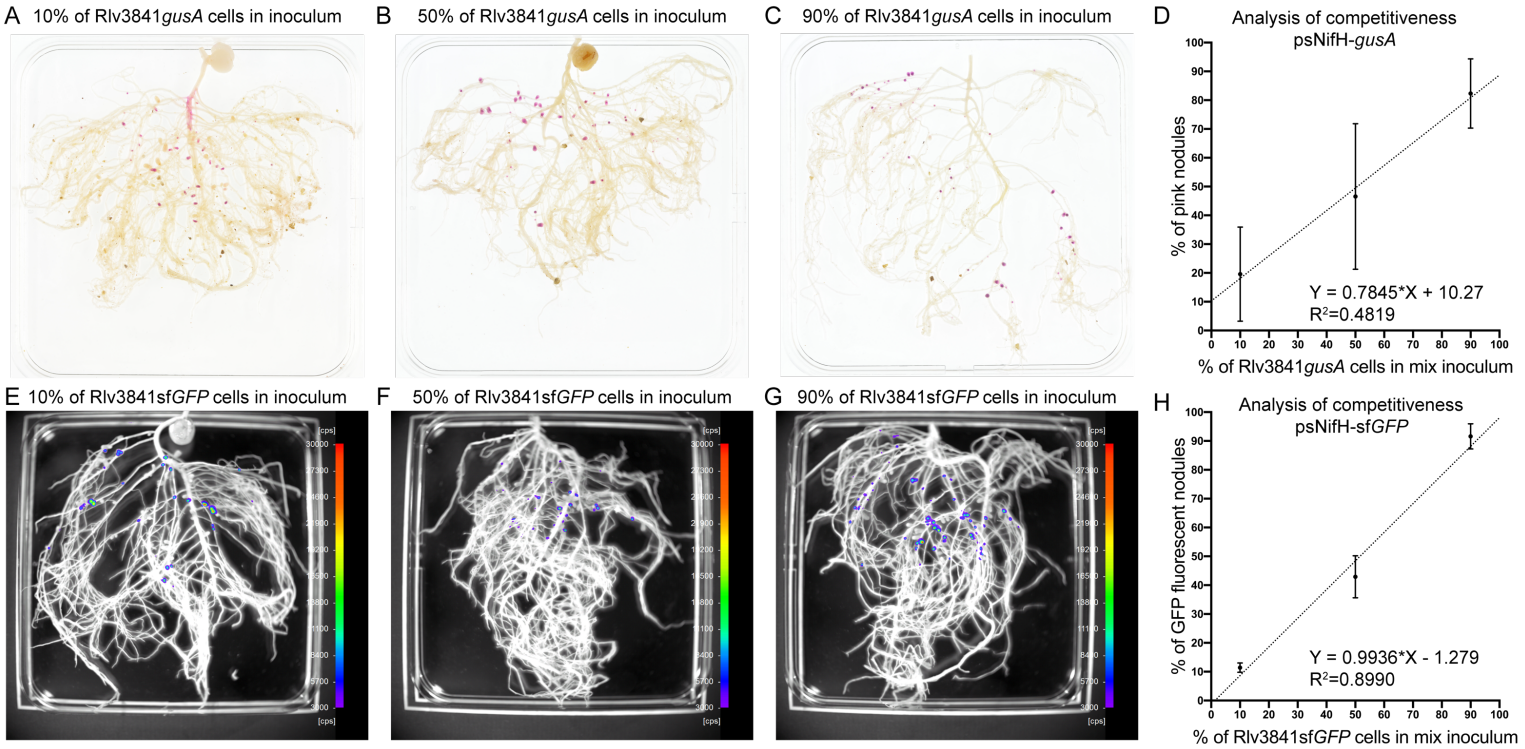


Figure 4. 13 Competition assays for nodule formation between marked and unmarked strains on pea roots. A-C) Following staining with Magenta-glcA, pink nodules formed by Rlv3841[pOPS0263] (*gusA* gene under control of psNifH) and white nodules formed by Rlv3841 (WT) from inoculation at the following ratios of Rlv3841(*gusA*); A) 10%, B) 50% and C) 90% in mix inoculum. E-G) GFP detection (scale, 3,000–30,000 cps) in nodules from Rlv3841[pOPS0381] (sfGFP under control of psNifH), white nodules formed by Rlv3841 (WT) from inoculation at the following ratios of Rlv3841[sfGFP]; E) 10%, F) 50% and G) 90% in mix inoculum. D-H) Analysis of competitiveness was performed by comparing % of marked nodules and % of marked strain cells in mix inoculum. Results from D) Rlv3841-psNifH-*gusA* with three independent biological replicates, and H) Rlv3841-psNifH-sfGFP with five independent biological replicates, where bars are the SEM, showed that marked nodules were not significantly different from unmarked nodules.

In both experiments, the percentage of nodules from marked strains, either *gusA* or sfGFP, reflected the percentage of marked cells in the inoculum. In the case of nodules marked with *gusA* (Fig. 4. 13D) the linear regression of the slope was 0.7845 ± 0.3074 (95% CI [0.05755 to 1.511]) and this is not significantly different from 1.0 (*p*=0.5059). The Y-intercept = 10.27 is not significantly different from 0 (*p*=0.5935) but with a low R^2=0.4819. In the case of nodules marked with sfGFP (Fig. 4. 13H) the linear regression of the slope was 0.9936 ± 0.09614 (95% CI [0.7841 to 1.203]) and this is not significantly different from 1.0 (*p*=0.9476). The Y-intercept = -1.279 is not significantly different from 0 (*p*=0.8192). In this case we obtained a R^2= 0.8990.

Although having a marked or unmarked strain had not significantly impacted on the proportion of nodules formed, we can observe that results from GFP fluorescent nodules were more accurate since they were detected and counted automatically using Indigo® software compared to the pink nodules that were counted manually. Additionally, competition assays using sfGFP are cheaper and require less time to obtain the results compared to using *gusA*, because staining with Magenta-glcA is not required

Detection of sfGFP under control of synthetic consensus promoter psNifH as a proxy for nitrogenase activity assays

To test if nitrogenase activity correlates with GFP expression under the psNifH promoter we needed to select *R. leguminosarum* strains with different levels of effectiveness at fixing nitrogen. A collection of 17 *R. leguminosarum* strains was obtained from Professor John Howieson from the Centre for Rhizobium Studies (CRS),

Perth, Western Australia. We evaluated the effectiveness of these strains by acetylene reduction assays (ARA) (Hardy *et al.*, 1968) and placed them into broad classes (Table S.1 in Supplemental Material). The best candidates to test in an experiment to see if GFP expression could be used as a proxy for nitrogenase activity were strains that showed the lowest and highest ARAs in our evaluations, which were WSM1475 and WSM1521, which were also those identified in the lowest and highest effectiveness class by analysis of dry weight biomass (using method of Howieson *et al.*, 1995)[4]. Additionally, Rlv3841 and UPM791 were added as host backgrounds since they are well-studied strains in our lab.

These four strains were conjugated with reporter plasmids containing a cloned sfGFP under the control of psNifH WSM1475[pOPS0515], WSM1521[pOPS0504], Rlv3841[pOPS0503] and UPM791[pOPS0491]. Each of these reporter plasmids are almost identical to pOPS0381 the only difference is the inclusion of a plasmid ID of 43 bp upstream of sfGFP replacing the terminator (T module) T-pharma (Fig. 4. 14A-B). In this experiment no use is made of the plasmid id and the plasmids are for all intents and purposes identical to pOPS0381. The purpose of this plasmid ID will be explained in greater detail in Chapter 5.

[4] Information of effectiveness class by email communication with Professor John Howieson

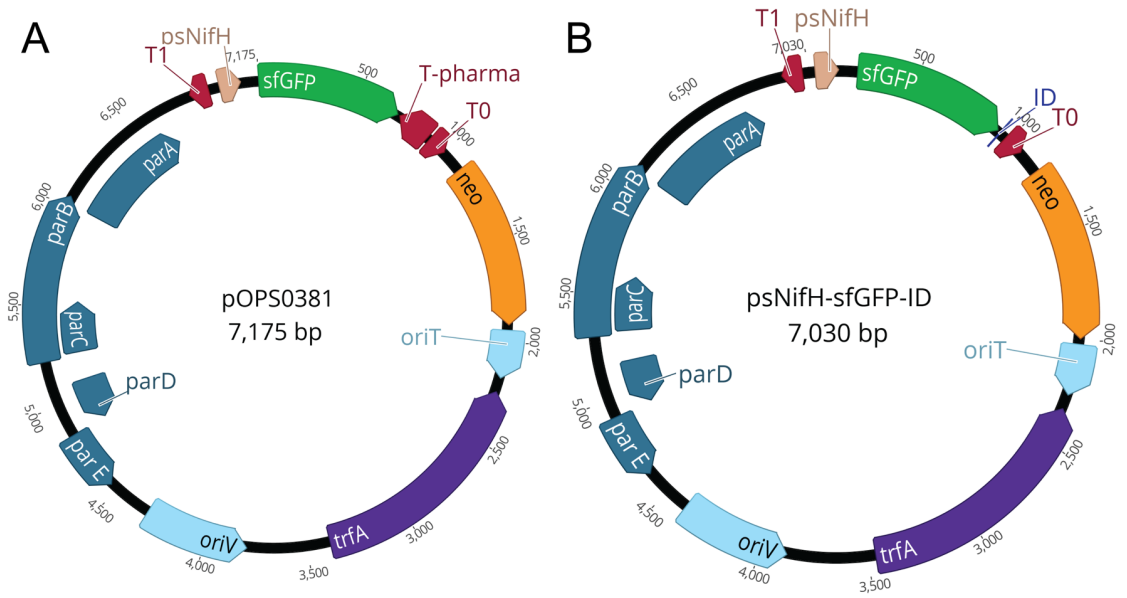


Figure 4.14. Schematic maps of pOPS0381 and psNifH-sf-ID plasmids (pOPS0491, pOPS0503, pOPS0504 and pOPS0515) which containing a unique plasmid ID (12 nt) instead of the T-pharma (T module for Golden Gate assembly).

The plasmids pOPS0491, pOPS0503, pOPS0504 and pOPS0515 differ by only a 12 nt sequence. At this moment we will assume they will perform in the same way as each other and pOPS0381 but more rigorous analysis of plasmid stability and gene expression assays is required. Strains containing plasmid pOPS0383 (lacking a promoter upstream of sfGFP) were used as negative controls.

Analysis of inoculated pea plants was made at three different time points: 28, 35 and 42dpi. Green fluorescence expression from nodules and nodule area were determined using the NightOWL CCD camera. Nitrogenase activity was determined by acetylene reduction assays (ARAs) (Hardy *et al.*, 1968; Lodwig *et al.*, 2003) in terms of number of nodules per plant.

The high effectiveness of strain WSM1521 has been confirmed before by Slattery & Pearce (2002) through glasshouse evaluation of dry matter yields. We have therefore decided to use WSM1521 as a positive control for nitrogenase activity. We expect that detection of GFP fluorescence from WSM1521 nodules should be statistically significantly higher than WSM1475, Rlv3841 and UPM791, and that ARAs at 28 dpi should show the same tendency since this is the time of peak N_2 fixation rate in pea plants, due to flowering point (Allaway & Lodwig, 2000; Mulley *et al.*, 2010). Symbiotic performance of the strains was analysed by collecting shoot dry biomass (DW) at 35 and 42 dpi. Full results are summarised in Table 4. 3.

Table 4. 3 Fluorescence from expression of sfGFP, acetylene reduced and shoot dry weight from nodules of pea plants inoculated with different rhizobial strains, harvested at 28, 35 and 42 dpi.

Strain	Nodule fluorescence (cps)(plant)$^{-1}$	Acetylene reduced (plant)$^{-1}$	Shoot DW (mg) (plant)$^{-1}$	No. of nodules (plant)$^{-1}$	Nodule size (mm²)
28dpi					
WSM1475[pOPS0515]	6.19E+06 ± 1.42E+06**	1189.35 ± 74.78**	ND	141.33 ± 24.92	1.70 ± 0.06*
Rlv3841[pOPS0503]	2.92E+08 ± 1.28E+07*	2758.50 ± 181.65**	ND	182.67 ± 24.88	2.23 ± 0.24
UPM791[pOPS0491]	3.78E+08 ± 5.35E+07*	4613.00 ± 562.91*	ND	276.00 ± 37.75	1.88 ± 0.27
WSM1521[pOPS0504]	5.02E+08 ± 8.67E+07	9600.97 ± 1984.58	ND	176.00 ± 35.12	2.89 ± 0.50
35dpi					
WSM1475[pOPS0515]	1.47E+07 ± 1.57E+06*	2849.84 ± 385.46	333.33 ± 66.67 *	266.67 ± 23.47	1.44 ± 0.09*
Rlv3841[pOPS0503]	1.43E+08 ± 2.67E+07	4531.97 ± 932.35	776.00 ± 167.29*	219.33 ± 44.36	2.16 ± 0.38
UPM791[pOPS0491]	1.55E+08 ± 1.22E+07	5481.30 ± 621.64	916.67 ± 101.38	293.67 ± 23.90	1.93 ± 0.08
WSM1521[pOPS0504]	1.51E+08 ± 5.88E+07	3305.21 ± 525.85	1033.33 ± 185.59	138.00 ± 15.52	3.50 ± 0.74
42dpi					
WSM1475[pOPS0515]	2.01E+07 ± 2.12E+06	ND	800.00 ± 100 **	269.67 ± 18.85	1.67 ± 0.21*
Rlv3841[pOPS0503]	3.67E+08 ± 5.34E+07	ND	850.00 ± 28.87**	295.67 ± 32.63	2.21 ± 0.28
UPM791[pOPS0491]	1.79E+08 ± 7.04E+07	ND	1233.33 ± 185.59*	289.00 ± 41.58	2.58 ± 0.34
WSM1521[pOPS0504]	7.42E+07 ± 3.29E+07	ND	1750.00 ± 132.29	152.00 ± 13.58	2.96 ± 0.26

Nodule fluorescence in counts per second (cps)

Acetylene reduced in nanomoles of C_2H_2 reduced per hour per plant.

Values are the mean from at least three biological replicates ± SEM. In the case of ARA values were obtained from three biological and three technical replicates.

Statistical analysis for CPS, ARA and Shoot DW. Activities for each strain within the same time point were compared by One-way ANOVA (Dunnett's multiple comparisons test) compared to WSM1521 and those that are significantly different are indicated with * ($P<0.05$) and with ** ($P<0.005$).

ND, not determined.

GFP expression, nitrogenase activity and shoot dry weight biomass were normalised by number of nodules per plant. All normalised data points for GFP fluorescence and nitrogenase activity at 28 and 35 dpi were plotted in Fig. 4. 15A. Analysing the correlation of GFP fluorescence and ARA by a Two-tailed Pearson test shows a positive correlation at 28 dpi of r=0.8894 (95% CI [0.6444 to 0.9688]), *P*=0.0001 and a R^2=0.7910. At 35 dpi the positive correlation was r=0.8160 (95% CI [0.4234 to 0.9506]), *P*=0.0022 and a R^2=0.6659. Results from 28 dpi were grouped by strains in figure 4. 15B and biomass collected at 35 and 42 dpi are shown in figure 4. 15C.

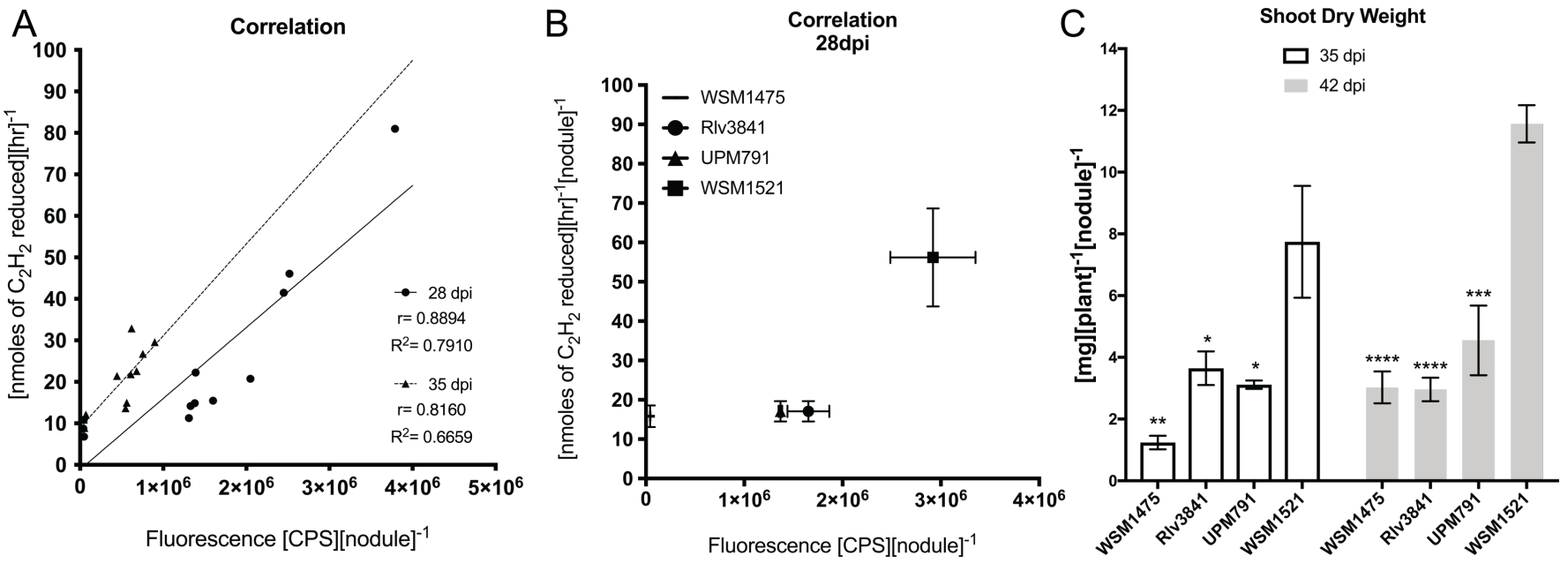


Figure 4.15 GFP fluorescence detection, nitrogenase activity and shoot dry weight biomass are presented after being normalised by the number of nodules on each pea plant. A) Correlation of fluorescence and nitrogenase activity at 28 and 35 dpi, each dot represents a plant. B) Correlation in data grouped per strain from plants at 28 dpi. C) Pea shoot dry weight at 35 and 42 dpi. Each strain within the same time point was compared to WSM1521 by One-way ANOVA using a post-hoc Dunnett's multiple comparisons test. * indicates *P value* <0.05, ** *P*<0.005, * *P*<0.0005, and ****P<0.0001. In all the cases results are from three independent biological and bars represent the SEM.**

At both 28 dpi and 35 dpi there was a statistical correlation (Fig. 4.15A), but the analysis at 28 dpi shows the strongest correlation, agreeing with the literature which recommends ARAs at 28 dpi in the pea plants (Allaway & Lodwig, 2000; Mulley *et al.*, 2010). Therefore, it is correct to consider 28 dpi as the time point for future GFP florescence detection assays. In figure 4.15B we can observe that both GFP fluorescence and ARA correspond to the effectiveness class in pea plants of the strains WSM1521 (high) and WSM1475 (low) previously observed by Professor John Howieson. Strains Rlv3841 and UPM791 are more difficult to pull apart but this is because they have a more similar effectiveness class compared to WSM1521 and WSM1475. This does not invalidate the use of psNifH as a proxy for nitrogenase activity, but it sets the limitation of assay.

Biomass measurements at 35 and 42 dpi (Fig. 4. 15C) support the suggestion that strain WSM1521 is highly effective, corroborating the observations of both nitrogenase activity measured by ARA and GFP florescence detection (Fig 4. 15B). Furthermore, WSM1521 also had the largest nodule size compared to the other three strains at all time points. A positive correlation of nodule size (or nodule mass) with symbiotic effective strains has been publish several times in our field, most recently by (Oono *et al.*, 2011; Melino *et al.*, 2012; Regus *et al.*, 2017).

More accurate measurements of total nitrogenase activity certainly exist, such as flow-through ARA systems and $^{15}N_2$ assays, but the disadvantage of these assays is that they require special equipment (mass spectrometer or emission spectrometer), making it a relatively expensive technique and not accessible for low income

laboratories (Unkovich *et al.*, 2008). As stated in Vessey (1994) and Giller (2001), it is essential to emphasise that static ARA is an indirect method of measuring N_2-fixation and does not calculate the total nitrogenase activity, although it is a very sensitive diagnostic tool and undoubtedly very useful as a rapid method for detecting nitrogenase activity (Giller, 2001; Unkovich *et al.*, 2008). Furthermore it is a good indicator of relative changes of nitrogenase activity if the plants to be assessed are grown in the same conditions and at the same time, and if the level of disturbance is minimised by retaining the whole nodulated root system and shoot (Minchin *et al.*, 1994; Vessey, 1994; Howieson & Dilworth, 2016), as has been done in this experiment.

It is true that there is theoretical and empirical support for a conversion ratio of C_2H_2 reduced to fixed N_2 of 3–4 to 1 (Hardy *et al.*, 1973; Giller, 2001) but as recommend by Giller (2001) and Unkovich et al. (2008), static ARA has been calibrated by $^{15}N_2$ in our group by Dr Carmen Sánchez-Cañizares in collaboration with Terpollili's lab in pea plants using *R. leguminosarum* bv. *viciae* A34 and in bean plants using *R. leguminosarum* bv. *phaseoli* 4292. These results were obtained from both pea and bean plants grown for 4 and 6 weeks, respectively, in vermiculite in a growth chamber. From these plants, a broad range of parameters were measured, all reflecting the same tendency of N_2 fixation: shoot and nodules fresh and dry weight, nodule number, bacteroid protein, ARA or$^{15}N_2$ among others (manuscript in preparation)[5]

[5] Personal communication with Dr Carmen Sánchez-Cañizares

Therefore, based on our findings, we can used GFP florescence detection under the control of psNifH as a proxy of nitrogenase activity in individual nodules.

Conclusion

In this chapter we designed and made a synthetic consensus promoter (psNifH), compatible with Golden Gate cloning, to drive nodule-specific expression in biovars and strains of *R. leguminosarum* and potentially other species of rhizobia (e.g. *R. tropici, S. fredii*). Expression of reporter proteins CelB, GusA, mCherry and sfGFP confirmed that psNifH can be used in a broad range of *R. leguminosarum* strains, although the plasmid pOGG026, made from minimal essential parts, was incompatible with psNifH driving nodule-specific *lux* expression.

We found that psNifH does not affect the competitiveness of rhizobial strains or change their ability to form nodules on roots. Fluorescence from sfGFP can be used as an easily detected and quantified proxy for nitrogenase activity in individual nodules.

Now that we have a vector, pOGG026, suitable for experiments in environments where there is no antibiotic selection (Chapter 3) and a functional psNifH which can be used in high-throughput screening (this chapter), all is now set for the high-throughput analysis of rhizobial isolates, which will be the focus of Chapter 5.

Chapter 5: Design and testing of tagging bacterial strains using Plasmid IDs

Introduction

For a rhizobial strain to be selected as a suitable inoculant it must first be evaluated for its tolerance to environmental stresses, its effectiveness to fix N_2, its soil persistence and, most importantly, its ability to compete against high numbers of indigenous strains for nodule formation (Dilworth *et al.*, 2008). Examples of techniques to assess rhizobial competitiveness include: i) integration of marker genes into the host genome, demonstrated in a recent study by (Sánchez-Cañizares & Palacios, 2013); ii) identification of rhizobial strains from nodules by the use of native or induced antibiotic resistance (Poysti *et al.*, 2007; Geddes & Oresnik, 2012); and iii) the use of PCR-fingerprints which target specific genes (Laguerre *et al.*, 1996; Thies *et al.*, 2001; Duodu *et al.*, 2009; Mishra *et al.*, 2012). Despite the success of these techniques for rhizobial identification, they do not lend themselves to rapid analysis of a large number of samples since individual analysis of nodules is required, which is labour intensive and thus limits the number of strains that can be analysed in competition. Even in the most recent studies using specific primers, a maximum of only six strains has been used in competition assays (Ji *et al.*, 2017; Lardi *et al.*, 2017).

Metagenomics approaches for high-throughput identification of strains do exist, such as the small-subunit ribosomal RNA (16S rRNA) gene locus that is highly conserved across microbes (Pace *et al.*, 1986), but unfortunately, use of this method is not possible here since there is only low sensitivity due to the inability to distinguish

between closely related strains as they all have the same or very similar 16S rRNA. Additionally, bacterial genomes often present several copies of the 16S rRNA genes (Větrovský & Baldrian, 2013).

On the other hand, Shotgun metagenomic sequencing has the ability to monitor strain-level changes but has limitations in terms of potential experimental biases and very complex computational analyses that can lead to different interpretations (Quince *et al.*, 2017).

Despite extensive progress in the study of *Rhizobium*-legume symbioses and metagenomics techniques, we still lack techniques that can diagnose large numbers of rhizobia strains for nodule occupancy. However, without substantially increasing the number of rhizobial strains that can be evaluated simultaneously, it will be difficult and laborious to improve rhizobial inoculants.

To tackle this problem, in this chapter we present the design of reporter plasmids, which include unique synthetic nucleotide sequences, referred to as Plasmid IDs, in order that multiple *R. leguminosarum* strains can be marked and identified from one another in a single experiment through the use of next generation sequencing (NGS).

Ion Torrent's PGM (Ion Torrent) was evaluated as a High-Throughput NGS method since an Ion Torrent Ion Chef System can be easily accessed in our institution. However, due to low sequence data accuracy previously reported (Quail *et al.*, 2012; Bragg *et al.*, 2013; Speranskaya *et al.*, 2018) we decided to do parallel controls by Illumina MiSeq (MiSeq) sequencing.

Our results demonstrate: i) the design and construction of ninety-five plasmids, each with the same reporter gene and a different Plasmid ID; ii) the use of those Plasmid IDs

in the analysis of rhizobial isolates to determine nodule occupancy and iii) that Plasmid IDs (or the plasmids that they label), were not transferred between rhizobial strains under the selection pressure of a sterile pea rhizosphere.

Results and discussion

Design of a Plasmid ID for bacterial identification at strain level

The use of Golden Gate cloning as a strategy for plasmid construction allows easy modular construction of plasmids. To take advantage of this we designed our tagging method of a 12-nucleotide specific sequence (Plasmid ID) within a Golden Gate cloning module. The strategy used to build pOPS0381 (vector module design and Level 1 assembly, described in Chapter 4, Fig. 5. 1) resulted in two contiguous terminators: T-pharma from Level 0 T-module and the *rho*-independent T_0 lambda phage transcription terminator (Lutz & Bujard, 1997; Silva-Rocha *et al.*, 2013) coming from vector pOGG026. We decided to replace the Level 0 T module (T-pharma) with one containing the Plasmid ID. By designing a Level 0 T module that includes an ID, coded plasmids can be rapidly constructed.

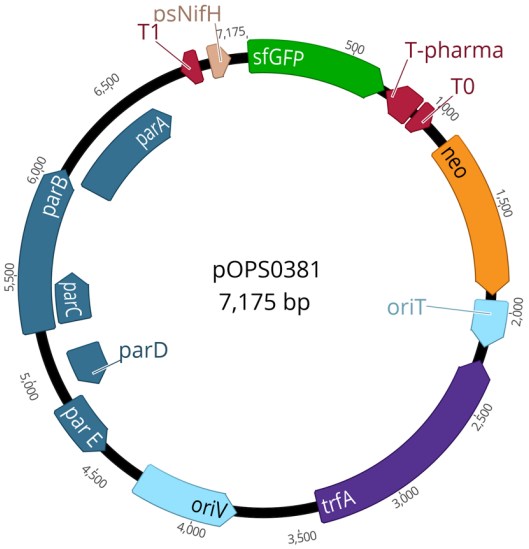


Figure 5. 1 Schematic map of pOPS0381. Two terminator regions are next each other, shown in red. These are T-pharma, a Level 0 T module, and T0 from the construction of vector pOGG026.

We searched in the literature for foreign nucleotide sequences that could be used as Plasmid IDs. A unique 12-nt, error-correcting Golay barcode described by Caporaso *et al.*, (2012) was selected as suitable for our need of strain identification. The ninety-six selected Golay barcodes are given in Table 5. 1.

Table 5. 1 Golay barcode sequences and assigned position in a 96-well pate to be used as Plasmid IDs

Golay barcode	Sequence	Assigned position	Golay barcode	Sequence	Assigned position	Golay barcode	Sequence	Assigned position	Golay barcode	Sequence	Assigned position
806rcbc0	TCCCTTGTCTCC	A1	**806rcbc24**	GCGATATATCGC	C1	**806rcbc48**	TACTACGTGGCC	E1	**806rcbc72**	GGTGACTAGTTC	G1
806rcbc1	ACGAGACTGATT	A2	**806rcbc25**	CGAGCAATCCTA	C2	**806rcbc49**	GGCCAGTTCCTA	E2	**806rcbc73**	ATGGGTTCCGTC	G2
806rcbc2	GCTGTACGGATT	A3	**806rcbc26**	AGTCGTGCACAT	C3	**806rcbc50**	GATGTTCGCTAG	E3	**806rcbc74**	TAGGCATGCTTG	G3
806rcbc3	ATCACCAGGTGT	A4	**806rcbc27**	GTATCTGCGCGT	C4	**806rcbc51**	CTATCTCCTGTC	E4	**806rcbc75**	AACTAGTTCAGG	G4
806rcbc4	TGGTCAACGATA	A5	**806rcbc28**	CGAGGGAAAGTC	C5	**806rcbc52**	ACTCACAGGAAT	E5	**806rcbc76**	ATTCTGCCGAAG	G5
806rcbc5	ATCGCACAGTAA	A6	**806rcbc29**	CAAATTCGGGAT	C6	**806rcbc53**	ATGATGAGCCTC	E6	**806rcbc77**	AGCATGTCCCGT	G6
806rcbc6	GTCGTGTAGCCT	A7	**806rcbc30**	AGATTGACCAAC	C7	**806rcbc54**	GTCGACAGAGGA	E7	**806rcbc78**	GTACGATATGAC	G7
806rcbc7	AGCGGAGGTTAG	A8	**806rcbc31**	AGTTACGAGCTA	C8	**806rcbc55**	TGTCGCAAATAG	E8	**806rcbc79**	GTGGTGGTTTCC	G8
806rcbc8	ATCCTTTGGTTC	A9	**806rcbc32**	GCATATGCACTG	C9	**806rcbc56**	CATCCCTCTACT	E9	**806rcbc80**	TAGTATGCGCAA	G9
806rcbc9	TACAGCGCATAC	A10	**806rcbc33**	CAACTCCCGTGA	C10	**806rcbc57**	TATACCGCTGCG	E10	**806rcbc81**	TGCGCTGAATGT	G10
806rcbc10	ACCGGTATGTAC	A11	**806rcbc34**	TTGCGTTAGCAG	C11	**806rcbc58**	AGTTGAGGCATT	E11	**806rcbc82**	ATGGCTGTCAGT	G11
806rcbc11	AATTGTGTCGGA	A12	**806rcbc35**	TACGAGCCCTAA	C12	**806rcbc59**	ACAATAGACACC	E12	**806rcbc83**	GTTCTCTTCTCG	G12
806rcbc12	TGCATACACTGG	B1	**806rcbc36**	CACTACGCTAGA	D1	**806rcbc60**	CGGTCAATTGAC	F1	**806rcbc84**	CGTAAGATGCCT	H1
806rcbc13	AGTCGAACGAGG	B2	**806rcbc37**	TGCAGTCCTCGA	D2	**806rcbc61**	GTGGAGTCTCAT	F2	**806rcbc85**	GCGTTCTAGCTG	H2
806rcbc14	ACCAGTGACTCA	B3	**806rcbc38**	ACCATAGCTCCG	D3	**806rcbc62**	GCTCGAAGATTC	F3	**806rcbc86**	GTTGTTCTGGGA	H3
806rcbc15	GAATACCAAGTC	B4	**806rcbc39**	TCGACATCTCTT	D4	**806rcbc63**	AGGCTTACGTGT	F4	**806rcbc87**	GGACTTCCAGCT	H4
806rcbc16	GTAGATCGTGTA	B5	**806rcbc40**	GAACACTTTGGA	D5	**806rcbc64**	TCTCTACCACTC	F5	**806rcbc88**	CTCACAACCGTG	H5
806rcbc17	TAACGTGTGTGC	B6	**806rcbc41**	GAGCCATCTGTA	D6	**806rcbc65**	ACTTCCAACTTC	F6	**806rcbc89**	CTGCTATTCCTC	H6
806rcbc18	CATTATGGCGTG	B7	**806rcbc42**	TTGGGTACACGT	D7	**806rcbc66**	CTCACCTAGGAA	F7	**806rcbc90**	ATGTCACCGCTG	H7
806rcbc19	CCAATACGCCTG	B8	**806rcbc43**	AAGGCGCTCCTT	D8	**806rcbc67**	GTGTTGTCGTGC	F8	**806rcbc91**	TGTAACGCCGAT	H8
806rcbc20	GATCTGCGATCC	B9	**806rcbc44**	TAATACGGATCG	D9	**806rcbc68**	CCACAGATCGAT	F9	**806rcbc92**	AGCAGAACATCT	H9
806rcbc21	CAGCTCATCAGC	B10	**806rcbc45**	TCGGAATTAGAC	D10	**806rcbc69**	TATCGACACAAG	F10	**806rcbc93**	TGGAGTAGGTGG	H10
806rcbc22	CAAACAACAGCT	B11	**806rcbc46**	TGTGAATTCGGA	D11	**806rcbc70**	GATTCCGGCTCA	F11	**806rcbc94**	TTGGCTCTATTC	H11
806rcbc23	GCAACACCATCC	B12	**806rcbc47**	CATTCGTGGCGT	D12	**806rcbc71**	CGTAATTGCCGC	F12	**806rcbc95**	GATCCCACGTAC	H12

Since terminator T0 is exactly downstream of the ID position, it could interfere with the isolation and amplification of this region of interest, due to the strong DNA fold formation of T0. In order to prevent this problem, a Universal Primer binding site (5'-CGTTTTACAACGTCGTGACTGGG-3') was added between the 12-nt Golay barcode and the T0. *In silico* analysis (with up to 5 mismatches) of the Universal primer was done in the full plasmid maps to ensure that this primer only binds to this specific region. The final construction was completed with the addition of type IIS restriction enzyme recognition sites and the specific overhangs at either end of a Golden Gate T module.

Incorporating the Plasmid ID, the final design of the T module is 57-bp in size (Figure 5. 2) (this is compared to 180-bp T-pharma T module). The features are BsaI enzyme recognition sites at either end (nucleotides 1-6 and 52-57) with the 5'-GCTT-3' and 5'-CGCT-3' overhangs of a Golden Gate T module, the Plasmid ID (Golay barcode) at nucleotides 12- 23 and a universal primer binding site at nucleotides 24-46.

The Plasmid IDs were ordered as sense and antisense oligonucleotides, so they can be used directly in a Golden Gate reaction after annealing, without requiring digestion with BsaI. The sequences of these oligonucleotides are given in Supplemental Material Table S.2.

Each Plasmid ID pair was assigned to a position in a 96-well plate (Table 5. 1) and oligonucleotide annealing was performed in this plate, rendering each as a T module.

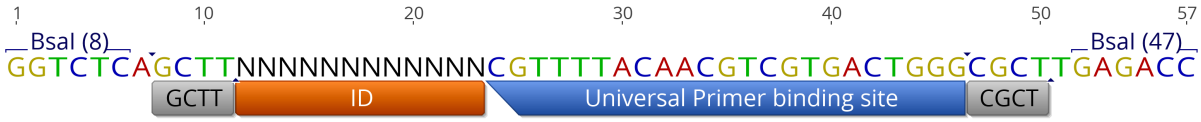


Figure 5. 2 Plasmid ID T module with terminal BsaI restriction sites and overhangs (grey, sequence 5' to 3') for Golden Gate assembly. Plasmid ID is due to 12-nt Golay barcode (orange) and the universal primer binding site (blue).

A total of ninety-six individual and unique reporter plasmids were built by Golden Gate Level 1 cloning one-tube one-step reaction (BsaI digestion and ligation) (Engler *et al.*, 2008) with the following modules utilising a 'master-mix' method for 100 reactions: vector; pOGG026, PU module; psNifH from pOGG043, SC module; sfGFP from pOGG037. The mix was distributed into a 96-well plate and the product from annealing the oligonucleotide Plasmid IDs was added (T module) according to the position designated in Table 5.1 to make the ninety-six individual clones (Fig. 5. 3).

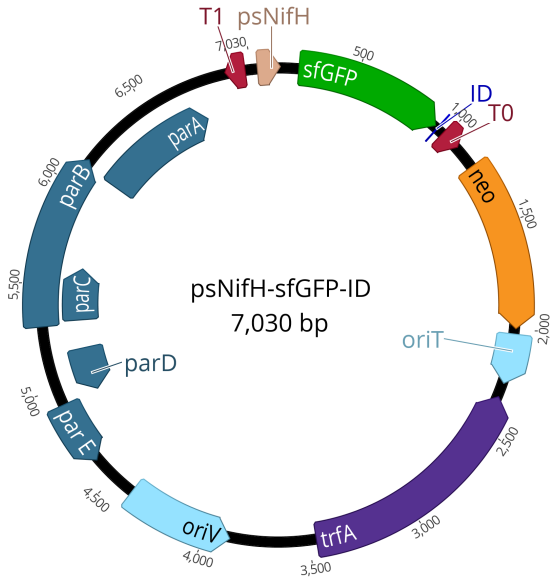


Figure 5. 3 Schematic map of ninety-six plasmids constructed with unique Plasmid IDs showing the Plasmid ID (in blue) in T module position. Components for construction of all plasmids were identical, with the exception of the T module Plasmid ID which was different in each case.

The step of transforming the Golden Gate cloning products into *E. coli* cells is normally done in individual Eppendorf tubes but since we had a high number of constructs, a new methodology was developed to allow high-throughput transformation of all ninety-six products at the same time (for more details see Chapter 2). From counting the colonies obtained on X-gal selection plates (where white colonies show an insert), we observed that cloning efficiency was similar to that described by Weber *et al.*, 2011. One colony per transformation was selected to verify the correct plasmid construction by restriction digest. To confirm that both the plasmid constructions and the expected Plasmid ID were found in a well, all ninety-six products were sent for sequencing, using primers oxp1363 and Universal Primer, to confirm the correct

constructions. From all products, just that from well H9 did not show the excepted sequence. Therefore, we decided to leave well H9 empty and to use it in the future as an experimental negative control. The final transformation products were stored in a -80^{o}C freezer as a plasmid library in a 96-well plate, retaining the same well position as assigned to Golay barcode in Table 5. 1.

Use of a Plasmid ID for high-throughput identification of rhizobial isolates for nodule occupancy

A proof-of-concept experiment was designed to identify rhizobial nodule occupancy using the Plasmid ID. The strains UPM791*gusA* and *Rlv*3841*celB*, as used by Sánchez-Cañizares & Palacios, (2013), were selected as they have a constitutive promoter driving the reporter gene with the *gusA* and *celB* marker genes introduced into the genome. UPM791*gusA* was conjugated with the Plasmid ID-E10 (pOPS0548) and *Rlv*3841*celB* was conjugated with Plasmid ID-G2 (pOPS0564) (Fig 5. 4A). We performed a competition assay on pea plants between UPM791*gusA*-E10[pOPS0548] and *Rlv*3841*celB*-G2[pOPS0564] (Fig 5. 4B) and the strain occupying the nodules formed was identified after sequential double-staining with Magenta-glcA and X-gal after thermal treatment (Fig 5. 4C).

While legumes typically are infected by a single clone, mixed clonal infections of their symbiont can be found but it is difficult to analyse them since the proportion of mixed nodules is very low (Gage, 2004). Nodule staining with Magenta-glcA and X-gal has facilitated post harvesting analysis of mixed nodules (Sessitsch *et al.*, 1997; Sánchez-Cañizares & Palacios, 2013) and in this this study, using this staining technique, we

were able to recover both single infection nodules and mixed nodules to investigate the possibility of differentiating the proportion of sequence reads.

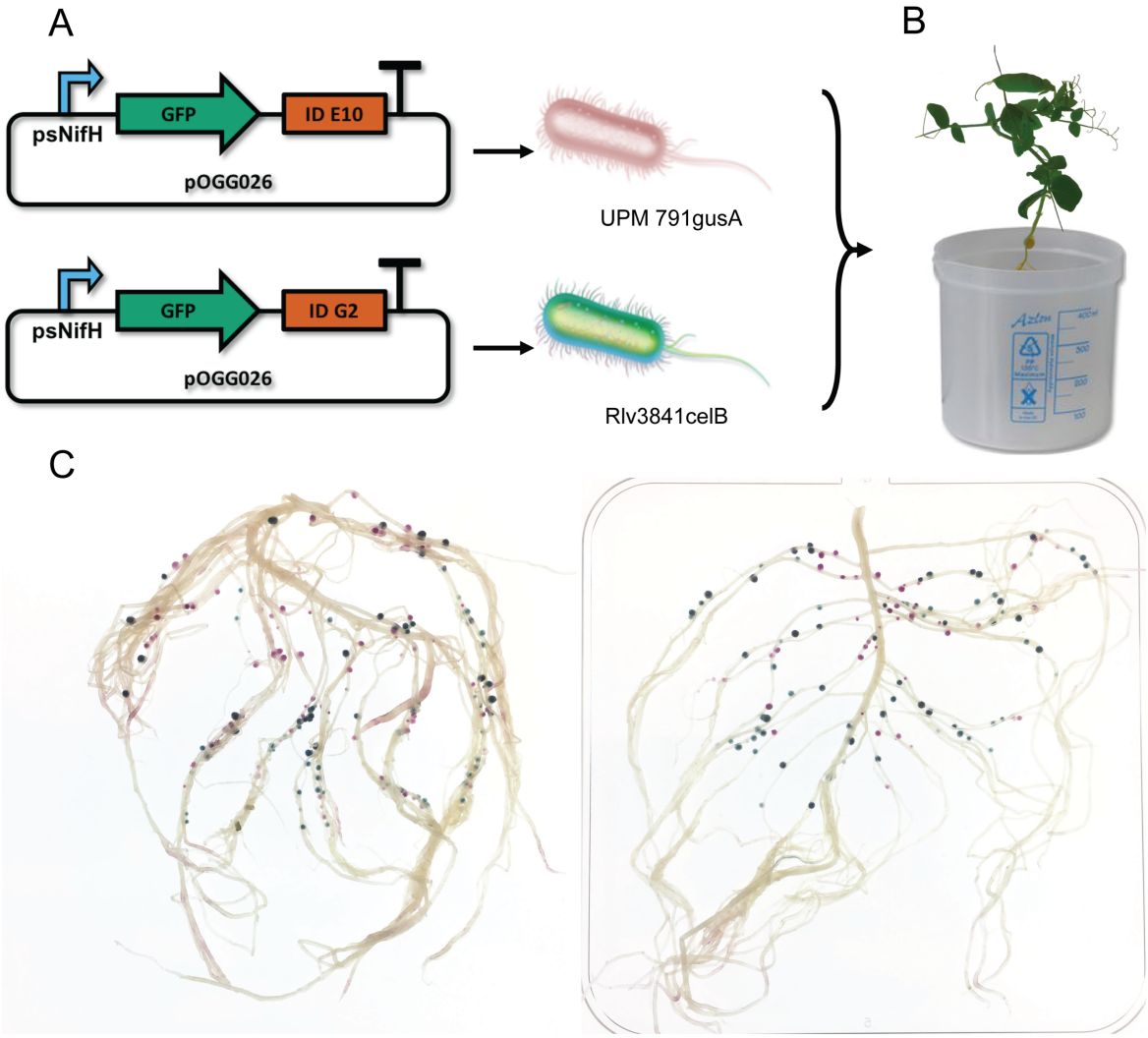


Figure 5. 4 A) Plasmid ID-E10 (pOPS0548) and Plasmid ID-G2 (pOPS0564) were conjugated into UPM791*gusA* and *Rlv*3841*celB*, respectively and used to B) co-inoculate pea plants. C) Examples of pea roots that were sequential double-stained with Magenta-glcA, and X-gal after thermal treatment, pink nodules formed by UPM791*gusA*-E10[pOPS0548] (*gusA* constitutively expressed) and blue nodules formed by *Rlv*3841*celB*-G2[pOPS0564] (*celB* constitutively expressed).

Forty nodules were selected for analysis by NGS with Ion Torrent. These included 18 single infection nodules from each strain and four mixed nodules (i.e. all the mixed nodules that were found) (Fig 5. 5).

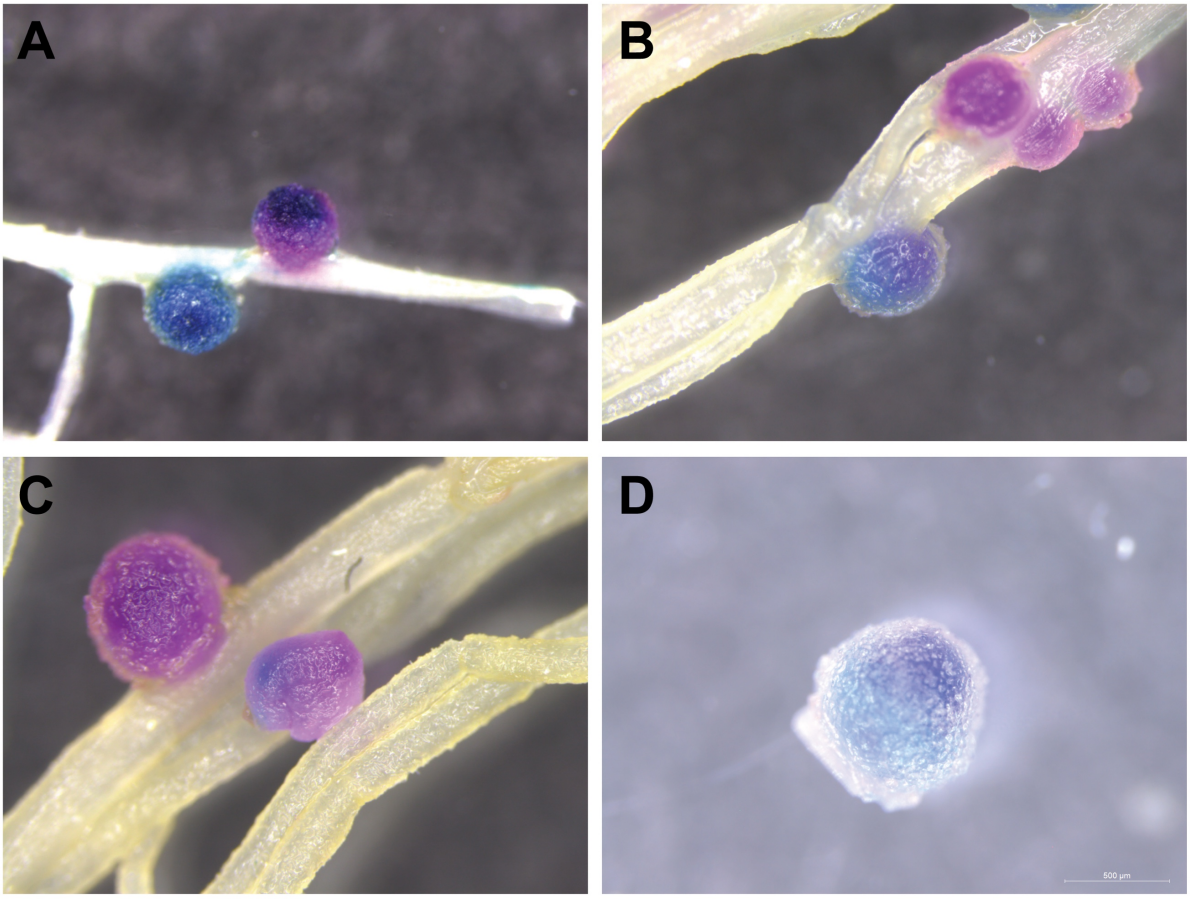


Figure 5. 5 A-D) Microscopic view of the four mixed nodule that were found. Pink colour in nodules due to UPM791*gusA*-E10 [pOPS0548] (*gusA* constitutively expressed) and blue colour due to *Rlv*3841*celB*-G2[pOPS0564] (*celB* constitutively expressed).

Primer design for multiplex Ion Torrent sequencing strategy

We designed primers for a two-step polymerase chain reaction (PCR) (Berry et al., 2011; Kozich et al., 2013; Fadrosh et al., 2014) taking into account the following

restrictions: 1) First-step PCR primers needed to add specific barcodes to our amplicon sequence with the aim of identifying column and row samples in a 96-well plate, 2) should insert 'landing pads' for the second-step PCR preventing hairpin formation or binding to our amplicon sequence, 3) Second-step forward primers should add Ion Xpress™ Barcodes for library identification, 4) Second-step reverse primer should introduce the trP1 sequencing adapter and 5) the final sequencing template should not exceed 200bp in length.

The size of the First-step PCR forward primers (Table 5. 2) was minimised to include just 18 nt from the original 29 nt of the 'landing pad' compatible with Second-step PCR Forward primers from Wheatley *et al.*, 2017 (Table 5. 3). The reverse 'landing pad' for the Second-step PCR reverse primer used before by (Perry & Yost, 2014; Wheatley *et al.*, 2017) was forming a hairpin with our amplicon sequence, and for this reason, we substituted the 'landing pad' with a truncated version of the Illumina Nextera transposase sequence presented by de Souza *et al.*, 2016 (Table 5. 2). The final Second-step PCR reverse primer length is 41 nt (Table 5. 3) instead of 49 nt from the original primer published by (Perry & Yost, 2014).

All the primers, landing pads, adapters and barcodes sequences were analysed *In silico*, as individual units and as a complete sequence, and no hairpin formations were found with up to 5 mismatches. All these modifications allowed us to have a final sequencing template of 195-bp in length.

Table 5. 2 First- step PCR primers for Ion Torrent sequencing

Well tagging			
Forward primers (column tagging). Sequence (5'-3')			
Name	**Forward landing pad[a]**	**Barcode**	**Primer**
1_IT	GCCCAGTCTACTCGAGGG	GCTA	CATCACGCATGGTATGGA
2_IT	GCCCAGTCTACTCGAGGG	TGTGT	CATCACGCATGGTATGGA
3_IT	GCCCAGTCTACTCGAGGG	AGTCTG	CATCACGCATGGTATGGA
4_IT	GCCCAGTCTACTCGAGGG	ATCA	CATCACGCATGGTATGGA
5_IT	GCCCAGTCTACTCGAGGG	GACGA	CATCACGCATGGTATGGA
6_IT	GCCCAGTCTACTCGAGGG	TCGTCG	CATCACGCATGGTATGGA
7_IT	GCCCAGTCTACTCGAGGG	TGCT	CATCACGCATGGTATGGA
8_IT	GCCCAGTCTACTCGAGGG	CAGTT	CATCACGCATGGTATGGA
9_IT	GCCCAGTCTACTCGAGGG	ACATGT	CATCACGCATGGTATGGA
10_IT	GCCCAGTCTACTCGAGGG	GCGG	CATCACGCATGGTATGGA
11_IT	GCCCAGTCTACTCGAGGG	GTTGA	CATCACGCATGGTATGGA
12_IT	GCCCAGTCTACTCGAGGG	GTGGCT	CATCACGCATGGTATGGA
Reverse primers (row tagging). Sequence (5'-3')			
Name	**Reverse landing pad[b]**	**Barcode**	**Primer**
A_IT	ATCTCGGTGGTCGCCGTA	GCTC	CCCAGTCACGACGTTGTAAAACG
B_IT	ATCTCGGTGGTCGCCGTA	CTAGT	CCCAGTCACGACGTTGTAAAACG
C_IT	ATCTCGGTGGTCGCCGTA	TAGATC	CCCAGTCACGACGTTGTAAAACG
D_IT	ATCTCGGTGGTCGCCGTA	TCGC	CCCAGTCACGACGTTGTAAAACG
E_IT	ATCTCGGTGGTCGCCGTA	CCTTA	CCCAGTCACGACGTTGTAAAACG
F_IT	ATCTCGGTGGTCGCCGTA	CATAAC	CCCAGTCACGACGTTGTAAAACG
G_IT	ATCTCGGTGGTCGCCGTA	CAGA	CCCAGTCACGACGTTGTAAAACG
H_IT	ATCTCGGTGGTCGCCGTA	TGTTC	CCCAGTCACGACGTTGTAAAACG

[a] Modified from (Wheatley *et al.*, 2017)

[b] Modified from Illumina Nextera transposase sequence

Table 5. 3 Second- step PCR primers for Ion Torrent sequencing

Plate tagging			
Forward primers (Library tagging). Sequence (5'-3')[a]			
Name	**IT adapter**	**Ion Xpress™ Barcode**	**Primer for Forward landing pad**
IT_A_FP_1	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CTAAGGTAAC	GATATAAAACCGCCCAGTCTACTCGAGGG
IT_A_FP_2	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TAAGGAGAAC	GATATAAAACCGCCCAGTCTACTCGAGGG
IT_A_FP_3	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AAGAGGATTC	GATATAAAACCGCCCAGTCTACTCGAGGG
IT_A_FP_4	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TACCAAGATC	GATATAAAACCGCCCAGTCTACTCGAGGG
IT_A_FP_5	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CAGAAGGAAC	GATATAAAACCGCCCAGTCTACTCGAGGG
IT_A_FP_6	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CTGCAAGTTC	GATATAAAACCGCCCAGTCTACTCGAGGG
IT_A_FP_7	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TTCGTGATTC	GATATAAAACCGCCCAGTCTACTCGAGGG
IT_A_FP_8	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TTCCGATAAC	GATATAAAACCGCCCAGTCTACTCGAGGG
IT_A_FP_9	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TGAGCGGAAC	GATATAAAACCGCCCAGTCTACTCGAGGG
IT_A_FP_10	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CTGACCGAAC	GATATAAAACCGCCCAGTCTACTCGAGGG
Reverse primer compatible with synthetic amplicons. Sequence (5'-3')			
Name	**trP1 sequencing adapter**	**Primer for reverse landing pad**	
IT_trP1_FP_Synthetic	CCTCTCTATGGGCAGTCGGTGAT	ATCTCGGTGGTCGCCGTA	

[a] From (Wheatley *et al.*, 2017)

DNA from surface-sterilised nodules was prepared for multiplex sequencing in a 96-well plate. Mixed nodules were prepared in the following wells: 1A (Fig 5. 5A), 1B (Fig 5. 5B), 1C (Fig 5. 5C) and 1D (Fig 5. 5D). Figure 5. 6A indicates the wells in which UPM791*gusA*-E10[pOPS0548] nodules were prepared (in pink) and the wells in which *Rlv*3841*celB*-G2[pOPS0564] nodules were prepared (in blue). Samples were tagged in the First-step PCR (Fig. 5. 6B) and pooled at equal molarity, and samples from different 96-well plates were kept separate (Fig. 5. 6C). Libraries were tagged and a trP1 sequencing adapter was introduced in the Second-step PCR (Fig. 5. 6D). With more than one library, each library needs to be prepared for quality and equal molarity so that they can be pooled before sequencing (Fig. 5. 6E).

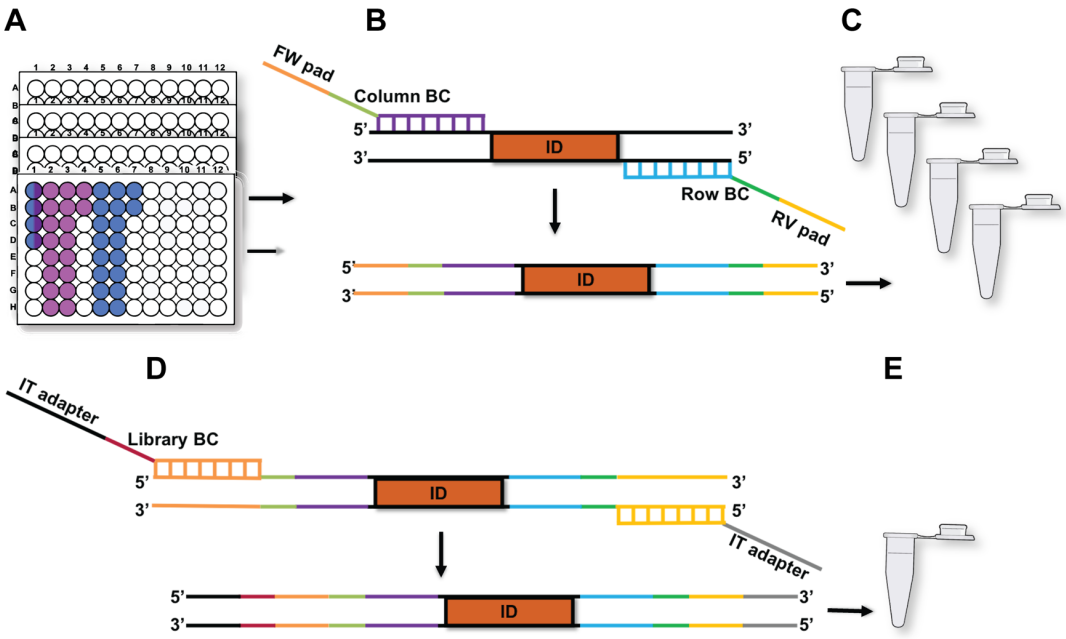


Figure 5. 6 Workflow for Plasmid ID sequencing. A) DNA extraction from nodules prepared for multiplex sequencing, mixed nodules prepared in wells 1A to 1D, UPM791*gusA*-E10[pOPS0548] nodules prepared in pink wells and Rlv3841*celB*-G2[pOPS0564] nodules prepared in blue wells. B) First-step PCR primers target Plasmid ID for amplicon sequencing and add specific barcodes to identify the sample well. C) Pooled 96 samples per library. D) Second-PCR step primers tagging each library preparation and added sequencing adapters, E) libraries are pooled for sequencing.

Sequencing results were analysed following the workflow in Figure 5. 7. Pre-processing steps were carried out with Trimmomatic-0.36 (Bolger *et al.*, 2014) and we developed a script in Biopython (Cock *et al.*, 2009) (Supplemental Material, Figure S.1) to separate all the reads by amplicon (Plasmid ID) and sample-specific barcodes. Data manipulation was carried out in Microsoft Excel.

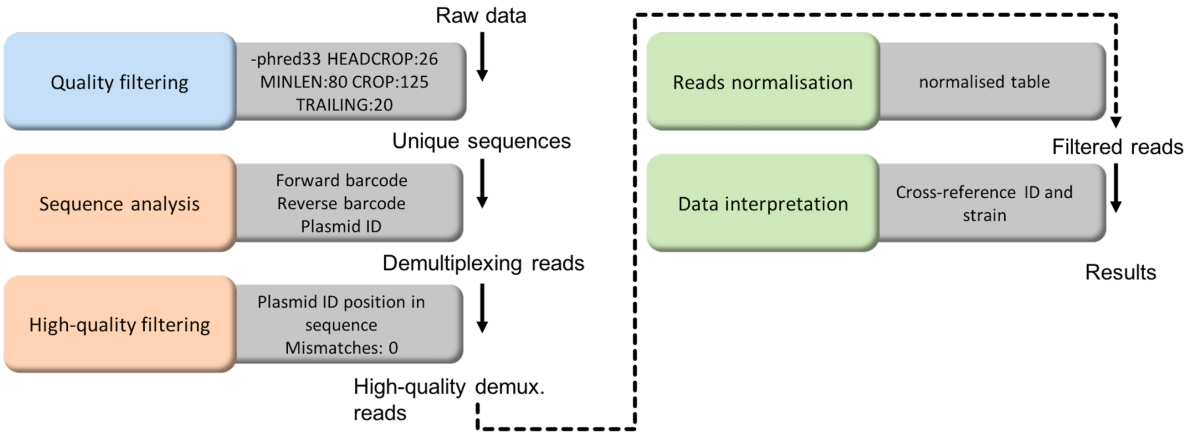


Figure 5. 7 Workflow for data analysis from Ion Torrent reads. Pre-processing steps in blue rectangles, sequence analysis in orange rectangles and data manipulation in green rectangles.

After the high-quality filtering, where no mismatches were allowed, all samples identified with reads below 1% of the total number of unique sequences were removed to reduce false positive reads (Murray *et al.*, 2015). A total of 13 million reads were identified for sample-specific and Plasmid ID. PivotTables were used to cluster the reads of Plasmid IDs to their corresponding column and row barcodes mimicking the original 96-well plate. Results are shown as total counts per well (Table 5. 4) and as a percentage of counts per well (Table 5. 5).

Table 5. 4 Total counts of plasmid ID E10 and G2 assigned to column and row barcodes. Results are classified in graded coloured scale where green is the highest value, yellow is the midpoint percentile and red is the minimum value, cells in white do not show reads.

		Column Barcode																					
Row		1		2		3		4		5		6		7		8		9		10		11		12	
Barcode	ID	E10	G2	E10	G2	E10	G2	E10	G2	E10	G2	E10	G2	E10	G2	E10	G2	E10	G2	E10	G2	E10	G2	E10	G2
A		3E+04	2E+03	5E+04		6E+04		3E+04			1E+05		2E+05		3E+05										
B		4E+04	2E+05	2E+05	5E+03	2E+05	4E+03	1E+05			4E+05		4E+05		4E+05										
C		1E+05	2E+04	2E+05	2E+03	1E+05	3E+03				4E+05		5E+05												
D		1E+04	1E+05	2E+05	3E+03	1E+05	2E+03				3E+05		4E+05												
E				1E+05		1E+05					2E+05		4E+05												
F				2E+05		1E+05					3E+05	4E+03	4E+05												
G				2E+05		1E+05					3E+05		4E+05												
H				1E+05		9E+04					3E+05		5E+05												

Table 5. 5 Percentage of plasmid ID E10 and G2 counts from total counts assigned to column and row barcodes. Results are classified in graded coloured scale where green is the highest value, yellow is the midpoint percentile and red is the minimum value, cells in white do not show reads.

		Column Barcode																					
Row		1		2		3		4		5		6		7		8		9		10		11		12	
Barcode	ID	E10	G2	E10	G2	E10	G2	E10	G2	E10	G2	E10	G2	E10	G2	E10	G2	E10	G2	E10	G2	E10	G2	E10	G2
A		94%	6%	100%	0%	100%	0%	100%	0%	0%	100%	0%	100%	0%	100%										
B		18%	82%	98%	2%	98%	2%	100%	0%	0%	100%	0%	100%	0%	100%										
C		90%	10%	99%	1%	98%	2%			0%	100%	0%	100%												
D		9%	91%	98%	2%	98%	2%			0%	100%	0%	100%												
E				100%	0%	100%	0%			0%	100%	0%	100%												
F				100%	0%	100%	0%			0%	100%	1%	99%												
G				100%	0%	100%	0%			0%	100%	0%	100%												
H				100%	0%	100%	0%			0%	100%	0%	100%												

Plasmid ID sequencing results from all forty samples agreed entirely with the strain expected from the colour of the nodule from which they were prepared (Table 5. 4). For the mixed nodules (wells 1A, 1B, 1C and 1D, Table 5. 5), different percentages of the two barcodes were found. Theses ranged from 9-94% and, reciprocally, from 6-91%. Although it is known that Ion Torrent PGM platform has a higher raw error rate (~1.38% after quality filtering) (Quail *et al.*, 2012; Bragg *et al.*, 2013) the percentages obtained in mixed nodules were higher than this rate and the proportions of the two strains matched the approximate colour of each mixed nodule (Fig 5. 5). Sequencing data revealed that analysed samples corresponded in 100% of the cases with the predicted Plasmid ID. Furthermore, this method is sensitive enough to identify mixed nodules, unlike techniques for nodule infection identification performed using antibiotic resistance (Oresnik *et al.*, 1999; Poysti *et al.*, 2007; Geddes & Oresnik, 2012) or primer specificity (Ji *et al.*, 2017; Lardi *et al.*, 2017).

Plasmid ID data analysis by MiSeq sequencing

A sample was sent for MiSeq sequencing as a control. MiSeq was selected as a control technique since it has better sequence data accuracy compared to Ion Torrent (Quail *et al.*, 2012; Bragg *et al.*, 2013; Speranskaya *et al.*, 2018). Primers previously used to validate plasmid ID constructions, oxp1363 and Universal Primer, were complemented with the same specific barcodes used in the First- step PCR primers for Ion Torrent sequencing (Table 5. 6).

Table 5. 6 Primers for MiSeq multiplex sequencing. Specific barcodes for sample tagging.

Well tagging **Forward primers (column tagging). Sequence (5'-3')**		
Name	**Barcode**	**Primer**
1_MiSeq	GCTA	GTTGAGGATGGCAGCGTG
2_MiSeq	TGTGT	GTTGAGGATGGCAGCGTG
3_MiSeq	AGTCTG	GTTGAGGATGGCAGCGTG
4_MiSeq	ATCA	GTTGAGGATGGCAGCGTG
5_MiSeq	GACGA	GTTGAGGATGGCAGCGTG
6_MiSeq	TCGTCG	GTTGAGGATGGCAGCGTG
7_MiSeq	TGCT	GTTGAGGATGGCAGCGTG
8_MiSeq	CAGTT	GTTGAGGATGGCAGCGTG
9_MiSeq	ACATGT	GTTGAGGATGGCAGCGTG
10_MiSeq	GCGG	GTTGAGGATGGCAGCGTG
11_MiSeq	GTTGA	GTTGAGGATGGCAGCGTG
12_MiSeq	GTGGCT	GTTGAGGATGGCAGCGTG

Reverse primers (row tagging). Sequence (5'-3')		
Name	**Barcode**	**Primer**
A_MiSeq	GCTC	CCCAGTCACGACGTTGTAAAACG
B_MiSeq	CTAGT	CCCAGTCACGACGTTGTAAAACG
C_MiSeq	TAGATC	CCCAGTCACGACGTTGTAAAACG
D_MiSeq	TCGC	CCCAGTCACGACGTTGTAAAACG
E_MiSeq	CCTTA	CCCAGTCACGACGTTGTAAAACG
F_MiSeq	CATAAC	CCCAGTCACGACGTTGTAAAACG
G_MiSeq	CAGA	CCCAGTCACGACGTTGTAAAACG
H_MiSeq	TGTTC	CCCAGTCACGACGTTGTAAAACG

DNA from 17 nodules was prepared for multiplex sequencing in a 96-well plate. Figure 5. 8A indicates the wells in which UPM791*gusA*-E10[pOPS0548] nodules were prepared (in pink) and the wells in which *Rlv*3841*celB*-G2[pOPS0564] nodules were prepared (in blue). Samples were tagged in First-step PCR (Fig. 5. 8B) and we pooled samples at equal molarity (Fig. 5. 8C). The second-step PCR to generate amplicon

libraries and sequencing by Illumina MiSeq 300PE was done at MR DNA (Molecular Research LP, Shallowater, TX, USA). The final sequencing template was 265 bp in length.

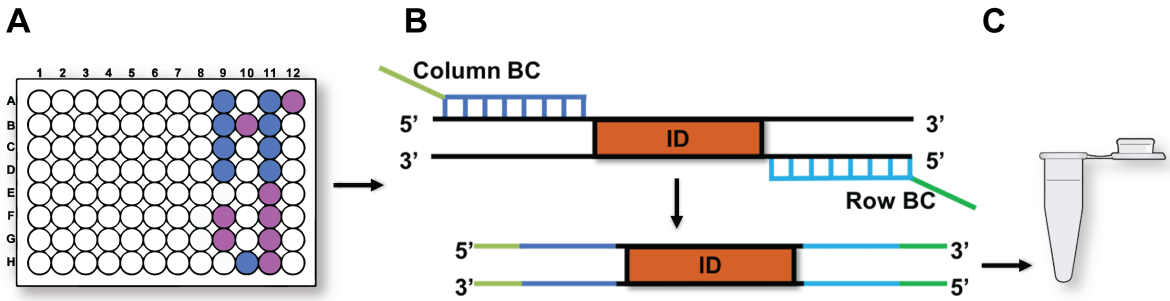


Figure 5. 8 Workflow for MiSeq sequencing. A) DNA extraction from nodule samples prepared for multiplex sequencing. UPM791*gusA*-E10[pOPS0548] nodules prepared in pink wells and Rlv3841*celB*-G2[pOPS0564] nodules prepared in blue wells B) First-step PCR primers target Plasmid ID for amplicon sequencing and add specific barcodes to identify the sample well. C) Pooled samples were MR DNA laboratory to be sequence by Illumina MiSeq 300PE.

Sequencing results were analysed following workflow from Figure 5. 9. This analysis was very similar to the one followed for Ion Torrent sequencing, the only differences were the Paired-end merging done with usearch10.0.240 (Edgar, 2010) and minimum length in the quality filtering step.

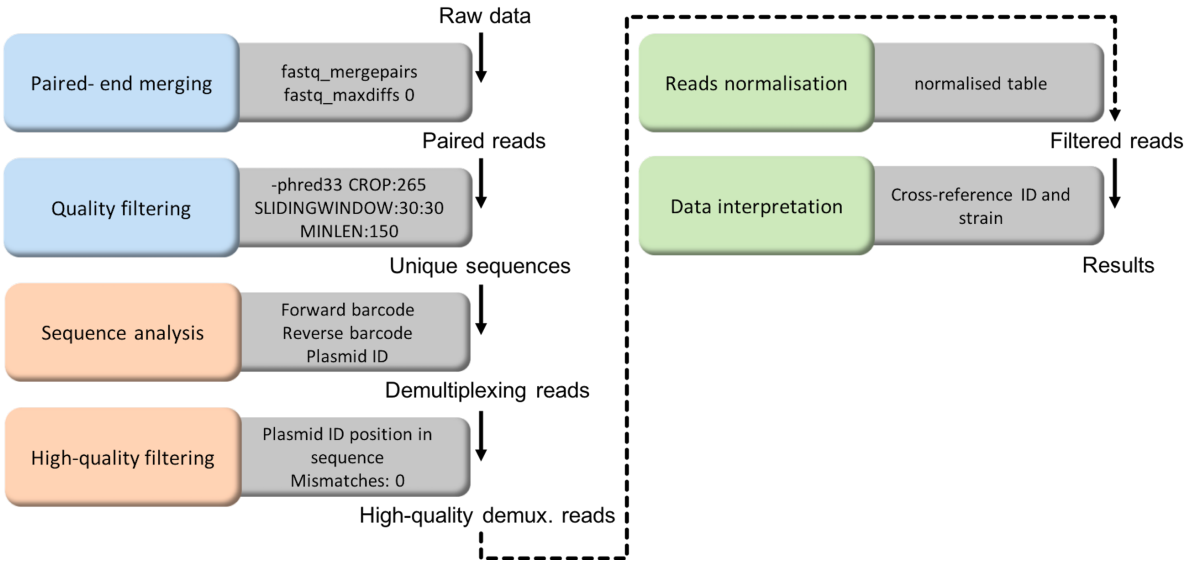


Figure 5. 9 Workflow for data analysis from MiSeq reads. Pre-processing steps in blue rectangles, sequence analysis in orange rectangles and data manipulation in green rectangles.

After the high-quality filtering, where no mismatches were allowed, all samples identified with reads below 1% of the total number of unique sequences were removed to reduce false positive reads (Murray *et al.*, 2015). A total of 450,000 MiSeq reads were identified for sample-specific and Plasmid ID. Reads of Plasmid IDs were clustered by column and row barcodes. Results are shown mimicking the original 96-well plate as total counts per well (Table 5. 7) and as a percentage of counts per well (Table 5. 8). After the high-quality filtering, MiSeq sequencing do not generated error rate compared to the Ion Torrent platform which showed up to a 2% error rate, similar rates to those reported in the literature (Quail *et al.*, 2012; Bragg *et al.*, 2013; Quince *et al.*, 2017).

Sequencing results revealed that the 17 analysed samples used as controls corresponded in all the cases with the predicted Plasmid ID.

Table 5. 7 Total counts of plasmid ID E10 and G2 assigned to column and row barcodes. Results are classified in graded coloured scale where green is the highest value, yellow is the midpoint percentile and red is the minimum value, cells in white do not show reads.

		Column Barcode																					
		1		2		3		4		5		6		7		8		9		10		11		12	
Row Barcode	ID	E10	G2	E10	G2	E10	G2	E10	G2	E10	G2	E10	G2	E10	G2	E10	G2	E10	G2	E10	G2	E10	G2	E10	G2
A																			2E+04				3E+04	1E+04	
B																			2E+04	3E+03			3E+04		
C																			4E+04				3E+04		
D																			2E+04				3E+04		
E																						2E+04			
F																		2E+04				2E+04			
G																		1E+04				3E+04			
H																					1E+04	4E+04			

Table 5. 8 Percentage of plasmid ID E10 and G2 counts from total counts assigned to column and row barcodes. Results are classified in graded coloured scale where green is the highest value, yellow is the midpoint percentile and red is the minimum value, cells in white do not show reads.

		Column Barcode																					
		1		2		3		4		5		6		7		8		9		10		11		12	
Row Barcode	ID	E10	G2	E10	G2	E10	G2	E10	G2	E10	G2	E10	G2	E10	G2	E10	G2	E10	G2	E10	G2	E10	G2	E10	G2
A																		0%	100%			0%	100%	100%	0%
B																		0%	100%	100%	0%	0%	100%		
C																		0%	100%			0%	100%		
D																		0%	100%			0%	100%		
E																						100%	0%		
F																		100%	0%			100%	0%		
G																		100%	0%			100%	0%		
H																				0%	100%	100%	0%		

Although MiSeq sequencing results shown better definition in the identification of the Plasmid IDs, the coverage and number of reads obtained with Ion Torrent are sufficient to be confident with results obtained by the Ion Torrent platform. We decided to continue using Ion torrent sequencing for future experiments since the facilities of Ion Torrent Ion Chef System are accessible to us, thus reducing the waiting time of sequencing results. Additionally, having access to these facilities gave us the opportunity to prepare our samples so that different libraries could be pooled to be sequenced in the same run, reducing sequencing cost. Most importantly though the Ion Torrent is much cheaper than using MiSeq for sequencing.

Conclusions

In this chapter we simultaneously constructed ninety-five individually-tagged reporter plasmids using Golden Gate assembly. These plasmids, assembled using high-throughput cloning, used vector pOGG026 (designed in Chapter 3), reporter sfGFP expressed under the control of the nodule-specific promoter psNifH (designed and tested in Chapter 4) and a unique Plasmid ID which can be used to identify rhizobial nodule occupancy. We have demonstrated that Plasmid IDs can be used in a competition assay for high-throughput identification of rhizobial isolates through the use of NGS (Ion Torrent and MiSeq sequencing), even identifying the percentage of two strains within a mixed nodule.

Ion Torrent and MiSeq are quite similar in terms of efficacy and workflow simplicity, therefore Plasmid IDs can be used by other groups and the decision on which platform to use for sequencing will just depend on personal experience and existing infrastructure.

Summarising our results, we have now built the molecular tools for high-throughput identification of the most competitive and effective strains (or biovars) of *R. leguminosarum* in competition experiments to assess up to ninety-five strains simultaneously.

In Chapter 6 these tools will be used in a large competition assay to evaluate the performance of a high number of *Rhizobium leguminosarum* strains against native rhizobia.

Chapter 6: Application of a molecular tool for a high-throughput identification of competitive and effective rhizobial strains

Introduction

Legumes grown under starvation of nitrogen and sterile conditions can show strong growth differences between inoculated and non-inoculated treatments. However, real agricultural conditions are never entirely devoid of nitrogen and usually some indigenous rhizobia are present (Wilson *et al.*, 1995a). Therefore, once rhizobial strains are applied to legumes under field conditions they often fail to nodulate legumes compared to indigenous strains (López-García *et al.*, 2002; Naeem *et al.*, 2004; Yates *et al.*, 2011; Drew *et al.*, 2012).

For a rhizobial strain to be considered as a potential inoculant it is easiest to first study it in the laboratory to demonstrate its effectiveness at nitrogen fixation, its ability to successfully fight for rhizosphere colonisation, and its ability to compete against indigenous rhizobia for nodule occupancy. It is then crucial to test the rhizobial strain in greenhouse and field trials (Werner & Newton, 2005; Lupwayi *et al.*, 2006). There must be many more elite strains than those currently used as inoculants, and despite the first rhizobial competition assays being presented in 1941 (Nicol & Thornton, 1941), there has been limited progress in this area due to the complexity of directly observing and identifying interactions between bacteria and plants. This compounded by the time-consuming nature of investigations using these methods.

In Chapter 5 we presented a molecular tool for incorporating a library of ninety-five Plasmid IDs for high-throughput identification of rhizobial nodule occupancy through

the use of NGS. Plasmid IDs expressed sfGFP under the control of the nodule-specific promoter psNifH which can be used as a proxy for nitrogenase activity assays in biovars and strains of *R. leguminosarum* (Chapter 4). Plasmid IDs were assembled in the broad host range vector pOGG026 (Chapter 3) which makes them ideal for experiments in environments where antibiotic selection is not possible.

The aim of this chapter is to test that each of the parts presented in previous chapters are working as a whole by tagging ninety-five *R. leguminosarum* strains using the Plasmid IDs library and performing a high-throughput competition assay with all the Tagged-strains using pea plants. Growing plant hosts in sterile conditions that are starved of nitrogen are not the most relevant conditions to analyse *Rhizobium* strains for their effectiveness and competitiveness. We therefore performed our competition assay in non-sterilised nutrient-rich field soil which contain indigenous rhizobia.

Identification of rhizobial nodule occupancy through Ion Torrent sequencing showed: i) nearly 45% of the nodules infected by a single clone were occupied by the same strain; ii) 15% of the total sequenced nodules were mixed nodules; iii) up to four different Tagged-strains were identified within individual mixed nodules; iv) under the NightOWL CCD camera, the most competitive strain produced nodules with the highest level of green fluorescence detection.

Results and discussion

Analysis of selected soil

Soil was collected from Yatesbury House Farm (Yatesbury, Calne, Wiltshire, UK) with the help of the owner Richard Gantlett. This is a 600-ha farm maintained under EU organic regulations and certified by The Soil Association Certification Limited (SA Certification).

Their crops of wheat, barley, oats, field beans and lentils are cultivated with legume-based mixtures of clover. Soil was analysed for chemical composition by James Hutton Limited (Aberdeen, Scotland) (Table 6. 1).

Table 6. 1 Chemical composition of Yatesbury soil

Sample	pH	Ca	K	Mg	P	NO_3	NH_4
Yatesbury soil	7.75	91601	226.5	390.7	115.6	64.02	13.05

With the exception of pH, values are expressed in [mg][kg]$^{-1}$

For assays to compare the competitiveness of different strains of *R. leguminosarum*, we wanted to have, at least approximately, the same number of cells from each of the strains tested (including indigenous rhizobia from the soil), so that they have the same opportunity, at least with respect to cell number, to compete for plant infection. For this reason, we determined the number of indigenous rhizobia with are symbionts of pea in Yatesbury soil. The cell number was estimated by the indirect plant infection method known as Most Probable Number (MPN) (Somasegaran & Hoben, 1994; Howieson & Dilworth, 2016) with pea plants as the trap host. Results from four soil

samples indicated that Yatesbury soil contained around 100 cells of indigenous rhizobia per gram.

Results showed that our pots contained 620.4 ± 1.955 gr of Yatesbury soil, meaning there were approximately $6.20x10^4$ cells of indigenous rhizobia per pot.

Ninety-five strains were selected from a well-characterised collection of *R. leguminosarum* bv. *viciae* (Rlv) strains obtained from different collaborators (listed in Table 6. 2). For this study, selection required meeting at least one of the followed criteria: i) they had been tested for competitiveness, ii) they had been analysed for effectiveness at fixing nitrogen or iii) we have genomic information (published or unpublished) from each strain.

Strains well-studied in our lab in competition assays

Fulfilling the first of our chosen criteria, strain Rlv UPM791, provided by Professor Jose Manuel Palacios from Centro de Biotecnología y Genómica de Plantas (CBGP) (Madrid, Spain), and strain Rlv3841, the model strain studied in our lab were selected not only because we had already tested them in a competition assay in Chapter 3 and found that they showed different competitiveness when assessed in sterile conditions, but also because their genomes are published (Sánchez-Cañizares *et al.*, 2018) (Young *et al.*, 2006).

Strains analysed for effectiveness at fixing nitrogen

Fulfilling the second of our criteria, are seventeen Rlv strains from Professor John Howieson from the Centre for Rhizobium Studies (CRS) (Perth, Australia). From this collection we selected WSM1475 and WSM1521, since we had previously studied

them and have data on their effectiveness (Chapter 4). Additional strains WSM1529, SU303 and WSM1488 were selected because they showed high, medium and low ARA rates in the laboratory (Table S. 1 in Supplemental Material). Strain WSM1481 was chosen as its complete genome is available at the Joint Genome Institute (JGI) database (Project ID: 404038).

The group of Dr Euan James from the James Hutton Institute (Dundee, UK) has tested more than one hundred different Rlv strains on peas in field conditions. Out of these, the twenty-two strains that showed the highest shoot dry biomass were selected. In addition, two commercial inoculants, CellTech and LegTech, were chosen, even though they were not among the best ranked bacteria in these field experiments (Euan James, personal communication).

Strains from the same soil population with phylogenetic analysis

With respect to strains fulfilling the final criteria, we used the collection of thirty-six Rlv strains obtained from Professor Peter Young (Department of Biology, University of York, UK). The strains were isolated from *V. sativa* nodules grown on a 1 m^2 area at the University of York. Kumar *et al.*, 2015 analysed 305 genes that were present in all the strains. From an alignment of all these genes among the species represented by a Neighbour Joining tree, they found five putative genospecies (gs). Rlv were only present in gsB, gsC and gsE, with gsC being the biggest group (Kumar *et al.*, 2015). For our study we selected a total of eight strains, one from each of the eight different sub-clusters within gsC (see Fig. S3, Kumar *et al.*, 2015). We did not take any strains from gsB since it was already represented by Rlv3841, nor from gsE, which had a high phylogenetic distance from gsB and gsC.

A collection of fifty-four Rlv strains (Jorrin & Imperial, 2015) was obtained from Professor Juan Imperial (CBGP, Madrid, Spain). These strains were selected from a population of 400 different bacteria isolated from the same soil (Institute for Agronomy Research Experimental Station of Époisses (Bretenière, Côte d'Or, France)) using four different plants: pea, lentil, broad bean and *V. sativa*. We chose this collection since the group of strains within each subpopulation (pea (G), lentil (L), faba bean (H) and vetch (V)) had the full genome sequenced (Jorrin Rubio, 2016).

All the selected strains (Table 6. 2) were tested for antibiotic resistance with gentamycin (Gm), neomycin (Nm), nitrofurantoin (Nit), streptomycin (St) and tetracycline (Tet) (concentrations are found in Table 2. 5). We found that all the strains showed Nit^r, with the exception of UPM791, which only had St^r. Furthermore, none showed Nm^r, which is the antibiotic carried by the vector into which Plasmid ID is cloned (Chapter 5). These results are of great importance since it allowed us to grow our Rlv strains in TY-Nit and UPM791 in TY-St^r and thus reduce the chance of contamination with *E. coli*.

Two biological replicates of all ninety-five selected rhizobial strains were tested for nodulation of pea plants grown in boiling tubes and sterile vermiculite after 21 dpi (Fig. 6. 1A). We observed that all ninety-five strains had formed nodules and in every case the nodules showed the typical pink colour indicative of an effective symbiosis (Fig 6. 1B).

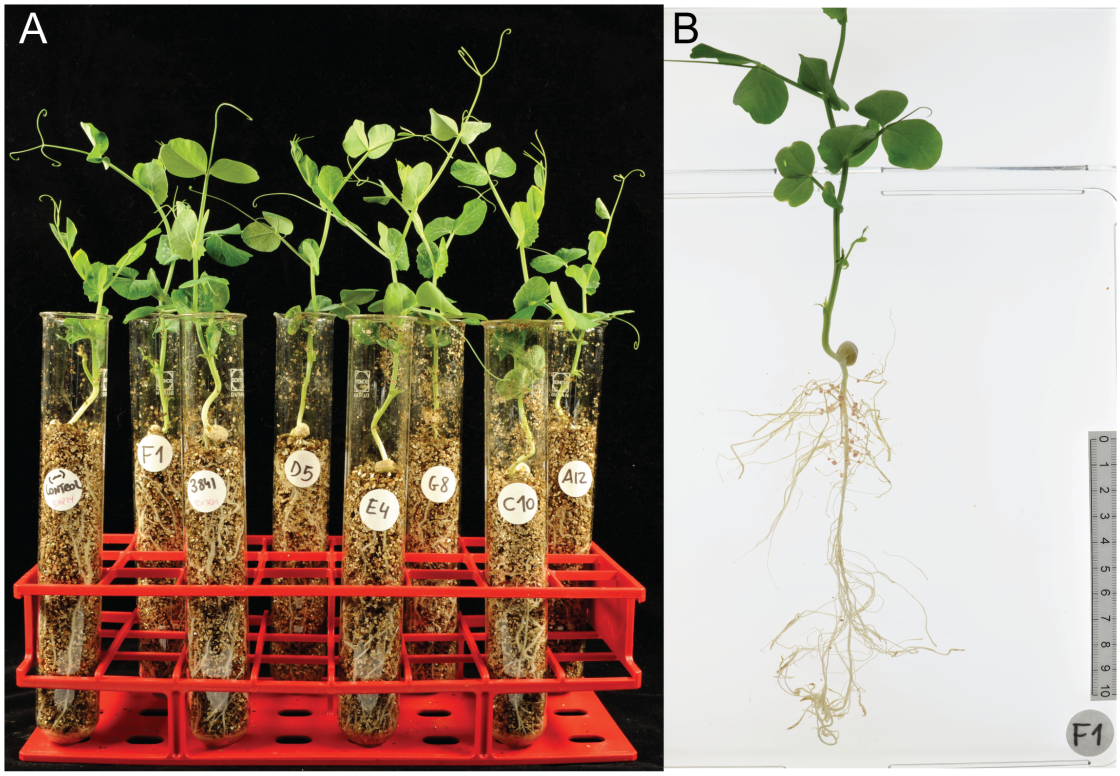


Figure 6. 1 A) Nodulation test in pea plants using ninety-five selected rhizobial strains (Table 6. 2). B) Example of a pea plant inoculated with a selected strain showing typical root nodules indicative of an effective symbiosis.

Based on these results we decided to grow and maintain for long-term storage these selected Rlv strains as a library (Rlv-library) in a 96-well plate in a -80°C freezer (Table 6.2).

Tagging Rlv-library with Plasmid IDs

We developed a new methodology for high-throughput conjugation of all ninety-five strains at the same time (for more details see Chapter 2). Each plasmid from the Plasmid IDs library (Table 6. 3) was conjugated into a strain from the Rlv-library (Table 6. 2) in its corresponding position of a 96-well plate. For example, Plasmid ID pOPS0491 (well A1) was conjugated into UPM791 of the Rlv-library (well A1).

Table 6. 2 Final selected Rlv strains (Rlv-library). Strains in yellow were studied in competition assays. Strains in blue were studied for effectiveness at fixing nitrogen in laboratory conditions, strains in grey were studied for effectiveness at fixing nitrogen in field conditions, strains in green have been phylogenetically analysed and strains in orange came from the same soil but were selected by different hosts.

	1	2	3	4	5	6	7	8	9	10	11	12
A	UPM791	Rlv3841	WSM1475	WSM1481	WSM1529	WSM1488	WSM1521	370	V060	V014	G011	VSX1
B	CellTech	JED2	364	PB6-3B	VF2	CORUS 16	CORUS 1	388	V068	V043	V002	VSX7
C	13	VF5	SU303	LegTech	G004	G077	G073	23--1	V067	H012	V004	VSX10
D	L102	L104	G007	G051	V116	G008	V006	387	V069	H009	V008	VSX11
E	24	E2-1B	PB3-3	536	42	PeanNW	367	BeanLA	V074	H127	V010	VSX16
F	G093	G099	G083	G109	G108	G094	L008	L019	V100	H130	V030	VSX25
G	H011	H082	H174	G028	G067	G088	L002	2B-1	G016	H178	V050	VSX28
H	L010	L018	L070	L074	L079	L082	L111	E2-1A	Control	H031	V057	VSX32

Table 6. 3 Plasmid IDs library (Chapter 5). All plasmids are identical, apart from the ID (a unique 12-nt, error-correcting Golay barcode) which is different in each case.

	1	2	3	4	5	6	7	8	9	10	11	12
A	pOPS0491	pOPS0492	pOPS0493	pOPS0494	pOPS0495	pOPS0496	pOPS0497	pOPS0498	pOPS0499	pOPS0500	pOPS0501	pOPS0502
B	pOPS0503	pOPS0504	pOPS0505	pOPS0506	pOPS0507	pOPS0508	pOPS0509	pOPS0510	pOPS0511	pOPS0512	pOPS0513	pOPS0514
C	pOPS0515	pOPS0516	pOPS0517	pOPS0518	pOPS0519	pOPS0520	pOPS0521	pOPS0522	pOPS0523	pOPS0524	pOPS0525	pOPS0526
D	pOPS0527	pOPS0528	pOPS0529	pOPS0530	pOPS0531	pOPS0532	pOPS0533	pOPS0534	pOPS0535	pOPS0536	pOPS0537	pOPS0538
E	pOPS0539	pOPS0540	pOPS0541	pOPS0542	pOPS0543	pOPS0544	pOPS0545	pOPS0546	pOPS0547	pOPS0548	pOPS0549	pOPS0550
F	pOPS0551	pOPS0552	pOPS0553	pOPS0554	pOPS0555	pOPS0556	pOPS0557	pOPS0558	pOPS0559	pOPS0560	pOPS0561	pOPS0562
G	pOPS0563	pOPS0564	pOPS0565	pOPS0566	pOPS0567	pOPS0568	pOPS0569	pOPS0570	pOPS0571	pOPS0572	pOPS0573	pOPS0574
H	pOPS0575	pOPS0576	pOPS0577	pOPS0578	pOPS0579	pOPS0580	pOPS0581	pOPS0582	Control	pOPS0584	pOPS0585	pOPS0586

All conjugated strains were grown in a 96-deep well plate in liquid TY-Nit, just UPM791 was grown in TY-St. Subsequently, we evaluated the fitness performance of conjugated strains by measuring optical density (Table 6. 4) at mid-exponential growth. Results showed that eighty-four strains had desired growth, indicated by green and yellow cells in Table 6. 4. We believe that the remaining eleven strains, indicated by red cells, did not show growth because of the harsh conditions of low oxygen concentration combined with the use of antibiotics. To eliminate possible spontaneous resistance to antibiotics, we performed colony PCR on all of the samples to verify that they had acquired the expected plasmid.

Table 6. 4. Evaluation of fitness performance of conjugated Plasmid IDs library into Rlv-library by measurement of OD_{600} at mid-exponential growth. Results are classified in graded coloured scale where green is the highest value, yellow is the midpoint percentile and red the minimum value.

	1	2	3	4	5	6	7	8	9	10	11	12
A	0.412	0.378	0.349	0.367	0.076	0.405	0.01	0.021	0.365	0.365	0.346	0.388
B	0.478	0.356	0.362	0.386	0.346	0.348	0.353	0.018	0.356	0.027	0.345	0.349
C	0.462	0.437	0.375	0.404	0.361	0.41	0.421	0.4	0.412	0.352	0.374	0.374
D	0.433	0.438	0.406	0.373	0.357	0.406	0.413	0.411	0.395	0.34	0.365	0.351
E	0.411	0.446	0.413	0.415	0.026	0.36	0.386	0.027	0.373	0.351	0.362	0.379
F	0.492	0.463	0.367	0.4	0.349	0.357	0.402	0.342	0.364	0.379	0.356	0.023
G	0.008	0.364	0.37	0.414	0.354	0.382	0.031	0.342	0.372	0.356	0.371	0.349
H	0.497	0.365	0.387	0.407	0.38	0.372	0.028	0.393	0	0.376	0.341	0.373

We decided to continue our experiment with these eighty-four strains since using 96-deep well plates and antibiotic selection because this is the only way to achieve high-throughput growth of bacteria without contamination. Final conjugated strains were grown and maintained for long-term storage in a -80°C freezer in a 96-well plate as a library (Tagged-strains) (Table 6. 5).

Table 6. 5 Final library of tagged Rlv strains (Tagged-strains) with Plasmid IDs. Empty cells do not contain any bacteria.

	1	2	3	4	5	6	7	8	9	10	11	12
A	UPM791 [pOPS0491]	Rlv3841 [pOPS0492]	WSM1475 [pOPS0493]	WSM1481 [pOPS0494]		WSM1488 [pOPS0496]			V060 [pOPS0499]	V014 [pOPS0500]	G011 [pOPS0501]	VSX1 [pOPS0502]
B	CellTech [pOPS0503]	JED2 [pOPS0504]	364 [pOPS0505]	PB6-3B [pOPS0506]	VF2 [pOPS0507]	CORUS 16 [pOPS0508]	CORUS 1+2 [pOPS0509]		V068 [pOPS0511]		V002 [pOPS0513]	VSX7 [pOPS0514]
C	13 [pOPS0515]	VF5 [pOPS0516]	SU303 [pOPS0517]	LegTech [pOPS0518]	G004 [pOPS0519]	G077 [pOPS0520]	G073 [pOPS0521]	23--1 [pOPS0522]	V067 [pOPS0523]	H012 [pOPS0524]	V004 [pOPS0525]	VSX10 [pOPS0526]
D	L102 [pOPS0527]	L104 [pOPS0528]	G007 [pOPS0529]	G051 [pOPS0530]	V116 [pOPS0531]	G008 [pOPS0532]	V006 [pOPS0533]	387 [pOPS0534]	V069 [pOPS0535]	H009 [pOPS0536]	V008 [pOPS0537]	VSX11 [pOPS0538]
E	24 [pOPS0539]	E2-1B [pOPS0540]	PB3-3 [pOPS0541]	536 [pOPS0542]		PeanNW [pOPS0544]	367 [pOPS0545]		V074 [pOPS0547]	H127 [pOPS0548]	V010 [pOPS0549]	VSX16 [pOPS0550]
F	G093 [pOPS0551]	G099 [pOPS0552]	G083 [pOPS0553]	G109 [pOPS0554]	G108 [pOPS0555]	G094 [pOPS0556]	L008 [pOPS0557]	L019 [pOPS0558]	V100 [pOPS0559]	H130 [pOPS0560]	V030 [pOPS0561]	
G		H082 [pOPS0564]	H174 [pOPS0565]	G028 [pOPS0566]	G067 [pOPS0567]	G088 [pOPS0568]		2B-1 [pOPS0570]	G016 [pOPS0571]	H178 [pOPS0572]	V050 [pOPS0573]	VSX28 [pOPS0574]
H	L010 [pOPS0575]	L018 [pOPS0576]	L070 [pOPS0577]	L074 [pOPS0578]	L079 [pOPS0579]	L082 [pOPS0580]		E2-1A [pOPS0582]		H031 [pOPS0584]	V057 [pOPS0585]	VSX32 [pOPS0586]

Competition assays among Tagged-strains in presence of indigenous rhizobia

The Tagged-strains library (Table 6. 5) was grown in liquid culture in a 96-deep well plate until late exponential phase. Cell density was estimated and each culture was adjusted, by appropriate dilutions, in order to have the same cell number as the indigenous rhizobia (~$6.20x10^4$ cells) contained in a 0.5L pot of Yatesbury soil. We inoculated pea seedlings in 0.5 L Yatesbury soil pots with the 1:1 inoculum and the same procedure was repeated in thirty pots. Seven pea plants were grown in pots without the Tagged-strains mixed as negative controls (-ve control).

At 21 dpi, pea plants were harvested (Fig. 6. 2A), and root systems were washed and placed in 10-cm square petri dishes (Fig. 6. 2B-C). Root systems photos were exposed to a blue-light transilluminator to check if nodules expressed green fluorescence belonging to Tagged-strains (Fig. 6. 2D-E). Results showed that all the inoculated plants had nodules formed by both, indigenous rhizobia, nodules without green fluorescence expression, and nodules formed by Tagged-strains expressing GFP under the control of psNifH, nodules with green fluorescence expression (Fig. 6. 2D). None of the seven negative controls showed any nodules with green fluorescence (Fig. 6.2 E).

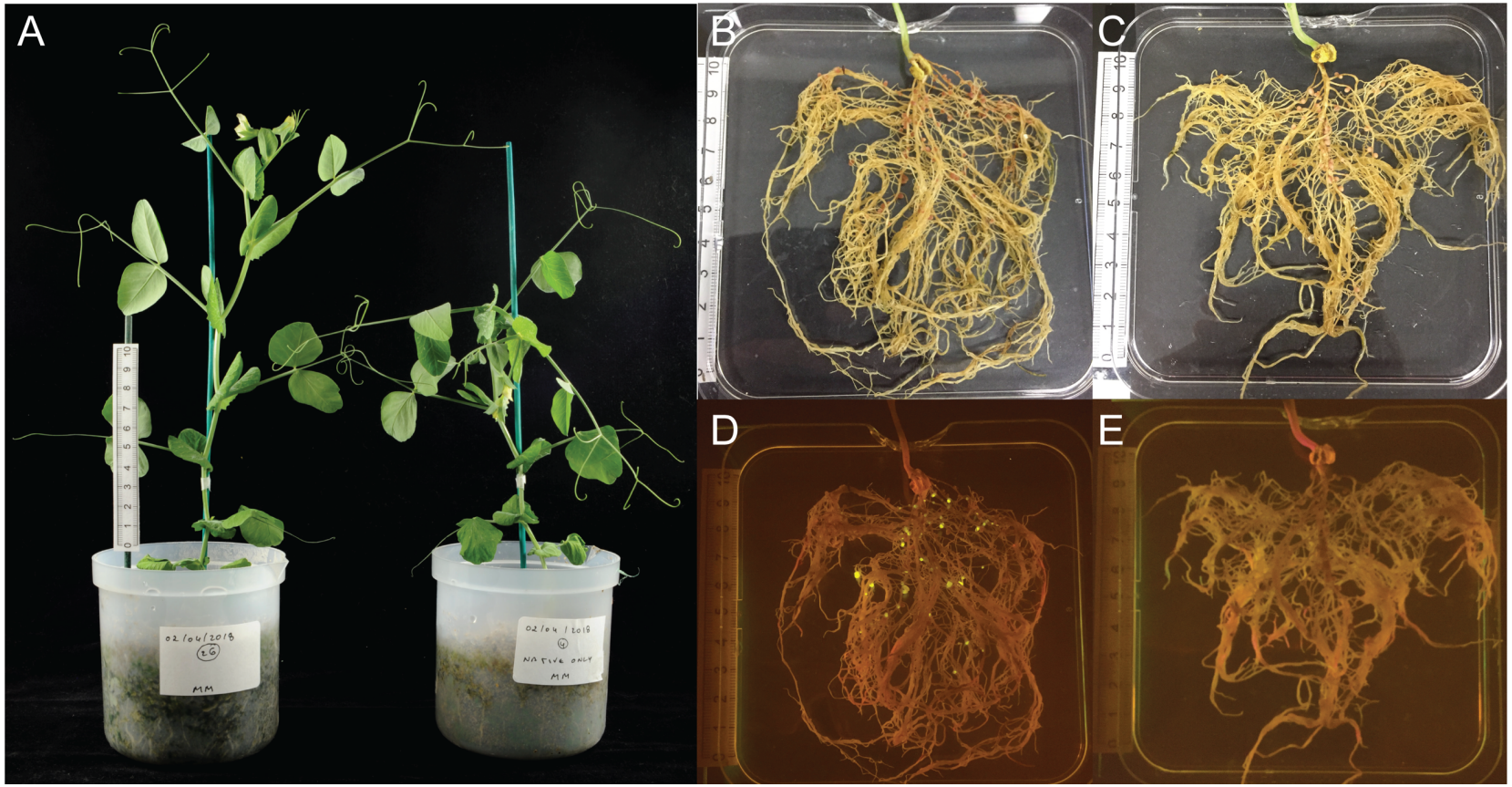


Figure 6. 2 A) Pea plants grown in 0.5 L Yatesbury soil pots. Example of plant at 21 dpi with Tagged-strain inoculum on the left and example of -ve control on the right. B-C) Root systems of inoculated (B) and -ve control (C) plants showing typical symbiotically-effective root nodules. D-E) Root systems exposed to a blue-light transilluminator to check green fluorescence from nodules formed by Tagged-strains expressing GFP under the control of psNifH. Photographs were taken with a camera fitted with a blue light filter. D) Plant inoculated with Tagged-strains showing nodules with green fluorescence. E) Nodules from -ve controls show no green fluorescence.

In Chapter 4 we showed that fluorescence from sfGFP expression can be used as an easily detected and quantifiable proxy for nitrogenase activity in individual nodules. Therefore, we decided to identify the eight brightest nodules from each plant as a selection of the most effective nodules.

Pea root systems were analysed with the NightOWL CCD camera accompanied by blue-light transilluminator as a high-throughput quantification of green fluorescence expression in nodules. Those expressing GFP were counted automatically using Indigo® software to identify the eight brightest nodules from each plant (Fig. 6. 3A). We picked the selected nodules and the rest of the plant was appropriately labelled and placed at -80 °C for long-term storage. All plants have been saved if analysis of all the nodules expressing GFP is desired.

We took a second NightOWL CCD camera photograph, this time just from the selected nodules in a specific well position in a 96-well plate with white walls to prevent fluorescence reflection (Fig. 6. 3B) and we took note of each nodule position, fluorescence expression and plant of origin.

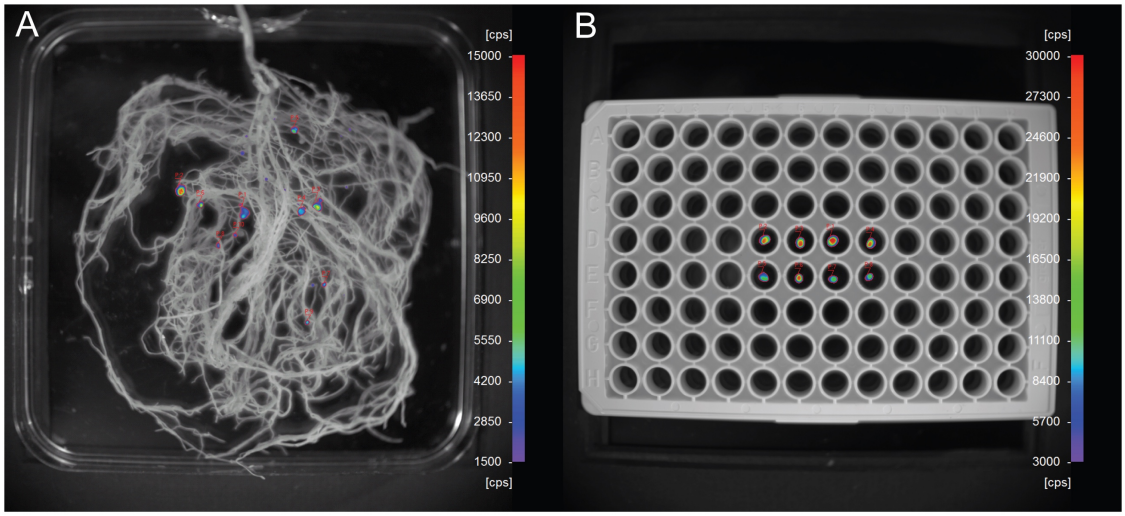


Figure 6. 3 A) Root systems were analysed to quantify levels of green fluorescence with a NightOWL CCD camera in order to identify the brightest nodules. B) Fluorescence was quantified of eight selected individual nodules in a plate reader.

The high-throughput identification of rhizobial isolates for nodule occupancy by Ion Torrent sequencing strategy presented in detail Chapter 5 was followed. A total of 232[6] nodules were surface-sterilised and DNA extractions were prepared for Ion Torrent multiplex sequencing. We used primers from Table 5. 2 and Table 5. 3 for two-step PCR and sequencing reads were identified for sample-specific and Plasmid ID. After high-quality filtering (where no mismatches were allowed) and reads normalisation, we crossed reference each nodule sample sequencing result with their specific fluorescence expression obtain from NightOWL CCD camera photos (Fig. 6. 3).

From sequencing analysis, we identified 197 nodules which had developed following a single clonal infection (Figure 6. 4A). We found thirty-five nodules with DNA sequences which have arisen from more than one strain (in fact up to four different strains in some nodules), implying infection with multiple strains (these are termed mixed nodules) (Figure 6. 5).

Results from a single clone infection

Single clonal infection results (Figure 6. 4A) showed that nearly 45% were occupied by the G083 (F3) strain alone. The next four strains to most commonly occupy a nodule with single clone infection were V116 (D5) with 13%, G076 (G5) with 6.5%, Rlv3841 (A2) with 5% and L008 (F7) with 4% of occupied nodules. The remaining 26.5% of nodules were a mixture of

[6] One of the 30 inoculated plants was harvested one day before to set-up the nodule screening/selection process. This plant was left out of sequencing analysis.

20 strains. We calculated competitiveness index (CI) of these top five strains (Figure 6. 4B) by comparing their own occurrence to the number of nodules analysed per plant:

$$CI = \frac{n_A}{n_T}$$

$$CI = competitiveness$$

$$n_A = nodules\ occupied\ by\ strain\ A$$

$$n_T = [total\ number\ of\ nodules\ analysed][plant]^{-1}$$

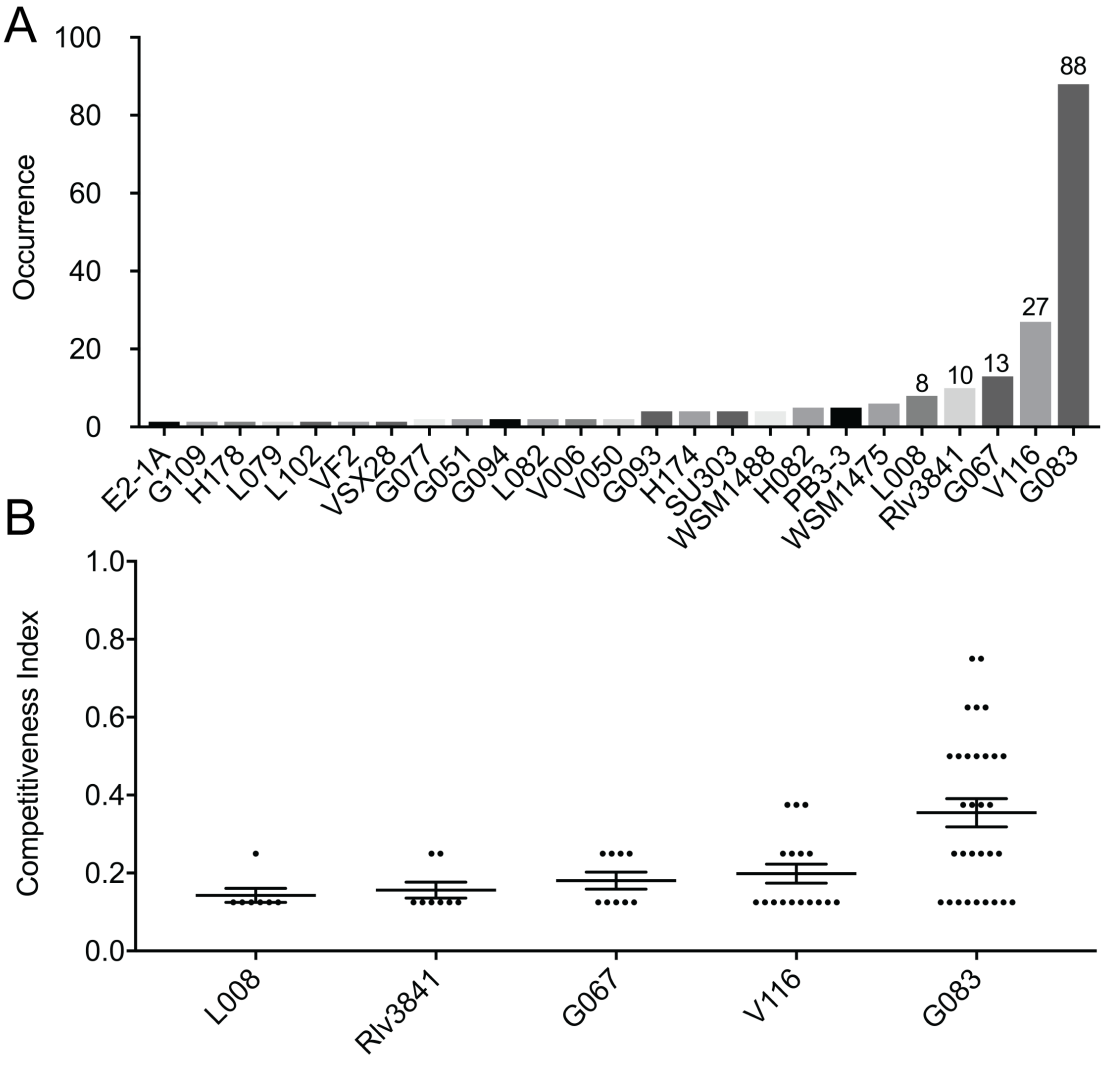


Figure 6. 4 A) Tagged-strains identified in single clone infection organised by occurrence from a total of 197 infections. The occurrence of the five strains found most often is shown above their respective columns. B) Competitiveness index (CI) of the top five strains. Dots represent CI per plant and bars represent the SEM.

It is important to mention that CI only considers single clonal infections, although we do not expect a drastically change when accounting for mixed nodules since the percentage of these is low. This analysis to determine competitiveness for nodule infection is one of the most preferred (e.g. Frederix *et al.*, 2014; Zhang *et al.*, 2014; Ji *et al.*, 2017) and results can be presented as a fraction or percentage.

Although other ways of analysing competitiveness are found in the literature, they are relevant when a different inoculum proportion is used (Amarger & Lobreau, 1982) or when just two strains are competing at the same time (Sessitsch *et al.*, 1997; Onishchuk *et al.*, 2017). These analyses are not appropriate for our experiment since we ensured all strains involved had similar proportion in mix inoculum, and we have eighty-five different strains in our inoculum.

After ordinary one-way ANOVA, a Tukey's multiple comparisons test results from CI shown that G083 was significant difference from the rest of the strains (full statistical analysis in Supplemental Material Table S. 3) and did not detect a statistically significant difference among V116, G067, Rlv3841 and L008, we cannot, therefore, categorise competitiveness of strains after G083 (Figure 6. 4B), even when there is a difference in the total nodule occurrence (Figure 6. 4A).

Soil pH has been shown to affect the competitiveness of *R. leguminosarum* strains (Vargas & Graham, 1989; Frey & Blum, 1994) and since G083 was trapped by the same host used in our competition assay, pea, and in a soil with very similar pH (Jorrin Rubio, 2016) than Yatesbury soil, this may explain its high competitiveness in our experiment. What we cannot explain at the moment is why G083 is more competitive than the rest of the sixteen strains from the same (G) collection. This is something interesting which could be explored further using the genomic information of the sequenced genomes obtained by Jorrin Rubio, 2016.

Having just one super competitive strain can be related to the process of autoregulation of nodulation (AON) triggering pathways involving long-distance root-shoot signalling inhibiting further nodulation events (Ferguson *et al.*, 2010; Mortier *et al.*, 2012; Ryu *et al.*, 2012). We can therefore presume that from the moment G083 began to form all the nodules that the plant can support at this stage, the plant then started to block the less competitive strains from making more nodules, or at least during this early stage of plant growth.

Nodules containing more than one strain (mixed infection)

In the case of mixed nodules (Figure 6. 5), the strains that showed higher occurrence were G077 (C6) which appeared 24 times, G109 (F4) 21 times and H031 (H10) 19 times.

From the thirty-five mixed nodules, among the most frequent combinations were, G077/G109 and G109/H031 with 6 times each, and G077/G109/H031 with 5 times. It is important to highlight that these strains were trapped from the same soil as G083 (the most competitive among single clonal infections) where G077 and G109 were also trapped by pea (as G083), and H031 by faba bean.

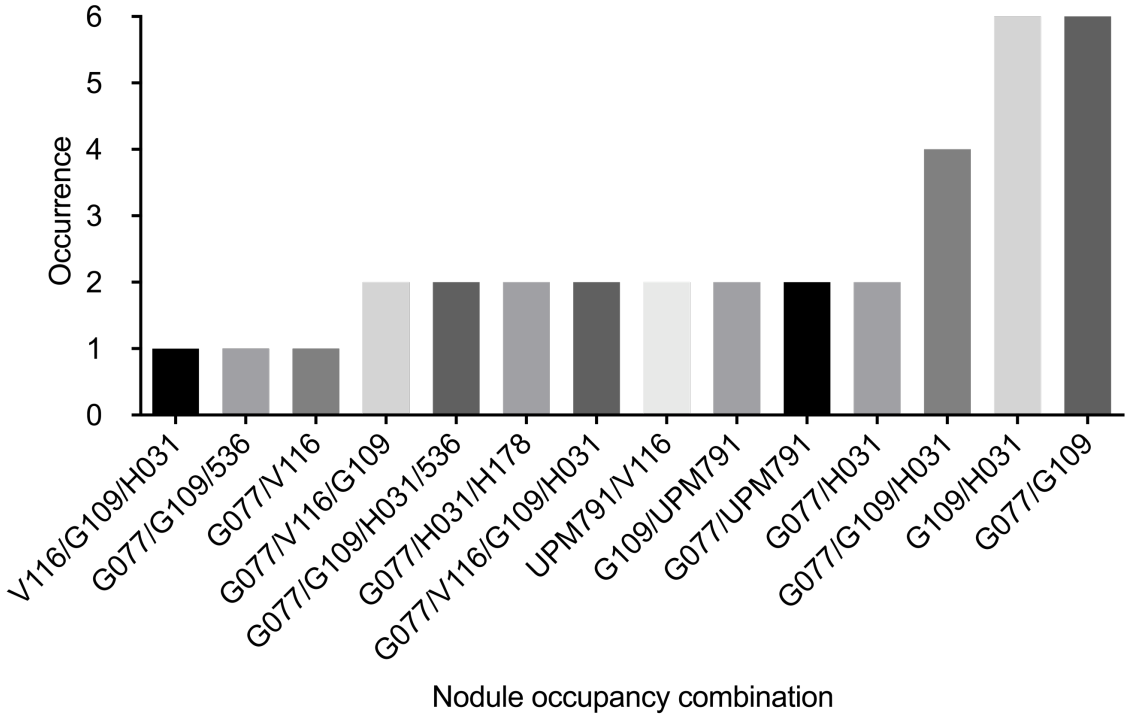


Figure 6. 5 Mixed nodules organised by occurrence of nodule occupancy combination. Nodule occupancy by a combination of up to four strains was observed.

Mixed nodules represent 15% of the total sequenced nodules. This percentage is much higher than the ~1% that we observed in the two competition assays where only two-strains were inoculated in sterile vermiculite/sand (Chapter 3 and Chapter 5). We believe that the high number of strains competing in this experiment (eighty-four Tagged-strains) increased the chance of having more than two cells of different rhizobia present at the time that root hairs are curled and infection threads (ITs) are initiated (Oldroyd *et al.*, 2011). Although our observation of 15% of mixed nodules is not as high as the 36% reported in lentils (May & Bohlool, 1983), there are a number of possible reasons for this. It may be that some of the nodules that only contain one Tagged-strain in our experiment do in fact also contain indigenous rhizobia that we simply did not identify (because they were not tagged and so NGS did not detect them). The difference may also be driven by the fact that we are using a

different *Rhizobium* strain and a different legume genotype. We also use a different method to identify rhizobial strains. Additionally, our method of strain identification after sequencing analysis is very conservative since we do not allow mismatches in either sample or Plasmid ID sequences, and we remove all reads below 1% of the total number of unique sequences. Furthermore, in our experiment, the inoculum was distributed homogenously within the pot so the bacterial competition starts in the rhizosphere, in contrast to May & Bohlool, (1983) where seeds received the inoculum.

More interestingly, even with such a conservative strain identification method, nodule occupancy with a combination of up to four strains was observed. More specifically, four nodules occupied by four strains and nine nodules occupied by three strains (Fig. 6. 5). After an exhaustive literature search, we can say that this is the first study where more than two strains have been identified from mixed nodules.

It has been presented previously by Denison, 2000 and Kiers *et al.*, 2003 that mixed nodules might reduce the success of whole-nodule sanctions by the plant. If these multi-strain mixed nodules are more prevalent than previously thought, then plant sanctioning of whole-nodules could be even more difficult to achieve, which could be one of the reasons why rhizobial parasitism occurs.

Another fascinating result is that the most competitive strain, G083, was not found in mixed nodules. One possibility for the competitive advantage of strain G083 is that it may migrate to root hairs and initiate infection much faster than any other tested strains. This result could be related to communication between the first strain that enters in the IT and the plant itself, triggering an immune response that obstructs entry of further strains (Mellor & Collinge, 1995). For instance, necrosis of cortical cells of ITs in alfalfa is observed after *S. meliloti* infection (Vasse *et al.*, 1993). Alternatively, it may have active mechanisms of

preventing other strains from infecting, or of disabling them in a competitive race. Examples of such mechanisms include a Type VI secretion system that might inject inhibitors into competing rhizobia (Gallique *et al.*, 2017). Interestingly one of the first identified Type VI secretion systems was in *R. leguminosarum* Rlv3841 (Bladergroen *et al.*, 2003). It will be very to interesting compare the G083 genome to the genomes of other strains used in this experiment to look for potential competition factors. It may also be possible to use global mutagenesis techniques such as INSeq (as used in our laboratory), to identify genes required for competition by GO83 competing against itself and against other strains (Wheatley *et al.*, 2017).

Another interesting observation from the use of multiple rhizobial strains in an agricultural soil, is that one of the most competitive strains was Rlv3841 and not UPM791. In previous studies of the two strains competing (Sánchez-Cañizares & Palacios, 2013) and repeated here in Chapter 3, UPM791 consistently out competes Rlv3841 for nodule occupancy under sterile plant growth conditions. This emphasises how much more complex true community competition is in agricultural soil compared to sterile sand/vermiculite. However, it also emphasises how exciting it will be to repeat the multi-strain competition in soil reported in this chapter but under sterile sand/vermiculite conditions. This will help determine how much of a competitive advantage is just due to intra-rhizobial strain competition versus how much is due to interaction with the soil environment.

The set-up of this experiment dictates that identification of the most competitive strain also identifies the most effective since the eight brightest nodules per plant were selected. However, we wanted to further investigate the nodule behaviour of each strain in terms of GFP expression from sequenced nodules. For this reason, we plotted quantification of GFP fluorescence [CPS] as individual nodules, without considering from which plant the nodules

were selected. Figure 6.6 shows the fluorescence from nodules of the five strains with the highest occurrence in single clonal infections and the three nodule occupancy combinations that most frequently occurred.

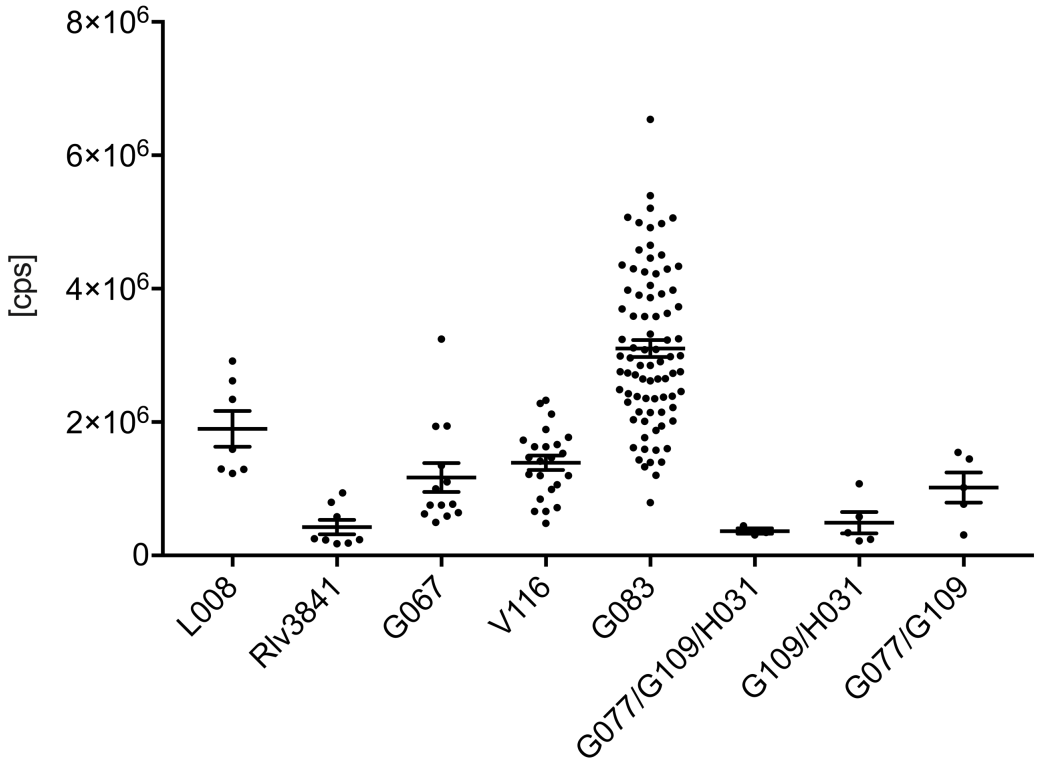


Figure 6. 6 Quantification of fluorescence in nodules due to GFP detection. Values in Counts per Second (CPS). Each nodule is represented by a dot and bars represent the SEM.

After ordinary one-way ANOVA, a Tukey's multiple comparisons test results showed that G083 was the strain with the most GFP fluorescence. There was no statistically significant difference among the rest of the comparisons (full statistical analysis in Supplemental Material Table S. 4). Therefore, we can be confident to say that G083, besides of being the most competitive amongst the strains, using a proxy for nitrogenase activity, also produces the most effective nodules.

By examining only the eight most fluorescent nodules from each root may mean that the percentage of mixed nodules (Fig. 6. 6) is underestimated since such infections might not necessarily form very bright nodules. This could be another reason for the lower percentage

of mixed nodules observed compared to other studies (May & Bohlool, 1983; Moawad & Schmidt, 1987). Although the set-up of the experiment was designed to select the most effective nodules (i.e. by taking the brightest eight nodules), since it is possible to observe GFP expression after plants have been stored at -80^{o}C and thawed (Supplemental Material Figure S. 2), in future work we will explore the true proportion of mixed nodules occupied by Tagged-strains and analyse the competitiveness index of each Tagged-strain from the full plant. Nodules from stored plants will also be used for sequencing to determine total strain occupancy.

Conclusions

In this chapter we selected ninety-five pea symbiotic, effective *R. leguminosarum* strains and tagged them simultaneously with our Plasmid IDs library using a high-throughput conjugation method. We had a final selection of eighty-four Tagged-strains that showed desired fitness performance in the 96-deep well plate growth system.

Using a high-throughput competition assay with pea plant hosts and a non-sterilised nutrient-rich agricultural soil occupied by indigenous rhizobia, we identified strain G083 as the most effective and competitive amongst the Tagged-strains, representing nearly 45% of all the single clonal infection nodules and 38% of the total sequenced nodules. Interestingly, G083 was not found in mixed nodules.

Mixed nodules, which represented nearly 15% of the total sequenced nodules, were mostly occupied by strains G077, G109 and H031. The excellent performance of G083, G077, G109 and H03 in this competition assay fits with the fact that they were trapped from the same soil with similar pH to the Yatesbury soil (Vargas & Graham, 1989; Anyango *et al.*, 1998) .

For a long time, it was accepted that nodules were formed by single clonal infection but experimental results have challenged this idea (Johnston & Beringer, 1975; May & Bohlool, 1983; Dowling & Broughton, 1986; Moawad & Schmidt, 1987), now using this molecular tool, it was possible to show for the first time that rhizobial nodule occupancy can be shared by as many as four Tagged-strains strains (which is possibly an underestimation as we know there are indigenous rhizobia was not detected by the NGS).

There is a general understanding in our field that effectiveness at nitrogen fixation and competitiveness for nodulation are two rhizobial traits that are not always linked. However, due to the complexity of competition assays there are few studies evaluating these two traits in the same experiment, and strains that contain both traits have therefore seldom been identified. This tool can help to identify more strains with both traits since we can now set to compete a much higher number of strains at the same time.

Chapter 7: General discussion and future perspectives

Each of the previous chapters contained their own relevant discussion. Here, I will briefly review the key results and the broader relation with future *Rhizobium* research.

The aim of this thesis was to engineer a molecular tool for the high-throughput identification of competitive and effective rhizobial strains. Major challenges in *Rhizobium*-legume research include the development of tools to study the spatial-temporal performance of rhizobial strains in the nodule and during earlier stages of symbiotic development. Transcriptomic, proteomic, and metabolic methods can help provide insights into the unique developmental stages (Ramachandran *et al.*, 2011; Czernic *et al.*, 2015; Wheatley *et al.*, 2017), and tools such as the bacterial biosensors from Pini et al., 2017 help for systems-level analyses. Ultimately though, rhizobial strains are not alone in the rhizosphere, and the rhizosphere is not an inert-sterile zone. It is therefore essential to build tools that combine the study of a large number of rhizobial strains for competition for nodule occupancy and symbiotic N_2 in agricultural soils whilst minimising financial and time costs with high test efficiency.

Through the use of Golden Gate Cloning and NGS we present here: i) the promoter-probe vector pOGG026, suitable for high-throughput cloning and studies where antibiotic selection is not possible or desirable; ii) the nodule-specific promoter psNifH which can be used as a proxy for nitrogenase activity assays in biovars and strains of *Rhizobium leguminosarum*; iii) a library of Plasmid IDs which can used in a competition assay for high-throughput identification of rhizobial isolates; and iv) a competition assay with the largest number of rhizobial strains (eighty-four Tagged-strains) performed to date.

We developed pOGG026 and showed it has excellent stability in rhizobia, even in the absence of antibiotics. This vector was not only crucial for this project due to its flexible and cost-effective multi-fragment assembly, but it has also been used within our group and by collaborators because of its excellent performance in the rhizosphere. For example, reporter plasmids under the control of a nodule-specific promoter with mCherry and sfGFP were used to mark Rlv3841-WT and Rlv3841-Δ*nifH* for nodule sanctioning analysis. In the group of Professor Miriam Gifford at The University of Warwick, our reporter plasmid constitutively expressing sfGFP has had excellent results for the screening of *S. medicae* WSM419 in competition experiments against indigenous rhizobia in different soil types.

The application of pOGG026 has permitted development of a plasmid-based system to mark bacteria with GusA/CelB and has been used to evaluate rhizobial competitiveness in forming nodules on legume roots. Pink and blue nodules have been counted manually, but in collaboration with Dr Charlotte Kirchhelle, from the Plant Sciences Department at the University of Oxford, we developing an automated nodule counting system using the open-source image analysis Fiji (Schindelin *et al.*, 2012). This new system also aims to analyse nodule size and identify mixed nodules. If the photographs are taken with the same settings, this software can be trained to detect proportions of blue, green and red with the aim of quantifying strain occupancy in mixed nodules, even when the pink and blue are not observed on the surface of the nodule. As a rapid and inexpensive methodology, the availability of this plasmid-based system may especially benefit researchers working with *Rhizobium*-legume symbiosis in developing countries.

Since our molecular tool is intended for the study of a large number of rhizobial strains, it was essential to standardise, as far as possible, nodule-specific expression in a broad range of *R. leguminosarum* strains. Using synthetic biology and sequence data available in public databases we have designed a consensus promoter (psNifH) to drive nodule-specific expression across a variety of *R. leguminosarum* biovars and strains. One of the limitations of this design is that it is not suitable for rhizobial species with a different *nifH* promoter and in these cases a new concensus promoter will therefore need to be determined. However, each component of these reporter plasmids - promoter, marker genes, barcodes and backbone vector - were constructed as modular segments for Golden Gate cloning. This allows flexibility to rapidly assemble new reporter plasmids with different components to analyse different rhizobial strains. Additionally, we want modules to be shared amongst the community by creating novel and useful parts to be added over time by an Open-Source approach. We have already taken the first step in this task by making all the components for the bacterial vector assembly system (described in Chapter 3) available in Addgene, and we will continue with this task with the remaining reporter plasmids.

Since N_2 fixation happens inside nodules, several indirect measurements of rhizobial effectiveness have been developed (Howieson & Dilworth, 2016). In this work, we used quantification of fluorescence from sfGFP expression under the psNifH promoter as a proxy for nitrogenase activity. Although we acknowledge that this is not the most accurate measurement of rhizobial effectiveness, the key is that this quantification of fluorescence can be used as a high-throughput and low-cost biosensor for analysis of individual nodules, making it a practical tool. We have quantified GFP using the NightOWL CCD camera, but to transfer these reporter plasmids to laboratories without this equipment, further research

will be necessary to determine if nitrogenase activity still correlates with their GFP quantification. We have already made the first improvements to ensure this strategy is low cost and widely available, by combining a blue-light transilluminator, a mobile camera (smartphone) and an orange filter normally used to take Black & White photographs (Figure 7.1). The photographs have now been greatly improved, making it possible to do a first test for nodule size and florescence intensity using Fiji software for image analysis.



Figure 7.5 Pea root systems 28dpi. Top photograph obtained using a mobile phone camera. In the bottom photograph roots were exposed to a blue-light transilluminator and the photograph was obtained with the mobile phone camera accompanied by an orange filter. Nodules expressing green fluorescence occupied by UPM791 [pOPS0491] and non-fluorescent nodules occupied by indigenous rhizobia from Yatesbury soil

A new low-cost and widely available strategy will integrate reporter plasmids and PCR techniques for a rapid strain identification strategy. The proposed idea (Figure 7.2) is to add synthetic long reads (i.e. 50bp difference) to reporter plasmids instead of the 12-nt specific sequence ID. Subsequently, WT strains would be conjugated with these different size Reporter plasmids and used for competition assays. Nodules would be selected with the mobile phone imaging strategy (Figure 7.1) combined with Fiji analysis for GFP expression. After PCR and DNA gel electrophoresis, strains would be identified by fragment size. Although this strategy will reduce the number of strains that can compete (typically 4-8 depending on the number of different fragment sizes synthesised), the benefit is that expensive sequencing will not be required. However, even the limitation of reduced sample number can be overcome by several alternatives such as adding different primer binding sites to the ends of the synthetic long reads to form several families of amplicons that amplify with different primer pairs.

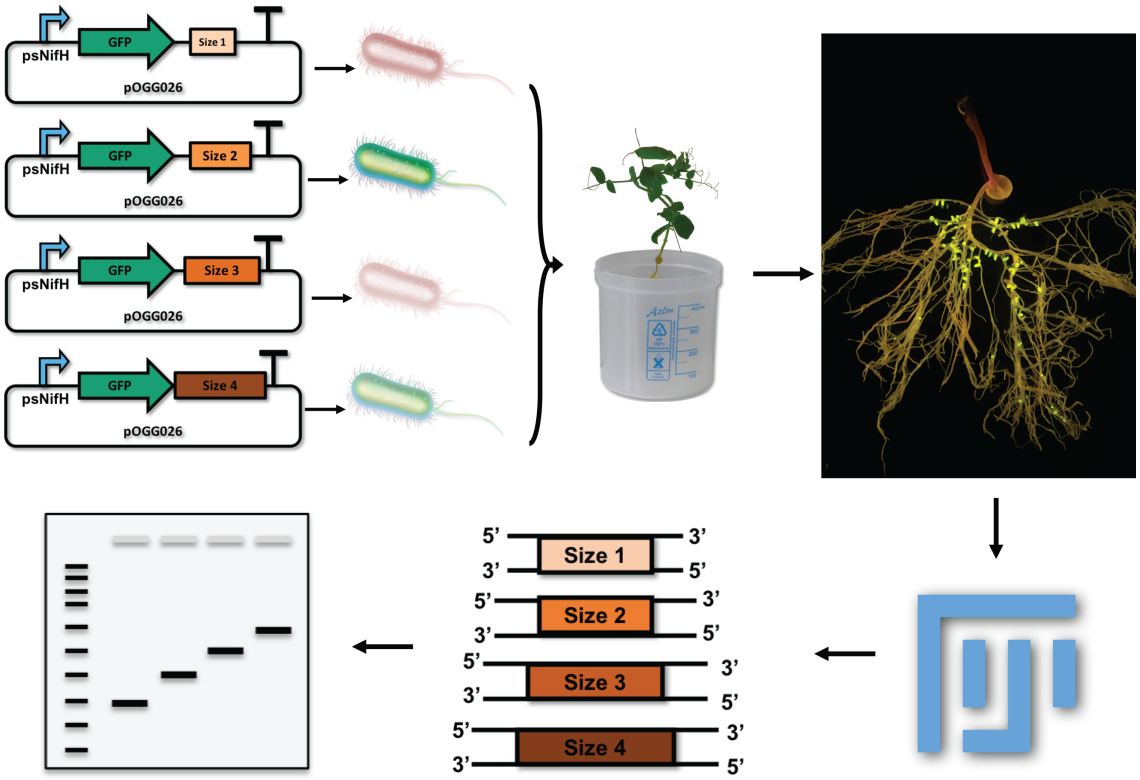


Figure 7.6 Workflow of the new proposed strategy for strain identification by PCR fragment size.

For very large sample sizes, pooling of samples can also be contemplated in a two-step PCR strategy. Second-step primers could be complemented with different length pad sizes and image gel systems software used. For example, using Image Lab™ Software from GelDoc EZ System (BioRad), slight differences in bp sizes were detected. In Figure 3, the first band in column 2 was identified with a size of 331bp and the second band with 297bp.

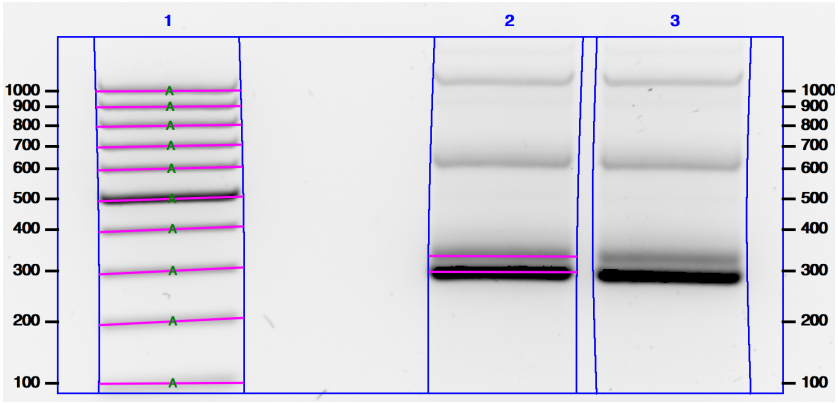


Figure7. 7. Column 1 contains a 100bp molecular ruler, columns 2 and 3 contain the same samples. Column 2 was analysed for molecular weight of bands, signalled by pink lines. Difference in band size can be observed in column 3.

Fluorescence expression under the control of psNifH not only allowed us to use high-throughput screening at a nodule level, but was also extremely helpful for confocal microscopy, allowing us to produce high resolution visualisations of the 3D structure of full bacteroids. This has led to a number of collaborations, including with Dr Charlotte Kirchhelle comparing bacteriod packing in determinate and indeterminate nodules, and with Professor Chris Hawes from the Department of Biological and Medical Sciences at Oxford Brookes University and Dr Euan James from the James Hutton Institute, studying the 3D structure of nodule sectioning using 3D Electron Microscopy.

Metagenomics approaches for high-throughput identification, such as the small-subunit ribosomal RNA (16S rRNA) gene, are not possible to use at a strain level due to their low sensitivity or replicability. Rhizobia identification at a strain level can be done by PCR DNA fingerprinting of genes (e.g. *nodD, thrC and ropA*) (Richardson *et al.*, 1995; Laguerre *et al.*, 1996; Zhang *et al.*, 2012) but the disadvantages include poor reproducibility, the fact that it requires strict standardisation of reaction parameters and, most importantly, it can only be

used with sequenced strains. Therefore, using a 12-nt specific sequence as Plasmid ID to mark *R. leguminosarum* strains and identify them directly from nodules by NGS, as done here, is the best option for un-sequenced strains.

Plasmid IDs were used to perform a competition assay using a collection of eighty-four Tagged-strains, in non-sterilised agricultural soil containing indigenous rhizobia. Interestingly, from this collection we identified an effective and super competitive strain G083. Although we were not expecting to find just one super competitive strain, further analysis of mixed nodules, where G083 was not found, led us to believe that this strain may be migrating to root hairs and initiating infection much faster than any other Tagged-strains and therefore forming all the nodules that the plant can support at this early stage. Additionally, the reason G083 is found alone could be related to the immune response of the plant, which starts necrosis of cortical cells of ITs after the first strain enters. Furthermore, G083 could activate a Type VI secretion system, conferring advantage during the rhizobial competition race. These possibilities should be investigated in future work by fluorescent tagging of G083 and other strains with different fluorescent markers and performing competition assays. This could potentially be obtained by combining Plasmid IDs with mini-Tn7 reporter plasmids developed in our group, where multiple coloured proteins allow different strains to be determined. Experiments can be designed using G083 and different combinations of strains to see if G083 really prevents all other strains from entering nodules by exploring each stage of colonisation of root surfaces and root hairs tracked over multiple time points using confocal microscopy. Further experiments can be run using the same inoculum mix used in Chapter 6, but this time in a soil with a different pH to analyse if G083 is still the most competitive. Last but not least, accurate effectiveness

analysis (ARA and/or dry weight) of G083 in a single inoculation needs to be done to really analyse its potential for field trials.

Alongside identification of this super competitive strain, sequencing of Plasmid IDs is so sensitive that it allowed us to show, for the first time, that nodules can have multi-strain occupancy with up to four Tagged-strains. Further studies can use GFP expression and multi-strain occupancy identification to assess sanctioning in mixed nodules at the plant cell and bacteroid level.

Finding elite indigenous rhizobia is advantageous since they avoid the legal and socioeconomic problems associated with genetically modified strains. Until now the search for such strains has been a time-consuming process involving numerous studies in the laboratory before the strains even can be tested in greenhouse and field trials. Here however, for the first time, a competition assay using agricultural soil and eighty-four rhizobial strains has been performed and a super competitive strain has been identified, which can potentially now be used, as a WT strain without Plasmid ID, in field trials. This means that this molecular tool has substantially reduced the laboratory and greenhouse studies required to advance to field trials, saving time and money.

Now that we have this tool it is time to consider how it can be transferred to other *Rhizobium* species, how to improve its accuracy and even how to integrate these findings into coordinated multidisciplinary studies with ecologists, agronomists, policy makers, economists and others. At the same time, we need to consider how to make it even more transferable, so laboratories and farmers in developing countries can use it, and how we can use it to produce tailor-made inoculants where commercial inoculants have previously not worked due to different soil conditions and different cultivars.

References

Agmon N, Mitchell LA, Cai Y, Ikushima S, Chuang J, Zheng A, Choi W-J, Martin JA, Caravelli K, Stracquadanio G, *et al.* 2015. Yeast Golden Gate (yGG) for the Efficient Assembly of *S. cerevisiae* Transcription Units. *ACS Synthetic Biology* **4**: 853–859.

Allaway D, Lodwig EM, Crompton LA, Wood M, Parsons R, Wheeler TR, Poole PS. **2000**. Identification of alanine dehydrogenase and its role in mixed secretion of ammonium and alanine by pea bacteroids. *Molecular Microbiology* **36**: 508–515.

Amarger N, Lobreau JP. **1982**. Quantitative study of nodulation competitiveness in Rhizobium strains. *Applied and Environmental Microbiology* **44**: 583–588.

Anyango B, Wilson K, Giller K. **1998**. Competition in Kenyan soils between Rhizobium leguminosarum biovar phaseoli strain Kim5 and R. tropici strain CIAT899 using the gusA marker gene. *Plant and Soil* **204**: 69–78.

Bais HP, Weir TL, Perry LG, Gilroy S, Vivanco JM. **2006**. The role of root exudates in rhizosphere interactions with plants and other organisms. *Annual review of plant biology* **57**: 233–266.

Barbier de Reuille P, Routier-Kierzkowska A-L, Kierzkowski D, Bassel GW, Schüpbach T, Tauriello G, Bajpai N, Strauss S, Weber A, Kiss A, *et al.* 2015. MorphoGraphX: A platform for quantifying morphogenesis in 4D. *eLife* **4**.

Bedoussac L, Journet E-P, Hauggaard-Nielsen H, Naudin C, Corre-Hellou G, Jensen ES, Prieur L, Justes E. **2015**. Ecological principles underlying the increase of productivity achieved by cereal-grain legume intercrops in organic farming. A review. *Agronomy for*

Sustainable Development **35**: 911–935.

Beringer JE. **1974**. R factor transfer in *Rhizobium leguminosarum*. *Microbiology* **84**: 188–198.

Berry D, Ben Mahfoudh K, Wagner M, Loy A. **2011**. Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. *Applied and environmental microbiology* **77**: 7846–9.

Bertin C, Yang X, Weston LA. **2003**. The role of root exudates and allelochemicals in the rhizosphere. *Plant and Soil* **256**: 67–83.

Bladergroen MR, Badelt K, Spaink HP. **2003**. Infection-Blocking Genes of a Symbiotic *Rhizobium leguminosarum* Strain That Are Involved in Temperature-Dependent Protein Secretion. *Molecular Plant-Microbe Interactions* **16**: 53–64.

Blatny JM, Brautaset T, Winther-Larsen HC, Haugan K, Valla S. **1997**. Construction and use of a versatile set of broad-host-range cloning and expression vectors based on the RK2 replicon. *Applied and Environmental Microbiology* **63**: 370–379.

Bohlool BB, Ladha JK, Garrity DP, George T. **1992**. Biological nitrogen fixation for sustainable agriculture: A perspective. *Plant and Soil* **141**: 1–11.

Bolger AM, Lohse M, Usadel B. **2014**. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

Bontemps C, Elliott GN, Simon MF, Dos Reis Júnior FB, Gross E, Lawton RC, Neto NE, De FÁtima Loureiro M, De Faria SM, Sprent JI, *et al.* 2010. *Burkholderia* species are ancient symbionts of legumes. *Molecular Ecology* **19**: 44–52.

Bragg LM, Stone G, Butler MK, Hugenholtz P, Tyson GW. **2013**. Shining a Light on Dark Sequencing: Characterising Errors in Ion Torrent PGM Data (N Beerenwinkel, Ed.). *PLoS*

Computational Biology **9**: e1003031.

Brito B, Martínez M, Fernández D, Rey L, Cabrera E, Palacios JM, Imperial J, Ruiz-Argüeso T. **1997**. Hydrogenase genes from *Rhizobium leguminosarum* bv. *viciae* are controlled by the nitrogen fixation regulatory protein NifA. *Proceedings of the National Academy of Sciences of the United States of America* **94**: 6019–6024.

Bromfield ESP, Thurman NP, Whitwill ST, Barran LR. **1987**. Plasmids and Symbiotic Effectiveness of Representative Phage Types from Two Indigenous Populations of Rhizobium meliloti. *Microbiology* **133**: 3457–3466.

Broughton WJ, Dilworth MJ. **1971**. Control of leghaemoglobin synthesis in snake beans. *The Biochemical journal* **125**: 1075–80.

Broughton WJ, Perret X. **1999**. Genealogy of legume-Rhizobium symbioses. *Current Opinion in Plant Biology* **2**: 305–311.

Butterbach-Bahl K, Gundersen P, Ambus P, Augustin J, Beier C, Boeckx P, Dannenmann M, Gimeno BS, Ibrom A, Kiese R, *et al.* 2011. *The European Nitrogen Assessment* (MA Sutton, CM Howard, JW Erisman, G Billen, A Bleeker, P Grennfelt, H van Grinsven, and B Grizzetti, Eds.). Cambridge: Cambridge University Press.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, *et al.* 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal* **6**: 1621–1624.

Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC. **1994**. Green fluorescent protein as a marker for gene expression. *Science (New York, N.Y.)* **263**: 802–5.

Chomczynski P, Rymaszewski M. **2006**. Alkaline polyethylene glycol-based method for

direct PCR from bacteria, eukaryotic tissue samples, and whole blood. *BioTechniques* **40**: 454, 456, 458.

Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, *et al.* 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**: 1422–1423.

Conway GR, Barbie EB. **1988**. After the Green Revolution: Sustainable and equitable agricultural development. *Futures* **20**: 651–670.

Coordinators NR. **2013**. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **41**: D8–D20.

Corich V, Giacomini A, Vian P, Vendramin E, Carlot M, Basaglia M, Squartini A, Casella S, Nuti MP. **2001**. Aspects of marker/reporter stability and selectivity in soil microbiology. *Microbial Ecology* **41**: 333–340.

Czernic P, Gully D, Cartieaux F, Moulin L, Guefrachi I, Patrel D, Pierre O, Fardoux J, Chaintreuil C, Nguyen P, *et al.* 2015. Convergent Evolution of Endosymbiont Differentiation in Dalbergioid and Inverted Repeat-Lacking Clade Legumes Mediated by Nodule-Specific Cysteine-Rich Peptides. *Plant physiology* **169**: 1254–65.

Denison RF. **2000**. Legume Sanctions and the Evolution of Symbiotic Cooperation by Rhizobia. *The American naturalist* **156**: 567–576.

Denton MD, Reeve WG, Howieson JG, Coventry DR. **2003**. Competitive abilities of common field isolates and a commercial strain of *Rhizobium leguminosarum* bv. *trifolii* for clover nodule occupancy. *Soil Biology and Biochemistry* **35**: 1039–1048.

Dilworth MJ, James EK, Sprent JI, Newton WE. **2008**. *Nitrogen-fixing Leguminous Symbioses*

(MJ Dilworth, EK James, JI Sprent, and WE Newton, Eds.). Dordrecht: Springer Netherlands.

Dixon R, Kahn D. **2004**. Genetic regulation of biological nitrogen fixation. *Nature reviews. Microbiology* **2**: 621–31.

Dowling DN, Broughton WJ. **1986**. Competition for Nodulation of Legumes. *Annual Review of Microbiology* **40**: 131–157.

Downie JA. **2014**. Legume nodulation. *Current Biology* **24**: R184–R190.

Downie JA, Ma QS, Knight CD, Hombrecher G, Johnston AW. **1983**. Cloning of the symbiotic region of *Rhizobium leguminosarum*: the nodulation genes are between the nitrogenase genes and a nifA-like gene. *The EMBO journal* **2**: 947–52.

Dresler-Nurmi A, Fewer D, Räsänen L, Lindström K. **2009**. The Diversity and Evolution of Rhizobia. In: Pawlowski K, ed. Microbiology Monographs. Prokaryotic Symbionts in Plants SE - 99. Springer Berlin Heidelberg, 3–41.

Drew E, Herridge D, Ballard R, Hara GO, Deaker R, Denton M, Yates R, Gemell G, Hartley E, Phillips L, *et al.* 2012. *Inoculating Legumes : a Practical Guide*. Grains Research and Development Corporation.

Duodu S, Brophy C, Connolly J, Svenning MM. **2009**. Competitiveness of a native *Rhizobium leguminosarum* biovar *trifolii* strain for nodule occupancy is manifested during infection. *Plant and Soil* **318**: 117–126.

Edgar RC. **2010**. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461.

Engler C, Kandzia R, Marillonnet S. **2008**. A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE* **3**: e3647.

Engler C, Marillonnet S. **2011**. Generation of families of construct variants using golden gate shuffling. *Methods in molecular biology* **729**: 167–181.

Engler C, Youles M, Gruetzner R, Ehnert T-MM, Werner S, Jones JDG, Patron NJ, Marillonnet S. **2014**. A Golden Gate modular cloning toolbox for plants. *ACS Synthetic Biology* **3**: 839–843.

Fadrosh DW, Ma B, Gajer P, Sengamalay N, Ott S, Brotman RM, Ravel J. **2014**. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome* **2**: 6.

Fan J, Crooks C, Lamb C. **2008**. High-throughput quantitative luminescence assay of the growth in planta of *Pseudomonas syringae* chromosomally tagged with *Photorhabdus luminescens luxCDABE*. *The Plant journal : for cell and molecular biology* **53**: 393–9.

Fan R, Huang YC, Grusak MA, Huang CP, Sherrier DJ. **2014**. Effects of nano-TiO2 on the agronomically-relevant Rhizobium–legume symbiosis. *Science of The Total Environment* **466–467**: 503–512.

Farkas A, Maroti G, Durg H, Gyorgypal Z, Lima RM, Medzihradszky KF, Kereszt A, Mergaert P, Kondorosi E. **2014**. *Medicago truncatula* symbiotic peptide NCR247 contributes to bacteroid differentiation through multiple mechanisms. *Proceedings of the National Academy of Sciences* **111**: 5183–5188.

Ferguson BJ, Indrasumunar A, Hayashi S, Lin M-H, Lin Y-H, Reid DE, Gresshoff PM. **2010**. Molecular Analysis of Legume Nodule Development and Autoregulation. *Journal of Integrative Plant Biology* **52**: 61–76.

Figurski DH, Helinski DR. **1979**. Replication of an origin-containing derivative of plasmid RK2

dependent on a plasmid function provided in trans. *Proceedings of the National Academy of Sciences* **76**: 1648–1652.

Food and Agriculture Organization of the United Nations. 2015. *World fertilizer trends and outlook to 2018.* Food & Agriculture Organization of United Nations.

Franche C, Lindström K, Elmerich C. **2008**. Nitrogen-fixing bacteria associated with leguminous and non-leguminous plants. *Plant and Soil* **321**: 35–59.

Frederix M, Edwards A, Swiderska A, Stanger A, Karunakaran R, Williams A, Abbruscato P, Sanchez-Contreras M, Poole PS, Downie JA. **2014**. Mutation of praR in *Rhizobium leguminosarum* enhances root biofilms, improving nodulation competitiveness by increased expression of attachment proteins. *Molecular microbiology* **93**: 464–478.

Frey SD, Blum LK. **1994**. Effect of pH on competition for nodule occupancy by type I and type II strains of *Rhizobium leguminosarum* bv. *phaseoli*. *Plant and Soil* **163**: 157–164.

Gage DJ. **2004**. Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes. *Microbiology and molecular biology reviews : MMBR* **68**: 280–300.

Gallique M, Bouteiller M, Merieau A. **2017**. The Type VI Secretion System: A Dynamic System for Bacterial Communication? *Frontiers in Microbiology* **8**: 1454.

Galloway JN, Dentener FJ, Capone DG, Boyer EW, Howarth RW, Seitzinger SP, Asner GP, Cleveland CC, Green PA, Holland EA, *et al.* 2004. Nitrogen Cycles: Past, Present, and Future. *Biogeochemistry* **70**: 153–226.

García Mc, Marina Ml, Laborda F, Torre M. **1998**. Chemical characterization of commercial soybean products. *Food Chemistry* **62**: 325–331.

Geddes B, Mendoza-Suárez M, Poole P. **2018**. A Bacterial Expression Vector Archive (BEVA) for flexible modular assembly of Golden Gate-compatible vectors. *bioRxiv*: 392423.

Geddes BA, Oresnik IJ. **2012**. Inability to catabolize galactose leads to increased ability to compete for nodule occupancy in *Sinorhizobium meliloti*. *Journal of bacteriology* **194**: 5044–53.

Giller KE. **2001**. *Nitrogen fixation in tropical cropping systems* (KE Giller, Ed.). Wallingford: CABI.

Graham PH, Vance CP. **2003**. Legumes: importance and constraints to greater use. *Plant physiology* **131**: 872–877.

Graham PH, Viteri SE, Mackie F, Vargas AT, Palacios A. **1982**. Variation in acid soil tolerance among strains of *Rhizobium phaseoli*. *Field Crops Research* **5**: 121–128.

Hakkila K, Maksimow M, Karp M, Virta M. **2002**. Reporter Genes lucFF, luxCDABE, gfp, and dsred Have Different Characteristics in Whole-Cell Bacterial Sensors. *Analytical Biochemistry* **301**: 235–242.

Hardy RW, Holsten RD, Jackson EK, Burns RC. **1968**. The acetylene-ethylene assay for n(2) fixation: laboratory and field evaluation. *Plant physiology* **43**: 1185–207.

Hartmann A, Rothballer M, Schmid M. **2008**. Lorenz Hiltner, a pioneer in rhizosphere microbial ecology and soil bacteriology research. *Plant and Soil* **312**: 7–14.

Howieson JG, Dilworth MJ. 2016. Working with rhizobia. ACIAR Monograph No. 173. ACIAR, 314.

Hu K, Li Y, Chen W, Chen D, Wei Y, Edis R, Li B, Huang Y, Zhang Y. **2010**. Modeling nitrate leaching and optimizing water and nitrogen management under irrigated maize in desert

oases in Northwestern China. *Journal of environmental quality* **39**: 667–677.

Islam MR, Mao S, Xue X, Eneji a E, Zhao X, Hu Y. **2011**. A lysimeter study of nitrate leaching, optimum fertilisation rate and growth responses of corn (*Zea mays* L.) following soil amendment with water-saving super-absorbent polymer. *Journal of the science of food and agriculture* **91**: 1990–1997.

Ji ZJ, Yan H, Cui QG, Wang ET, Chen WF, Chen WX. **2017**. Competition between rhizobia under different environmental conditions affects the nodulation of a legume. *Systematic and Applied Microbiology* **40**: 114–119.

Johnston AWB, Beringer JE. **1975**. Identification of the rhizobium strains in pea root nodules using genetic markers. *Journal of general microbiology* **87**: 343–50.

Jorrin B, Imperial J. **2015**. Population Genomics Analysis of Legume Host Preference for Specific Rhizobial Genotypes in the *Rhizobium leguminosarum* bv. *viciae* Symbioses. *Molecular Plant-Microbe Interactions* **28**: 310–318.

Jorrin Rubio B. **2016**. Genomics of Specificity in the Symbiotic Interaction between *Rhizobium leguminosarum* and Legumes.

Karunakaran R, Mauchline TH, Hosie AHF, Poole PS. **2005**. A family of promoter probe vectors incorporating autofluorescent and chromogenic reporter proteins for studying gene expression in Gram-negative bacteria. *Microbiology* **151**: 3249–3256.

Kiers ET, Rousseau RA, West SA, Denlson RF. **2003**. Host sanctions and the legume-rhizobium mutualism. *Nature* **425**: 78–81.

Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. **2013**. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on

the MiSeq Illumina sequencing platform. **79**: 5112–5120.

Kumar N, Lad G, Giuntini E, Kaye ME, Udomwong P, Shamsani NJ, Young JPW, Bailly X. **2015**. Bacterial genospecies that are not ecologically coherent: population genomics of *Rhizobium leguminosarum*. *Open biology* **5**: 140133.

Laguerre G, Mavingui P, Allard MR, Charnay MP, Louvrier P, Mazurier SI, Rigottier-Gois L, Amarger N. **1996**. Typing of rhizobia by PCR DNA fingerprinting and PCR-restriction fragment length polymorphism analysis of chromosomal and symbiotic gene regions: application to *Rhizobium leguminosarum* and its different biovars. *Applied and environmental microbiology* **62**: 2029–36.

Lakzian A, Murphy P, Turner A, Beynon JL, Giller KE. **2002**. Rhizobium leguminosarum bv. viciae populations in soils with increasing heavy metal contamination: abundance, plasmid profiles, diversity and metal tolerance. *Soil Biology and Biochemistry* **34**: 519–529.

Lang C, Smith LS, Long SR. **2018**. Characterization of Novel Plant Symbiosis Mutants Using a New Multiple Gene-Expression Reporter *Sinorhizobium meliloti* Strain. *Frontiers in Plant Science* **9**: 76.

Lardi M, de Campos SB, Purtschert G, Eberl L, Pessi G. **2017**. Competition experiments for legume infection identify *Burkholderia phymatum* as a highly competitive β-rhizobium. *Frontiers in Microbiology* **8**: 1527.

Lee CC, Ribbe MW, Hu Y. **2014**. Cleaving the N,N triple bond: The transformation of dinitrogen to ammonia by nitrogenases. *Metal Ions in Life Sciences* **14**–**113**: 147–176.

Lindemann WC, Schmidt EL, Ham GE. **1974**. Evidence for double infection within soybean nodules. *Soil Science* **118**.

Lodwig EM, Hosie a HF, Bourdès A, Findlay K, Allaway D, Karunakaran R, Downie J a, Poole PS. **2003**. Amino-acid cycling drives nitrogen fixation in the legume-Rhizobium symbiosis. *Nature* **422**: 722–726.

López-García SL, Vázquez TEE, Favelukes G, Lodeiro AR. **2002**. Rhizobial position as a main determinant in the problem of competition for nodulation in soybean. *Environmental microbiology* **4**: 216–24.

Lupwayi NZ, Clayton GW, Rice WA. **2006**. Rhizobial Inoculants for Legume Crops. *Journal of Crop Improvement* **15**: 97–125.

Lutz R, Bujard H. **1997**. Independent and tight regulation of transcriptional units in escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2regulatory elements. *Nucleic Acids Research* **25**: 1203–1210.

Maathuis FJM. **2009**. Physiological functions of mineral macronutrients. *Current opinion in plant biology* **12**: 250–8.

Markowitz VM, Chen IMA, Chu K, Szeto E, Palaniappan K, Grechkin Y, Ratner A, Jacob B, Pati A, Huntemann M, *et al.* 2012. IMG/M: The integrated metagenome data management and comparative analysis system. *Nucleic Acids Research* **40**.

Martínez-Garćía E, Aparicio T, Goñi-Moreno A, Fraile S, De Lorenzo V. **2015**. SEVA 2.0: An update of the Standard European Vector Architecture for de-/re-construction of bacterial functionalities. *Nucleic Acids Research* **43**: D1183–D1189.

Martinez E, Romero D, Palacios R. **2008**. The Rhizobium Genome. *Critical Reviews in Plant Sciences* **9**: 59–93.

Masson-Boivin C, Giraud E, Perret X, Batut J. **2009**. Establishing nitrogen-fixing symbiosis

with legumes: how many rhizobium recipes? *Trends in microbiology* **17**: 458–66.

Maunoury N, Kondorosi A, Kondorosi E, Mergaert P. **2008**. Cell Biology Of Nodule Infection And Development. Nitrogen Fixation: Origins, Applications, and Research Progress. Nitrogen-fixing Leguminous Symbioses SE - 6. Springer Netherlands, 153–189.

May SN, Bohlool BB. **1983**. Competition Among *Rhizobium leguminosarum* Strains for Nodulation of Lentils (*Lens esculenta*). *Applied and environmental microbiology* **45**: 960–5.

McIntosh GH, Topping DL. **2000**. Food legumes in Human Nutrition. In: Knight R, ed. Current Plant Science and Biotechnology in Agriculture. Linking Research and Marketing Opportunities for Pulses in the 21st Century SE - 63. Springer Netherlands, 655–660.

Meade HM, Long SR, Ruvkun GB, Brown SE, Ausubel FM. **1982**. Physical and genetic characterization of symbiotic and auxotrophic mutants of *Rhizobium meliloti* induced by transposon Tn5 mutagenesis. *Journal of bacteriology* **149**: 114–22.

Meena RS, Lal R. **2018**. Legumes and Sustainable Use of Soils. Legumes for Soil Health and Sustainable Management. Singapore: Springer Singapore, 1–31.

Melino VJ, Drew EA, Ballard RA, Reeve WG, Thomson G, White RG, O'Hara GW. **2012**. Identifying abnormalities in symbiotic development between *Trifolium* spp. and *Rhizobium leguminosarum* bv. *trifolii* leading to sub-optimal and ineffective nodule phenotypes. *Annals of botany* **110**: 1559–72.

Mellor RB, Collinge DB. **1995**. A simple model based on known plant defence reactions is sufficient to explain most aspects of nodulation. *Journal of Experimental Botany* **46**: 1–18.

Minchin FR, Witty JF, Mytton LR. **1994**. Reply to 'Measurement of nitrogenase activity in legume root nodules: In defense of the acetylene reduction assay' by J.K. Vessey. *Plant and*

Soil **158**: 163–167.

Mishra RPN, Tisseyre P, Melkonian R, Chaintreuil C, Miché L, Klonowska A, Gonzalez S, Bena G, Laguerre G, Moulin L. **2012**. Genetic diversity of *Mimosa pudica* rhizobial symbionts in soils of French Guiana: investigating the origin and diversity of *Burkholderia phymatum* and other beta-rhizobia. *FEMS Microbiology Ecology* **79**: 487–503.

Moawad M, Schmidt EL. **1987**. Occurrence and nature of mixed infections in nodules of field-grown soybeans (*Glycine max*). *Biology and Fertility of Soils* **5**: 112–114.

Moore SJ, Lai H-E, Kelwick RJR, Chee SM, Bell DJ, Polizzi KM, Freemont PS. **2016**. EcoFlex: A Multifunctional MoClo Kit for *E. coli* Synthetic Biology. *ACS Synthetic Biology* **5**: 1059–1069.

Morris JB. **1997**. Special-Purpose Legume Genetic Resources Conserved for Agricultural, Industrial, and Pharmaceutical Use. *Economic Botany* **51**: 251–263.

Mortier V, Holsters M, Goormachtig S. **2012**. Never too many? How legumes control nodule numbers. *Plant, Cell and Environment* **35**: 245–258.

Mulley G, Lopez-Gomez M, Zhang Y, Terpolilli J, Prell J, Finan T, Poole P. **2010**. Pyruvate is synthesized by two pathways in pea bacteroids with different efficiencies for nitrogen fixation. *Journal of bacteriology* **192**: 4944–4953.

Murray JD. **2011**. Invasion by invitation: rhizobial infection in legumes. *Molecular plant-microbe interactions : MPMI* **24**: 631–639.

Murray DC, Coghlan ML, Bunce M. **2015**. From benchtop to desktop: important considerations when designing amplicon sequencing workflows. *PloS one* **10**: e0124671.

Naeem FI, Ashraf MM, Malik K a., Hafeez FY. **2004**. Competitiveness of introduced Rhizobium strains for nodulation in fodder legumes. *Pakistan Journal of Botany* **36**: 159–

166.

Nicol H, Thornton HG. **1941**. Competition between Related Strains of Nodule Bacteria and Its Influence on Infection of the Legume Host. *Proceedings of the Royal Society B: Biological Sciences* **130**: 32–59.

NRA, CIRAD A and F. **2015**. Feedipedia - Animal Feed Resources Information System. *NRA, CIRAD, AFZ and FAO*.

O'Hara GW. **1998**. The Role of Nitrogen Fixation in Crop Production. *Journal of Crop Production* **1**: 115–138.

Oke V, Long SR. **1999**. Bacteroid formation in the Rhizobium–legume symbiosis. *Current Opinion in Microbiology* **2**: 641–646.

Oldroyd GED. **2013**. Speak, friend, and enter: Signalling systems that promote beneficial symbiotic associations in plants. *Nature Reviews Microbiology* **11**: 252–263.

Oldroyd GED, Murray JD, Poole PS, Downie JA. **2011**. The rules of engagement in the legume-rhizobial symbiosis. *Annual review of genetics* **45**: 119–44.

Oliveira MF da S, Mendes DQ, Ferrari LI, Vasconcelos ATR. **2004**. Ribosome binding site recognition using neural networks. *Genetics and Molecular Biology* **27**: 644–650.

Onishchuk OP, Vorobyov NI, Provorov NA. **2017**. Nodulation competitiveness of nodule bacteria: Genetic control and adaptive significance: Review. *Applied Biochemistry and Microbiology* **53**: 131–139.

Oono R, Anderson CG, Denison RF. **2011**. Failure to fix nitrogen by non-reproductive symbiotic rhizobia triggers host sanctions that reduce fitness of their reproductive clonemates. *Proceedings of the Royal Society B: Biological Sciences* **278**: 2698–2703.

Oresnik IJ, Twelker S, Hynes MF. **1999**. Cloning and Characterization of a *Rhizobium leguminosarum* Gene Encoding a Bacteriocin with Similarities to RTX Toxins. *Applied and environmental microbiology* **65**: 2833–2840.

Pace NR, Stahl DA, Lane DJ, Olsen GJ. **1986**. The Analysis of Natural Microbial Populations by Ribosomal RNA Sequences. Springer, Boston, MA, 1–55.

Parnell JJ, Berka R, Young HA, Sturino JM, Kang Y, Barnhart DM, DiLeo M V. 2016. From the Lab to the Farm: An Industrial Perspective of Plant Beneficial Microorganisms. *Frontiers in Plant Science* **7**: 1–12.

Pédelacq J-D, Cabantous S, Tran T, Terwilliger TC, Waldo GS. **2006**. Engineering and characterization of a superfolder green fluorescent protein. *Nature biotechnology* **24**: 79.

Peoples MB, Herridge DF, Ladha JK. **1995**. Biological nitrogen fixation: An efficient source of nitrogen for sustainable agricultural production? *Plant and Soil* **174**: 3–28.

Peralta H, Mora Y, Salazar E, Encarnación S, Palacios R, Mora J. **2004**. Engineering the nifH promoter region and abolishing poly-β -hydroxybutyrate accumulation in *Rhizobium etli* enhance nitrogen fixation in symbiosis with *Phaseolus vulgaris*. *Applied and Environmental Microbiology* **70**: 3272–3281.

Perry BJ, Yost CK. **2014**. Construction of a mariner-based transposon vector for use in insertion sequence mutagenesis in selected members of the Rhizobiaceae. *BMC Microbiology* **14**: 298.

Pini F, East AK, Appia-Ayme C, Tomek J, Karunakaran R, Mendoza-Suárez M, Edwards A, Terpolilli JJ, Roworth J, Downie JA, *et al.* 2017. Bacterial Biosensors for in Vivo Spatiotemporal Mapping of Root Secretion. *Plant Physiology* **174**: 1289–1306.

Poole P, Ramachandran V, Terpolilli J. **2018**. Rhizobia: from saprophytes to endosymbionts. *Nature Reviews Microbiology*.

Poole PS, Schofield N, Reid CJ, Drew EM, Walshaw DL. **1994**. Identification of chromosomal genes located downstream of dctD that affect the requirement for calcium and the lipopolysaccharide layer of *Rhizobium leguminosarum*. *Microbiology* **140**: 2797–809.

Poysti NJ, Loewen EDM, Wang Z, Oresnik IJ. **2007**. *Sinorhizobium meliloti* pSymB carries genes necessary for arabinose transport and catabolism. *Microbiology* **153**: 727–36.

Prell J, Boesten B, Poole PS, Priefer UB. **2002**. The *Rhizobium leguminosarum* bv. *viciae* VF39 γ-aminobutyrate (GABA) aminotransferase gene (gabT) is induced by GABA and highly expressed in bacteroids. *Microbiology* **148**: 615–623.

Pueppke SG, Broughton WJ. **1999**. *Rhizobium* sp. strain NGR234 and *R. fredii* USDA257 share exceptionally broad, nested host ranges. *Molecular plant-microbe interactions : MPMI* **12**: 293–318.

Puppo A, Groten K, Bastian F, Carzaniga R, Soussi M, Lucas MM, De Felipe MR, Harrison J, Vanacker H, Foyer CH. **2004**. Legume nodule senescence: roles for redox and hormone signalling in the orchestration of the natural aging process. *New Phytologist* **165**: 683–701.

Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y. **2012**. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* **13**: 341.

Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. **2017**. Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology* **35**: 833–844.

Quinto C, de la Vega H, Flores M, Fernández L, Ballado T, Soberón G, Palacios R. **1982**.

Reiteration of nitrogen fixation gene sequences in *Rhizobium phaseoli*. *Nature* **299**: 724–726.

Quinto C, De La Vega H, Flores M, Leemans J, Cevallos M a, Pardo M a, Azpiroz R, De Lourdes Girard M, Calva E, Palacios R. **1985**. Nitrogenase reductase: A functional multigene family in *Rhizobium phaseoli*. *Proceedings of the National Academy of Sciences of the United States of America* **82**: 1170–1174.

Ramachandran VK, East AK, Karunakaran R, Downie JA, Poole PS. **2011**. Adaptation of *Rhizobium leguminosarum* to pea, alfalfa and sugar beet rhizospheres investigated by comparative transcriptomics. *Genome biology* **12**: R106.

Reeve W, Chain P, O'Hara G, Ardley J, Nandesena K, Bräu L, Tiwari R, Malfatti S, Kiss H, Lapidus A, *et al.* 2010a. Complete genome sequence of the Medicago microsymbiont *Ensifer* (*Sinorhizobium*) *medicae* strain WSM419. *Standards in genomic sciences* **2**: 77–86.

Reeve W, O'Hara G, Chain P, Ardley J, Bräu L, Nandesena K, Tiwari R, Copeland A, Nolan M, Han C, *et al.* 2010b. Complete genome sequence of *Rhizobium leguminosarum* bv. *trifolii* strain WSM1325, an effective microsymbiont of annual Mediterranean clovers. *Standards in genomic sciences* **2**: 347–356.

Regus JU, Quides KW, O'Neill MR, Suzuki R, Savory EA, Chang JH, Sachs JL. **2017**. Cell autonomous sanctions in legumes target ineffective rhizobia in nodules with mixed infections. *American Journal of Botany* **104**: 1299–1312.

Richardson AE, Viccars LA, Watson JM, Gibson AH. **1995**. Differentiation of Rhizobium strains using the polymerase chain reaction with random and directed primers. *Soil Biology and Biochemistry* **27**: 515–524.

Roelvink P, Harmsen M. **1990**. The nifH promoter region of *Rhizobium leguminosarum*: nucleotide sequence and promoter elements controlling activation by NifA protein. *Gene* **87**: 31–36.

Ryan P, Delhaize E, Jones D. **2001**. Function and mechanism of organic anion exudation from plant roots. *Annual review of plant physiology and plant molecular biology* **52**: 527–560.

Ryu H, Cho H, Choi D, Hwang I. **2012**. Plant hormonal regulation of nitrogen-fixing nodule organogenesis. *Molecules and Cells* **34**: 117–126.

Sambrook J, Fritsch EF, Maniatis T. **1989**. *Molecular cloning: a laboratory manual*. Cold Spring Harbor.

Sánchez-Cañizares C, Jorrín B, Durán D, Nadendla S, Albareda M, Rubio-Sanz L, Lanza M, González-Guerrero M, Prieto R, Brito B, *et al.* 2018. Genomic Diversity in the Endosymbiotic Bacterium *Rhizobium leguminosarum*. *Genes* **9**: 60.

Sánchez-Cañizares C, Palacios J. **2013**. Construction of a marker system for the evaluation of competitiveness for legume nodulation in Rhizobium strains. *Journal of Microbiological Methods* **92**: 246–249.

Santi C, Bogusz D, Franche C. **2013**. Biological nitrogen fixation in non-legume plants. *Annals of botany* **111**: 743–67.

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, *et al.* 2012. Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**: 676–682.

Schmidhauser TJ, Helinski DR. **1985**. Regions of broad-host-range plasmid RK2 involved in

replication and stable maintenance in nine species of gram-negative bacteria. *Journal of bacteriology* **164**: 446–455.

Schwinghamer EA, Brockwell J. **1978**. Competitive advantage of bacteriocin and phage-producing strains of Rhizobium trifolii in mixed culture. *Soil Biology and Biochemistry* **10**: 383–387.

Sessitsch A, Hardarson G, Vos WM de, Wilson KJ. **1998**. Use of marker genes in competition studies of Rhizobium. *Plant and Soil* **204**: 35–45.

Sessitsch A, Jjemba PK, Hardarson G, Akkermans ADL, Wilson KJ. **1997**. Measurement of the competitiveness index of *Rhizobium tropici* strain CIAT899 derivatives marked with the gusA gene. *Soil Biology and Biochemistry* **29**: 1099–1110.

Sessitsch A, Wilson K, Akkermans A, de Vos W. **1996**. Simultaneous detection of different Rhizobium strains marked with either the *Escherichia coli* gusA gene or the *Pyrococcus furiosus* celB gene. *Applied and environmental microbiology* **62**: 4191–4194.

Shaner NC, Campbell RE, Steinbach PA, Giepmans BNG, Palmer AE, Tsien RY. **2004**. Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nature Biotechnology* **22**: 1567–1572.

Silva-Rocha R, Martínez-García E, Calles B, Chavarría M, Arce-Rodríguez A, De Las Heras A, Páez-Espino AD, Durante-Rodríguez G, Kim J, Nikel PI, *et al.* 2013. The Standard European Vector Architecture (SEVA): A coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Research* **41**: D666–D675.

Singer JT, Phennicie RT, Sullivan MJ, Porter LA, Shaffer VJ, Kim CH. **2010**. Broad-Host-Range Plasmids for Red Fluorescent Protein Labeling of Gram-Negative Bacteria for Use in the

Zebrafish Model System †. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY* **76**: 3467–3474.

Slattery J, Pearce D. **2002**. Development of Elite Inoculant Rhizobium Strains in Southeastern Australia. *ACIAR Proceedings 109e*: 86–94.

Somasegaran P, Hoben HJ. **1994**. *Handbook for Rhizobia*. New York, NY: Springer New York.

de Souza RSC, Okura VK, Armanhi JSL, Jorrín B, Lozano N, da Silva MJ, González-Guerrero M, de Araújo LM, Verza NC, Bagheri HC, *et al.* 2016. Unlocking the bacterial and fungal communities assemblages of sugarcane microbiome. *Scientific Reports* **6**: 28774.

Speranskaya AS, Khafizov K, Ayginin AA, Krinitsina AA, Omelchenko DO, Nilova M V., Severova EE, Samokhina EN, Shipulin GA, Logacheva MD. **2018**. Comparative analysis of Illumina and Ion Torrent high-throughput sequencing platforms for identification of plant components in herbal tea. *Food Control*.

Sprent JI. **2008**. Evolution and Diversity of Legume Symbiosis. In: Dilworth M,, In: James E,, In: Sprent J,, In: Newton W, eds. Nitrogen Fixation: Origins, Applications, and Research Progress. Nitrogen-fixing Leguminous Symbioses SE - 1. Springer Netherlands, 1–21.

Sprent JI, James EK. **2007**. Legume evolution: where do nodules and mycorrhizas fit in? *Plant physiology* **144**: 575–581.

Stein M, Bromfield ESP, Dye M. **1982**. An assessment of a method based on intrinsic antibiotic resistance for identifying Rhizobium strains. *Annals of Applied Biology* **101**: 261–267.

Sullivan JT, Brown SD, Ronson CW. **2013**. The NifA-RpoN Regulon of *Mesorhizobium loti* Strain R7A and Its Symbiotic Activation by a Novel LacI/GalR-Family Regulator (V de Crécy-Lagard, Ed.). *PLoS ONE* **8**: e53762.

Terfrüchte M, Joehnk B, Fajardo-Somera R, Braus GH, Riquelme M, Schipper K, Feldbrügge M. **2014**. Establishing a versatile Golden Gate cloning system for genetic engineering in fungi. *Fungal Genetics and Biology* **62**: 1–10.

Terpolilli J, Hill Y, Tian R, Howieson J, Bräu L, Goodwin L, Han J, Liolios K, Huntemann M, Pati A, *et al.* 2013. Genome sequence of Ensifer meliloti strain WSM1022; a highly effective microsymbiont of the model legume *Medicago truncatula* A17. *Standards in genomic sciences* **9**: 315–24.

Terpolilli JJ, Hood GA, Poole PS. **2012**. What determines the efficiency of $\mathbf{N_2}$-fixing Rhizobium-legume symbioses? *Advances in microbial physiology* **60**: 325–389.

Tett AJ, Rudder SJ, Bourdès A, Karunakaran R, Poole PS. **2012**. Regulatable vectors for environmental gene expression in Alphaproteobacteria. *Applied and environmental microbiology* **78**: 7137–7140.

Thies JE, Holmes EM, Vachot A. **2001**. Application of molecular techniques to studies in Rhizobium ecology: a review. *Australian Journal of Experimental Agriculture* **41**: 299–319.

Thoma S, Schobert M. **2009**. An improved *Escherichia coli* donor strain for diparental mating. *FEMS Microbiology Letters* **294**: 127–132.

Thöny B, Hennecke H. **1989**. The -24/-12 promoter comes of age. *FEMS Microbiology Letters* **63**: 341–357.

Tilman D, Cassman KG, Matson PA, Naylor R, Polasky S. **2002**. Agricultural sustainability and intensive production practices. *Nature* **418**: 671–677.

Timmers ACJ, Soupène E, Auriac M-C, de Billy F, Vasse J, Boistard P, Truchet G. **2000**. Saprophytic Intracellular Rhizobia in Alfalfa Nodules. *Molecular Plant-Microbe Interactions*

13: 1204–1213.

Tscharntke T, Clough Y, Wanger TC, Jackson L, Motzke I, Perfecto I, Vandermeer J, Whitbread A. **2012**. Global food security, biodiversity conservation and the future of agricultural intensification. *Biological Conservation* **151**: 53–59.

Unkovich M, Herridge D, Peoples M, Cadisch G, Boddey B, Giller K, Alves B, Chalk P. **2008**. *Measuring plant-associated nitrogen fixation in agricultural systems.* Canberra: Australian Centre for International Agricultural Research (ACIAR).

Turner TR, James EK, Poole PS. **2013**. The plant microbiome. *Genome biology* **14**: 209.

Vargas AAT, Graham PH. **1989**. Cultivar and pH effects on competition for nodule sites between isolates of Rhizobium in beans. *Plant and Soil* **117**: 195–200.

Vessey JK. **1994**. Measurement of nitrogenase activity in legume root nodules: In defense of the acetylene reduction assay. *Plant and Soil* **158**: 151–162.

Vasse J, Billy F, Truchet G. **1993**. Abortion of infection during the *Rhizobium meliloti*-alfalfa symbiotic interaction is accompanied by a hypersensitive reaction. *The Plant Journal* **4**: 555–566.

van Velzen R, Holmer R, Bu F, Rutten L, van Zeijl A, Liu W, Santuari L, Cao Q, Sharma T, Shen D, *et al.* 2018. Comparative genomics of the nonlegume Parasponia reveals insights into evolution of nitrogen-fixing rhizobium symbioses. *Proceedings of the National Academy of Sciences* **115**: E4700–E4709.

Větrovský T, Baldrian P. **2013**. The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses (J Neufeld, Ed.). *PLoS ONE* **8**: e57923.

Vijayakumar V. **2018**. Symbiotic Tripartism in the Model Plant Family of Legumes and Soil Sustainability. Legumes for Soil Health and Sustainable Management. Singapore: Springer Singapore, 173–203.

Walker TS, Bais HP, Grotewold E, Vivanco JM. **2003**. Root exudation and rhizosphere biology. *Plant physiology* **132**: 44–51.

Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S. **2011**. A modular cloning system for standardized assembly of multigene constructs (J Peccoud, Ed.). *PLoS ONE* **6**: e16765.

Weidner S, Becker A, Bonilla I, Jaenicke S, Lloret J, Margaret I, Pühler A, Ruiz-Sainz JE, Schneiker-Bekel S, Szczepanowski R, *et al.* 2012. Genome sequence of the soybean symbiont *Sinorhizobium fredii* HH103. *Journal of bacteriology* **194**: 1617–8.

Weinstein M, Roberts RC, Helinski DR. **1992**. A region of the broad-host-range plasmid RK2 causes stable in planta inheritance of plasmids in *Rhizobium meliloti* cells isolated from alfalfa root nodules. *Journal of bacteriology* **174**: 7486–7489.

Werner D, Newton WE. **2005**. *Nitrogen Fixation in Agriculture, Forestry, Ecology, and the Environment* (D Werner and WE Newton, Eds.). Springer Netherlands.

Wheatley RM, Ramachandran VK, Geddes BA, Perry BJ, Yost CK, Poole PS. **2017**. Role of O_2 in the Growth of *Rhizobium leguminosarum* bv. *viciae* 3841 on Glucose and Succinate. *Journal of bacteriology* **199**: e00572-16.

Wilson KJ. **1995**. Molecular techniques for the study of rhizobial ecology in the field. *Soil Biology and Biochemistry* **27**: 501–514.

Wilson KJ, Peoples MB, Jefferson RA. **1995a**. New techniques for studying competition by Rhizobia and for assessing nitrogen fixation in the field. *Plant and Soil* **174**: 241–253.

Wilson KJ, Sessitsch A, Corbo JC, Giller KE, Akkermans ADL, Jefferson RA. **1995b**. β-Glucuronidase (GUS) transposons for ecological and genetic studies of rhizobia and other Gram-negative bacteria. *Microbiology* **141**: 1691–1705.

Yates RJ, Howieson JG, Reeve WG, O'Hara GW. **2011**. A re-appraisal of the biology and terminology describing rhizobial strain success in nodule occupancy of legumes in agriculture. *Plant and Soil* **348**: 255–267.

Young CC, Chao CC. **1989**. Intrinsic antibiotic resistance and competition in fast- and slow-growing soybean rhizobia on a hybrid of Asian and US cultivars. *Biology and Fertility of Soils* **8**: 66–70.

Young J. **1996**. Phylogeny and taxonomy of rhizobia. *Plant and Soil*: 45–52.

Young JPW, Crossman LC, Johnston AWB, Thomson NR, Ghazoui ZF, Hull KH, Wexler M, Curson ARJ, Todd JD, Poole PS, *et al.* 2006. The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome biology* **7**: R34.

Zhang YM, Tian CF, Sui XH, Chen WF, Chen WX. **2012**. Robust markers reflecting phylogeny and taxonomy of rhizobia. *PloS ONE* **7**: e44936.

Zhang JJ, Yu T, Lou K, Mao PH, Wang ET, Chen WF, Chen WX. **2014**. Genotypic alteration and competitive nodulation of *Mesorhizobium muleiense* against exotic chickpea rhizobia in alkaline soils. *Systematic and Applied Microbiology* **37**: 520–524.

Supplemental material

Table S. 1 Evaluating effectiveness by Acetylene reduction assay and dry weight biomass.

Strain	Acetylene reduced (plant)$^{-1}$	SEM	Effectiveness class in pea by DW[a]
WSM4460	2.94	0.27	Medium
WSM4462	2.85	0.11	Low
WSM1529	2.59	0.15	High
WSM4458	2.27	0.11	Low
60 n1	2.26	0.04	Medium
WSM4461	2.22	0.07	High
WSM1521	2.19	0.13	High
WSM4459	2.14	0.14	Low
WSM1480	2.10	0.12	Medium
SU303	2.00	0.39	High
WSM1481	1.97	0.16	Medium
71 n4	1.89	0.24	Medium
54 n4	1.71	0.20	High
WSM1488	1.43	0.08	Low
SARDI 962	1.37	0.20	Low
WSM4645	1.12	0.18	High
WSM1475	0.20	0.05	Low

Acetylene reduced in μmoles of C_2H_2 reduced per hour per plant. Values ± SEM are shown from three biological and three technical replicates.

[a] Information obtained from Professor John Howieson from the Centre for Rhizobium Studies (CRS), Perth, Western Australia.

Table S. 2 IDs as sense and antisense oligonucleotides to be assemble as T modules.

Oligo name	Full Sequence (5' to 3')
806rcbc0	GCTTTCCCTTGTCTCCCGTTTTACAACGTCGTGACTGGG
806rcbc0_anti	AGCGCCCAGTCACGACGTTGTAAAACGGGAGACAAGGGA
806rcbc1	GCTTACGAGACTGATTCGTTTTACAACGTCGTGACTGGG
806rcbc1_anti	AGCGCCCAGTCACGACGTTGTAAAACGAATCAGTCTCGT
806rcbc2	GCTTGCTGTACGGATTCGTTTTACAACGTCGTGACTGGG
806rcbc2_anti	AGCGCCCAGTCACGACGTTGTAAAACGAATCCGTACAGC
806rcbc3	GCTTATCACCAGGTGTCGTTTTACAACGTCGTGACTGGG
806rcbc3_anti	AGCGCCCAGTCACGACGTTGTAAAACGACACCTGGTGAT
806rcbc4	GCTTTGGTCAACGATACGTTTTACAACGTCGTGACTGGG
806rcbc4_anti	AGCGCCCAGTCACGACGTTGTAAAACGTATCGTTGACCA
806rcbc5	GCTTATCGCACAGTAACGTTTTACAACGTCGTGACTGGG

806rcbc5_anti	AGCGCCCAGTCACGACGTTGTAAAACGTTACTGTGCGAT
806rcbc6	GCTTGTCGTGTAGCCTCGTTTTACAACGTCGTGACTGGG
806rcbc6_anti	AGCGCCCAGTCACGACGTTGTAAAACGAGGCTACACGAC
806rcbc7	GCTTAGCGGAGGTTAGCGTTTTACAACGTCGTGACTGGG
806rcbc7_anti	AGCGCCCAGTCACGACGTTGTAAAACGCTAACCTCCGCT
806rcbc8	GCTTATCCTTTGGTTCCGTTTTACAACGTCGTGACTGGG
806rcbc8_anti	AGCGCCCAGTCACGACGTTGTAAAACGGAACCAAAGGAT
806rcbc9	GCTTTACAGCGCATACCGTTTTACAACGTCGTGACTGGG
806rcbc9_anti	AGCGCCCAGTCACGACGTTGTAAAACGGTATGCGCTGTA
806rcbc10	GCTTACCGGTATGTACCGTTTTACAACGTCGTGACTGGG
806rcbc10_anti	AGCGCCCAGTCACGACGTTGTAAAACGGTACATACCGGT
806rcbc11	GCTTAATTGTGTCGGACGTTTTACAACGTCGTGACTGGG
806rcbc11_anti	AGCGCCCAGTCACGACGTTGTAAAACGTCCGACACAATT
806rcbc12	GCTTTGCATACACTGGCGTTTTACAACGTCGTGACTGGG
806rcbc12_anti	AGCGCCCAGTCACGACGTTGTAAAACGCCAGTGTATGCA
806rcbc13	GCTTAGTCGAACGAGGCGTTTTACAACGTCGTGACTGGG
806rcbc13_anti	AGCGCCCAGTCACGACGTTGTAAAACGCCTCGTTCGACT
806rcbc14	GCTTACCAGTGACTCACGTTTTACAACGTCGTGACTGGG
806rcbc14_anti	AGCGCCCAGTCACGACGTTGTAAAACGTGAGTCACTGGT
806rcbc15	GCTTGAATACCAAGTCCGTTTTACAACGTCGTGACTGGG
806rcbc15_anti	AGCGCCCAGTCACGACGTTGTAAAACGGACTTGGTATTC
806rcbc16	GCTTGTAGATCGTGTACGTTTTACAACGTCGTGACTGGG
806rcbc16_anti	AGCGCCCAGTCACGACGTTGTAAAACGTACACGATCTAC
806rcbc17	GCTTTAACGTGTGTGCCGTTTTACAACGTCGTGACTGGG
806rcbc17_anti	AGCGCCCAGTCACGACGTTGTAAAACGGCACACACGTTA
806rcbc18	GCTTCATTATGGCGTGCGTTTTACAACGTCGTGACTGGG
806rcbc18_anti	AGCGCCCAGTCACGACGTTGTAAAACGCACGCCATAATG
806rcbc19	GCTTCCAATACGCCTGCGTTTTACAACGTCGTGACTGGG
806rcbc19_anti	AGCGCCCAGTCACGACGTTGTAAAACGCAGGCGTATTGG
806rcbc20	GCTTGATCTGCGATCCCGTTTTACAACGTCGTGACTGGG
806rcbc20_anti	AGCGCCCAGTCACGACGTTGTAAAACGGGATCGCAGATC
806rcbc21	GCTTCAGCTCATCAGCCGTTTTACAACGTCGTGACTGGG
806rcbc21_anti	AGCGCCCAGTCACGACGTTGTAAAACGGCTGATGAGCTG
806rcbc22	GCTTCAAACAACAGCTCGTTTTACAACGTCGTGACTGGG
806rcbc22_anti	AGCGCCCAGTCACGACGTTGTAAAACGAGCTGTTGTTTG
806rcbc23	GCTTGCAACACCATCCCGTTTTACAACGTCGTGACTGGG
806rcbc23_anti	AGCGCCCAGTCACGACGTTGTAAAACGGGATGGTGTTGC
806rcbc24	GCTTGCGATATATCGCCGTTTTACAACGTCGTGACTGGG
806rcbc24_anti	AGCGCCCAGTCACGACGTTGTAAAACGGCGATATATCGC
806rcbc25	GCTTCGAGCAATCCTACGTTTTACAACGTCGTGACTGGG
806rcbc25_anti	AGCGCCCAGTCACGACGTTGTAAAACGTAGGATTGCTCG
806rcbc26	GCTTAGTCGTGCACATCGTTTTACAACGTCGTGACTGGG

806rcbc26_anti	AGCGCCCAGTCACGACGTTGTAAAACGATGTGCACGACT
806rcbc27	GCTTGTATCTGCGCGTCGTTTTACAACGTCGTGACTGGG
806rcbc27_anti	AGCGCCCAGTCACGACGTTGTAAAACGACGCGCAGATAC
806rcbc28	GCTTCGAGGGAAAGTCCGTTTTACAACGTCGTGACTGGG
806rcbc28_anti	AGCGCCCAGTCACGACGTTGTAAAACGGACTTTCCCTCG
806rcbc29	GCTTCAAATTCGGGATCGTTTTACAACGTCGTGACTGGG
806rcbc29_anti	AGCGCCCAGTCACGACGTTGTAAAACGATCCCGAATTTG
806rcbc30	GCTTAGATTGACCAACCGTTTTACAACGTCGTGACTGGG
806rcbc30_anti	AGCGCCCAGTCACGACGTTGTAAAACGGTTGGTCAATCT
806rcbc31	GCTTAGTTACGAGCTACGTTTTACAACGTCGTGACTGGG
806rcbc31_anti	AGCGCCCAGTCACGACGTTGTAAAACGTAGCTCGTAACT
806rcbc32	GCTTGCATATGCACTGCGTTTTACAACGTCGTGACTGGG
806rcbc32_anti	AGCGCCCAGTCACGACGTTGTAAAACGCAGTGCATATGC
806rcbc33	GCTTCAACTCCCGTGACGTTTTACAACGTCGTGACTGGG
806rcbc33_anti	AGCGCCCAGTCACGACGTTGTAAAACGTCACGGGAGTTG
806rcbc34	GCTTTTGCGTTAGCAGCGTTTTACAACGTCGTGACTGGG
806rcbc34_anti	AGCGCCCAGTCACGACGTTGTAAAACGCTGCTAACGCAA
806rcbc35	GCTTTACGAGCCCTAACGTTTTACAACGTCGTGACTGGG
806rcbc35_anti	AGCGCCCAGTCACGACGTTGTAAAACGTTAGGGCTCGTA
806rcbc36	GCTTCACTACGCTAGACGTTTTACAACGTCGTGACTGGG
806rcbc36_anti	AGCGCCCAGTCACGACGTTGTAAAACGTCTAGCGTAGTG
806rcbc37	GCTTTGCAGTCCTCGACGTTTTACAACGTCGTGACTGGG
806rcbc37_anti	AGCGCCCAGTCACGACGTTGTAAAACGTCGAGGACTGCA
806rcbc38	GCTTACCATAGCTCCGCGTTTTACAACGTCGTGACTGGG
806rcbc38_anti	AGCGCCCAGTCACGACGTTGTAAAACGCGGAGCTATGGT
806rcbc39	GCTTTCGACATCTCTTCGTTTTACAACGTCGTGACTGGG
806rcbc39_anti	AGCGCCCAGTCACGACGTTGTAAAACGAAGAGATGTCGA
806rcbc40	GCTTGAACACTTTGGACGTTTTACAACGTCGTGACTGGG
806rcbc40_anti	AGCGCCCAGTCACGACGTTGTAAAACGTCCAAAGTGTTC
806rcbc41	GCTTGAGCCATCTGTACGTTTTACAACGTCGTGACTGGG
806rcbc41_anti	AGCGCCCAGTCACGACGTTGTAAAACGTACAGATGGCTC
806rcbc42	GCTTTTGGGTACACGTCGTTTTACAACGTCGTGACTGGG
806rcbc42_anti	AGCGCCCAGTCACGACGTTGTAAAACGACGTGTACCCAA
806rcbc43	GCTTAAGGCGCTCCTTCGTTTTACAACGTCGTGACTGGG
806rcbc43_anti	AGCGCCCAGTCACGACGTTGTAAAACGAAGGAGCGCCTT
806rcbc44	GCTTTAATACGGATCGCGTTTTACAACGTCGTGACTGGG
806rcbc44_anti	AGCGCCCAGTCACGACGTTGTAAAACGCGATCCGTATTA
806rcbc45	GCTTTCGGAATTAGACCGTTTTACAACGTCGTGACTGGG
806rcbc45_anti	AGCGCCCAGTCACGACGTTGTAAAACGGTCTAATTCCGA
806rcbc46	GCTTTGTGAATTCGGACGTTTTACAACGTCGTGACTGGG
806rcbc46_anti	AGCGCCCAGTCACGACGTTGTAAAACGTCCGAATTCACA
806rcbc47	GCTTCATTCGTGGCGTCGTTTTACAACGTCGTGACTGGG

806rcbc47_anti	AGCGCCCAGTCACGACGTTGTAAAACGACGCCACGAATG
806rcbc48	GCTTTACTACGTGGCCCGTTTTACAACGTCGTGACTGGG
806rcbc48_anti	AGCGCCCAGTCACGACGTTGTAAAACGGGCCACGTAGTA
806rcbc49	GCTTGGCCAGTTCCTACGTTTTACAACGTCGTGACTGGG
806rcbc49_anti	AGCGCCCAGTCACGACGTTGTAAAACGTAGGAACTGGCC
806rcbc50	GCTTGATGTTCGCTAGCGTTTTACAACGTCGTGACTGGG
806rcbc50_anti	AGCGCCCAGTCACGACGTTGTAAAACGCTAGCGAACATC
806rcbc51	GCTTCTATCTCCTGTCCGTTTTACAACGTCGTGACTGGG
806rcbc51_anti	AGCGCCCAGTCACGACGTTGTAAAACGGACAGGAGATAG
806rcbc52	GCTTACTCACAGGAATCGTTTTACAACGTCGTGACTGGG
806rcbc52_anti	AGCGCCCAGTCACGACGTTGTAAAACGATTCCTGTGAGT
806rcbc53	GCTTATGATGAGCCTCCGTTTTACAACGTCGTGACTGGG
806rcbc53_anti	AGCGCCCAGTCACGACGTTGTAAAACGGAGGCTCATCAT
806rcbc54	GCTTGTCGACAGAGGACGTTTTACAACGTCGTGACTGGG
806rcbc54_anti	AGCGCCCAGTCACGACGTTGTAAAACGTCCTCTGTCGAC
806rcbc55	GCTTTGTCGCAAATAGCGTTTTACAACGTCGTGACTGGG
806rcbc55_anti	AGCGCCCAGTCACGACGTTGTAAAACGCTATTTGCGACA
806rcbc56	GCTTCATCCCTCTACTCGTTTTACAACGTCGTGACTGGG
806rcbc56_anti	AGCGCCCAGTCACGACGTTGTAAAACGAGTAGAGGGATG
806rcbc57	GCTTTATACCGCTGCGCGTTTTACAACGTCGTGACTGGG
806rcbc57_anti	AGCGCCCAGTCACGACGTTGTAAAACGCGCAGCGGTATA
806rcbc58	GCTTAGTTGAGGCATTCGTTTTACAACGTCGTGACTGGG
806rcbc58_anti	AGCGCCCAGTCACGACGTTGTAAAACGAATGCCTCAACT
806rcbc59	GCTTACAATAGACACCCGTTTTACAACGTCGTGACTGGG
806rcbc59_anti	AGCGCCCAGTCACGACGTTGTAAAACGGGTGTCTATTGT
806rcbc60	GCTTCGGTCAATTGACCGTTTTACAACGTCGTGACTGGG
806rcbc60_anti	AGCGCCCAGTCACGACGTTGTAAAACGGTCAATTGACCG
806rcbc61	GCTTGTGGAGTCTCATCGTTTTACAACGTCGTGACTGGG
806rcbc61_anti	AGCGCCCAGTCACGACGTTGTAAAACGATGAGACTCCAC
806rcbc62	GCTTGCTCGAAGATTCCGTTTTACAACGTCGTGACTGGG
806rcbc62_anti	AGCGCCCAGTCACGACGTTGTAAAACGGAATCTTCGAGC
806rcbc63	GCTTAGGCTTACGTGTCGTTTTACAACGTCGTGACTGGG
806rcbc63_anti	AGCGCCCAGTCACGACGTTGTAAAACGACACGTAAGCCT
806rcbc64	GCTTTCTCTACCACTCCGTTTTACAACGTCGTGACTGGG
806rcbc64_anti	AGCGCCCAGTCACGACGTTGTAAAACGGAGTGGTAGAGA
806rcbc65	GCTTACTTCCAACTTCCGTTTTACAACGTCGTGACTGGG
806rcbc65_anti	AGCGCCCAGTCACGACGTTGTAAAACGGAAGTTGGAAGT
806rcbc66	GCTTCTCACCTAGGAACGTTTTACAACGTCGTGACTGGG
806rcbc66_anti	AGCGCCCAGTCACGACGTTGTAAAACGTTCCTAGGTGAG
806rcbc67	GCTTGTGTTGTCGTGCCGTTTTACAACGTCGTGACTGGG
806rcbc67_anti	AGCGCCCAGTCACGACGTTGTAAAACGGCACGACAACAC
806rcbc68	GCTTCCACAGATCGATCGTTTTACAACGTCGTGACTGGG

806rcbc68_anti	AGCGCCCAGTCACGACGTTGTAAAACGATCGATCTGTGG
806rcbc69	GCTTTATCGACACAAGCGTTTTACAACGTCGTGACTGGG
806rcbc69_anti	AGCGCCCAGTCACGACGTTGTAAAACGCTTGTGTCGATA
806rcbc70	GCTTGATTCCGGCTCACGTTTTACAACGTCGTGACTGGG
806rcbc70_anti	AGCGCCCAGTCACGACGTTGTAAAACGTGAGCCGGAATC
806rcbc71	GCTTCGTAATTGCCGCCGTTTTACAACGTCGTGACTGGG
806rcbc71_anti	AGCGCCCAGTCACGACGTTGTAAAACGGCGGCAATTACG
806rcbc72	GCTTGGTGACTAGTTCCGTTTTACAACGTCGTGACTGGG
806rcbc72_anti	AGCGCCCAGTCACGACGTTGTAAAACGGAACTAGTCACC
806rcbc73	GCTTATGGGTTCCGTCCGTTTTACAACGTCGTGACTGGG
806rcbc73_anti	AGCGCCCAGTCACGACGTTGTAAAACGGACGGAACCCAT
806rcbc74	GCTTTAGGCATGCTTGCGTTTTACAACGTCGTGACTGGG
806rcbc74_anti	AGCGCCCAGTCACGACGTTGTAAAACGCAAGCATGCCTA
806rcbc75	GCTTAACTAGTTCAGGCGTTTTACAACGTCGTGACTGGG
806rcbc75_anti	AGCGCCCAGTCACGACGTTGTAAAACGCCTGAACTAGTT
806rcbc76	GCTTATTCTGCCGAAGCGTTTTACAACGTCGTGACTGGG
806rcbc76_anti	AGCGCCCAGTCACGACGTTGTAAAACGCTTCGGCAGAAT
806rcbc77	GCTTAGCATGTCCCGTCGTTTTACAACGTCGTGACTGGG
806rcbc77_anti	AGCGCCCAGTCACGACGTTGTAAAACGACGGGACATGCT
806rcbc78	GCTTGTACGATATGACCGTTTTACAACGTCGTGACTGGG
806rcbc78_anti	AGCGCCCAGTCACGACGTTGTAAAACGGTCATATCGTAC
806rcbc79	GCTTGTGGTGGTTTCCCGTTTTACAACGTCGTGACTGGG
806rcbc79_anti	AGCGCCCAGTCACGACGTTGTAAAACGGGAAACCACCAC
806rcbc80	GCTTTAGTATGCGCAACGTTTTACAACGTCGTGACTGGG
806rcbc80_anti	AGCGCCCAGTCACGACGTTGTAAAACGTTGCGCATACTA
806rcbc81	GCTTTGCGCTGAATGTCGTTTTACAACGTCGTGACTGGG
806rcbc81_anti	AGCGCCCAGTCACGACGTTGTAAAACGACATTCAGCGCA
806rcbc82	GCTTATGGCTGTCAGTCGTTTTACAACGTCGTGACTGGG
806rcbc82_anti	AGCGCCCAGTCACGACGTTGTAAAACGACTGACAGCCAT
806rcbc83	GCTTGTTCTCTTCTCGCGTTTTACAACGTCGTGACTGGG
806rcbc83_anti	AGCGCCCAGTCACGACGTTGTAAAACGCGAGAAGAGAAC
806rcbc84	GCTTCGTAAGATGCCTCGTTTTACAACGTCGTGACTGGG
806rcbc84_anti	AGCGCCCAGTCACGACGTTGTAAAACGAGGCATCTTACG
806rcbc85	GCTTGCGTTCTAGCTGCGTTTTACAACGTCGTGACTGGG
806rcbc85_anti	AGCGCCCAGTCACGACGTTGTAAAACGCAGCTAGAACGC
806rcbc86	GCTTGTTGTTCTGGGACGTTTTACAACGTCGTGACTGGG
806rcbc86_anti	AGCGCCCAGTCACGACGTTGTAAAACGTCCCAGAACAAC
806rcbc87	GCTTGGACTTCCAGCTCGTTTTACAACGTCGTGACTGGG
806rcbc87_anti	AGCGCCCAGTCACGACGTTGTAAAACGAGCTGGAAGTCC
806rcbc88	GCTTCTCACAACCGTGCGTTTTACAACGTCGTGACTGGG
806rcbc88_anti	AGCGCCCAGTCACGACGTTGTAAAACGCACGGTTGTGAG
806rcbc89	GCTTCTGCTATTCCTCCGTTTTACAACGTCGTGACTGGG

806rcbc89_anti	AGCGCCCAGTCACGACGTTGTAAAACGGAGGAATAGCAG
806rcbc90	GCTTATGTCACCGCTGCGTTTTACAACGTCGTGACTGGG
806rcbc90_anti	AGCGCCCAGTCACGACGTTGTAAAACGCAGCGGTGACAT
806rcbc91	GCTTTGTAACGCCGATCGTTTTACAACGTCGTGACTGGG
806rcbc91_anti	AGCGCCCAGTCACGACGTTGTAAAACGATCGGCGTTACA
806rcbc92	GCTTAGCAGAACATCTCGTTTTACAACGTCGTGACTGGG
806rcbc92_anti	AGCGCCCAGTCACGACGTTGTAAAACGAGATGTTCTGCT
806rcbc93	GCTTTGGAGTAGGTGGCGTTTTACAACGTCGTGACTGGG
806rcbc93_anti	AGCGCCCAGTCACGACGTTGTAAAACGCCACCTACTCCA
806rcbc94	GCTTTTGGCTCTATTCCGTTTTACAACGTCGTGACTGGG
806rcbc94_anti	AGCGCCCAGTCACGACGTTGTAAAACGGAATAGAGCCAA
806rcbc95	GCTTGATCCCACGTACCGTTTTACAACGTCGTGACTGGG
806rcbc95_anti	AGCGCCCAGTCACGACGTTGTAAAACGGTACGTGGGATC

Figure S. 1 Biopython script to identify reads by amplicon (plasmid ID) and sample-specific barcodes.

```
import sys
import getopt
import csv
import gzip
import itertools
import pandas as pd
import numpy as np
from Bio import SeqIO
from Bio.Seq import Seq


def fasta_to_dict(path, reverse):
	ret_dict = SeqIO.to_dict(SeqIO.parse(path, "fasta"))
	ret_dict = {k:v.seq for k, v in ret_dict.items()}
	if(reverse):
		ret_dict = {k:v.reverse_complement() for k, v in ret_dict.items()}
	ret_dict = {k:str(v) for k, v in ret_dict.items()}
	#print(ret_dict)
	return(ret_dict)

def read_plasmids_table(path):
	plasmids = {}
	with open(path, 'r') as csvfile:
		table_reader = csv.DictReader(csvfile, delimiter=',', quotechar="'")
		for row in table_reader:
			plasmids[row["Barcode name"]]= row["ID plasmid sequence"]
	return(plasmids)

def find_hash_position(sequence, tags):
	for key, value in tags.items():
		position = sequence.find(value)
		if position != -1:
			return (key, position)
	return ("none", -1)

def find_plasmid_positions(sequence, forward_dict, reverse_dict, plasmids,
orientation):
	fw_name, fw_pos = find_hash_position(sequence, forward_dict)
	rv_name, rv_pos = find_hash_position(sequence, reverse_dict)
	pl_name, pl_pos = find_hash_position(sequence, plasmids)

	if(fw_pos + rv_pos + pl_pos == -3 ):
		orientation = "."

	return ([fw_name, str(fw_pos), rv_name, str(rv_pos), pl_name,str(pl_pos),
orientation])


def write_raw_counts(fq, output_prefix, forward_dict, reverse_dict, plasmids):
	fw_f = fq
```

```
	distances_f = output_prefix + "_distances_merge.txt"
	with gzip.open(fw_f, "rt") as r1,open(distances_f,"w") as f3:

f3.write("fw_name,fw_pos,rev_name,rev_pos,pl_name,pl_pos,orientation\n")
		for fw in SeqIO.parse(r1, "fastq") :
			str_seq = str(fw.seq)
			values = find_plasmid_positions(str_seq, forward_dict, reverse_dict,
plasmids, "+")

			if values[6] == ".":
				str_seq = str(fw.seq.reverse_complement())
				values = find_plasmid_positions(str_seq, forward_dict,
reverse_dict, plasmids, "-")

			f3.write(",".join(values))
			f3.write("\n")

def write_summary(output_prefix):
	distances_f = output_prefix + "_distances_merge.txt"
	df = pd.read_csv(distances_f)
	summ=df[['fw_name', 'rev_name', 'pl_name', 'pl_pos']].groupby(['fw_name',
'rev_name', 'pl_name']).agg( ['count','mean'])
	summ.to_csv(output_prefix + "summary_merge.csv", sep=',')

def usage():
	print ("Usage: " + sys.argv[0]  + " --plasmid=<plasmids.csv> --
forward=<forward_ids.fa> --reverse=<reverse_ids.fa> --
sequences=<sequences.fq.gz> --output_prefix=<file_prefix>")


def main(argv):
	try:
		opts, args = getopt.getopt(argv,'p:f:r:s:o:h',
['plasmid=','forward=','reverse=','sequences=', 'output_prefix=', 'help'])
	except getopt.GetOptError:
		usage()
		sys.exit(2)

	if not opts:
			print ('No options supplied')
			usage()

	for opt, arg in opts:
		if opt in ('h', '--help'):
			usage()
			sys.exit(2)
		elif opt in ('p', '--plasmid'):
			plasmid_path = arg
			plamids_dict = read_plasmids_tables_table(plasmid_path)
		elif opt in ('f', '--forward'):
```

```
                forward_path = arg
                forward_dict = fasta_to_dict(forward_path, False)
            elif opt in ('r', '--reverse'):
                reverse_path = arg
                reverse_dict = fasta_to_dict(reverse_path, True)
            elif opt in ('s', '--sequences'):
                sequences_path = arg
                print(arg)
            elif opt in ('o', '--output_prefix'):
                output_prefix = arg

    write_raw_counts(sequences_path, output_prefix, forward_dict, reverse_dict,
plamids_dict)
    write_summary(output_prefix)

if __name__ == '__main__':
    main(sys.argv[1:])
```

Table S. 3. Competitiveness index (CI) statistical analysis. Ordinary one-way ANOVA with Tukey's multiple comparisons test.

Number of families= 1						
Number of comparisons per family= 10						
Alpha = 0.05						
Tukey's multiple comparisons test	**Mean Diff.**	**95.00% CI of diff.**	**Significant?**	**Summary**	**Adjusted *P Value***	
L008 vs. Rlv3841	-0.01339	-0.2269 to 0.2002	No	ns	0.9998	A-B
L008 vs. G067	-0.0377	-0.2456 to 0.1702	No	ns	0.9863	A-C
L008 vs. V116	-0.05567	-0.241 to 0.1296	No	ns	0.9164	A-D
L008 vs. G083	-0.212	-0.3847 to -0.03931	Yes	**	0.0086	A-E
Rlv3841 vs. G067	-0.02431	-0.2248 to 0.1762	No	ns	0.9971	B-C
Rlv3841 vs. V116	-0.04228	-0.2192 to 0.1346	No	ns	0.9622	B-D
Rlv3841 vs. G083	-0.1986	-0.3622 to -0.03496	Yes	**	0.0096	B-E
G067 vs. V116	-0.01797	-0.1881 to 0.1521	No	ns	0.9983	C-D
G067 vs. G083	-0.1743	-0.3305 to -0.01805	Yes	*	0.0212	C-E
V116 vs. G083	-0.1563	-0.2808 to -0.03178	Yes	**	0.0068	D-E

Test details	**Mean 1**	**Mean 2**	**Mean Diff.**	**SE of diff.**	**n1**	**n2**	**q**	**DF**
L008 vs. Rlv3841	0.1429	0.1563	-0.01339	0.07617	7	8	0.2486	67
L008 vs. G067	0.1429	0.1806	-0.0377	0.07417	7	9	0.7188	67
L008 vs. V116	0.1429	0.1985	-0.05567	0.0661	7	17	1.191	67
L008 vs. G083	0.1429	0.3548	-0.212	0.06159	7	31	4.867	67
Rlv3841 vs. G067	0.1563	0.1806	-0.02431	0.07152	8	9	0.4806	67
Rlv3841 vs. V116	0.1563	0.1985	-0.04228	0.0631	8	17	0.9475	67
Rlv3841 vs. G083	0.1563	0.3548	-0.1986	0.05837	8	31	4.812	67
G067 vs. V116	0.1806	0.1985	-0.01797	0.06067	9	17	0.4189	67
G067 vs. G083	0.1806	0.3548	-0.1743	0.05573	9	31	4.423	67
V116 vs. G083	0.1985	0.3548	-0.1563	0.04442	17	31	4.977	67

Table S. 4 Nodule GFP expression. Ordinary one-way ANOVA with Tukey's multiple comparisons test.

Number of families = 1						
Number of comparisons per family = 28						
Alpha = 0.05						
Tukey's multiple comparisons test	**Mean Diff.**	**95.00% CI of diff.**	**Significant?**	**Summary**	**Adjusted *P Value***	
L008 vs. Rlv3841	1472999	-67243 to 3013241	No	ns	0.0718	A-B
L008 vs. G067	729099	-666086 to 2124283	No	ns	0.7445	A-C
L008 vs. V116	509420	-775231 to 1794071	No	ns	0.9246	A-D
L008 vs. G083	-1205271	-2377133 to -33409	Yes	*	0.0390	A-E
L008 vs. G077/G109/H031	1532054	-521602 to 3585710	No	ns	0.3033	A-F
L008 vs. G109/H031	1407157	-335428 to 3149742	No	ns	0.2105	A-G
L008 vs. G077/G109	880685	-861900 to 2623270	No	ns	0.7755	A-H
Rlv3841 vs. G067	-743900	-2081205 to 593404	No	ns	0.6794	B-C
Rlv3841 vs. V116	-963579	-2185124 to 257965	No	ns	0.2364	B-D
Rlv3841 vs. G083	-2678270	-3780588 to -1575953	Yes	****	<0.0001	B-E
Rlv3841 vs. G077/G109/H031	59055	-1955727 to 2073837	No	ns	>0.9999	B-F
Rlv3841 vs. G109/H031	-65842	-1762441 to 1630756	No	ns	>0.9999	B-G
Rlv3841 vs. G077/G109	-592314	-2288912 to 1104285	No	ns	0.9610	B-H
G067 vs. V116	-219679	-1252329 to 812971	No	ns	0.9980	C-D
G067 vs. G083	-1934370	-2822794 to -1045945	Yes	****	<0.0001	C-E
G067 vs. G077/G109/H031	802955	-1103230 to 2709140	No	ns	0.8988	C-F
G067 vs. G109/H031	678058	-888033 to 2244149	No	ns	0.8848	C-G
G067 vs. G077/G109	151586	-1414504 to 1717677	No	ns	>0.9999	C-H
V116 vs. G083	-1714691	-2416892 to -1012490	Yes	****	<0.0001	D-E
V116 vs. G077/G109/H031	1022634	-804201 to 2849470	No	ns	0.6723	D-F
V116 vs. G109/H031	897737	-570741 to 2366215	No	ns	0.5658	D-G

V116 vs. G077/G109	371265	-1097213 to 1839743	No	ns	0.9940	D-H		
G083 vs. G077/G109/H031	2737325	987965 to 4486686	Yes	***	0.0001	E-F		
G083 vs. G109/H031	2612428	1241530 to 3983326	Yes	****	<0.0001	E-G		
G083 vs. G077/G109	2085956	715059 to 3456854	Yes	***	0.0002	E-H		
G077/G109/H031 vs. G109/H031	-124897	-2298283 to 2048488	No	ns	>0.9999	F-G		
G077/G109/H031 vs. G077/G109	-651369	-2824754 to 1522016	No	ns	0.9834	F-H		
G109/H031 vs. G077/G109	-526471	-2408678 to 1355735	No	ns	0.9889	G-H		

Test details	**Mean 1**	**Mean 2**	**Mean Diff.**	**SE of diff.**	**n1**	**n2**	**q**	**DF**
L008 vs. Rlv3841	1898616	425616	1472999	500406	7	8	4.163	138
L008 vs. G067	1898616	1169517	729099	453278	7	13	2.275	138
L008 vs. V116	1898616	1389196	509420	417367	7	23	1.726	138
L008 vs. G083	1898616	3103887	-1205271	380723	7	82	4.477	138
L008 vs. G077/G109/H031	1898616	366561	1532054	667207	7	3	3.247	138
L008 vs. G109/H031	1898616	491459	1407157	566144	7	5	3.515	138
L008 vs. G077/G109	1898616	1017930	880685	566144	7	5	2.2	138
Rlv3841 vs. G067	425616	1169517	-743900	434474	8	13	2.421	138
Rlv3841 vs. V116	425616	1389196	-963579	396865	8	23	3.434	138
Rlv3841 vs. G083	425616	3103887	-2678270	358129	8	82	10.58	138
Rlv3841 vs. G077/G109/H031	425616	366561	59055	654578	8	3	0.1276	138
Rlv3841 vs. G109/H031	425616	491459	-65842	551204	8	5	0.1689	138
Rlv3841 vs. G077/G109	425616	1017930	-592314	551204	8	5	1.52	138
G067 vs. V116	1169517	1389196	-219679	335495	13	23	0.926	138
G067 vs. G083	1169517	3103887	-1934370	288638	13	82	9.478	138
G067 vs. G077/G109/H031	1169517	366561	802955	619296	13	3	1.834	138
G067 vs. G109/H031	1169517	491459	678058	508803	13	5	1.885	138
G067 vs. G077/G109	1169517	1017930	151586	508803	13	5	0.4213	138
V116 vs. G083	1389196	3103887	-1714691	228136	23	82	10.63	138
V116 vs. G077/G109/H031	1389196	366561	1022634	593516	23	3	2.437	138

V116 vs. G109/H031	1389196	491459	897737	477090	23	5	2.661	138
V116 vs. G077/G109	1389196	1017930	371265	477090	23	5	1.101	138
G083 vs. G077/G109/H031	3103887	366561	2737325	568345	82	3	6.811	138
G083 vs. G109/H031	3103887	491459	2612428	445388	82	5	8.295	138
G083 vs. G077/G109	3103887	1017930	2085956	445388	82	5	6.623	138
G077/G109/H031 vs. G109/H031	366561	491459	-124897	706106	3	5	0.2501	138
G077/G109/H031 vs. G077/G109	366561	1017930	-651369	706106	3	5	1.305	138
G109/H031 vs. G077/G109	491459	1017930	-526471	611506	5	5	1.218	138

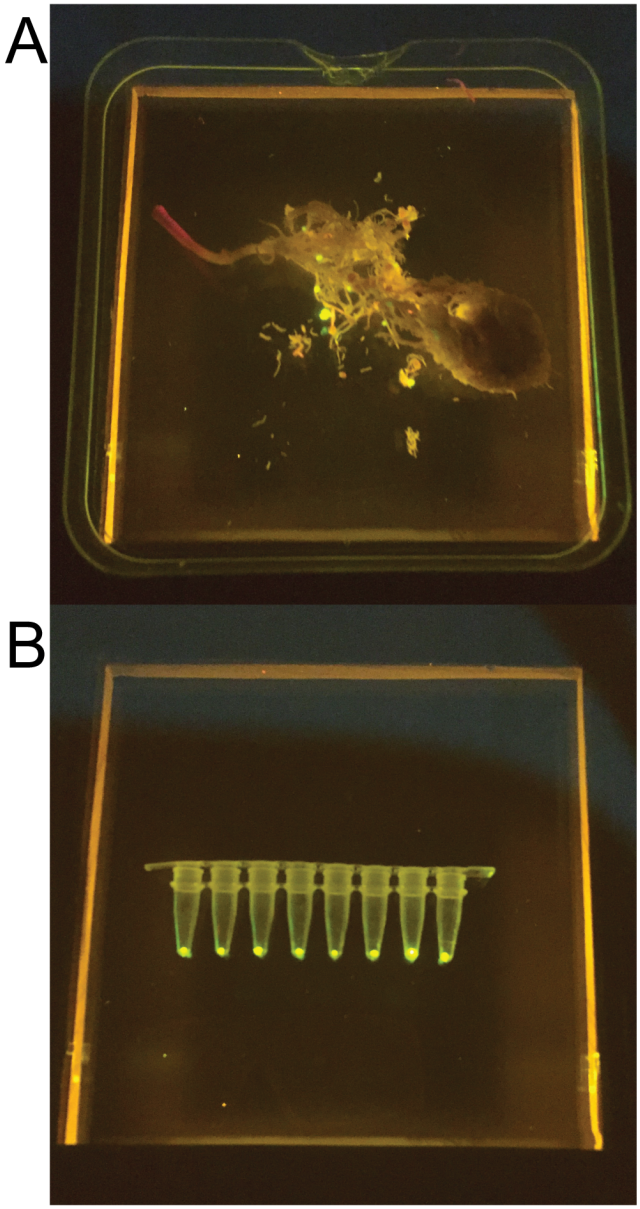


Figure S. 2 A) Pea root systems exposed to a blue-light transilluminator to check green fluorescence from nodule formed by tagged-strain expressing GFP under the control of psNifH after being storage at -80 °C for one month. B) Nodule GFP expression is still possible to observed after thaw.