Iron Homeostasis in the

Legume-Rhizobial Symbiosis

Jennifer Helen Walton

Thesis submitted to the University of East Anglia, for the degree of
Doctor of Philosophy

Department of Biological Chemistry
John Innes Centre
Norwich

September 2017

I confirm that the work in this thesis is my own original work, except where due reference is made. This thesis is submitted for the Doctor of Philosophy degree at the University of East Anglia and has not been submitted to this or any other university as part of another degree.

"The time has come," the Walrus said,
"To talk of many things"

Lewis Carroll (Through the Looking Glass
and What Alice Found There, 1872)

Abstract

All plants require iron to survive, but those that establish symbioses with nitrogen-fixing bacteria have a particularly high demand. The bacterial nitrogenase enzyme is dependent on iron-sulphur cofactors, and the host plant expresses large amounts of leghaemoglobin in infected cells to protect nitrogenase from oxygen inactivation. Using reverse genetics, this study aims to identify proteins involved in Fe transport and haem cofactor biosynthesis in the nodules of *Medicago truncatula* and its partner symbiont *Sinorhizobium meliloti*.

I identified the major (FECH1B) and minor (FECH1A) nodule ferrochelatases, enzymes responsible for the last stage of haem biosynthesis. I obtained a *Tnt1* insertion mutant in each of these and confirmed genotype by PCR. RT-PCR showed that transcript was absent in *fech1b* but not *fech1a,* most likely because the insertion in the latter is intronic. *fech1b* mutants produced fewer pink nodules than the wild type, with some morphological differences. FECH1B-eGFP localised to the plastids.

I also identified two highly upregulated nodule-specific vacuolar iron transporter-like (*VTL*) genes, *VTL4* and *SEN1*. I obtained two *Tnt1* insertion mutants in *VTL4,* and also a mutant, 13U, with a 30 kilobase deletion encompassing *VTL4* and *SEN1*; the only two nodule-expressed VTLs. 13U mutants only produced white nodules, while *vtl4* mutants produced more white nodules than the wild type. A novel bacterial iron reporter P_{mbfA}:*lux* showed that bacteria in 13U nodules perceived less iron than in wild-type nodules, with *vtl4* nodules showing an intermediate phenotype. VTL4-mCherry localised to the plasma membrane and infection thread membrane of cells in the infection zone, while SEN1-mCherry localised to the symbiosome membrane of interzone cells. Together, these data suggest that VTL4 and SEN1 mediate Fe transport to bacteroids, but at different stages of the infection process.

Acknowledgements

Firstly, I would like to thank Dr Janneke Balk, for her meticulous supervision and guidance over the last four years, and for always having an open door (at least until I start singing…) I have learnt a lot from working with her that will stand me in good stead for the future.

I would like to thank my secondary supervisors, Prof. Giles Oldroyd and Prof. Tony Maxwell, and my supervisory committee, Prof. Ray Dixon and Dr Andy Breakspear for their direction, encouragement and practical help.

The Balk group has been a really fun and friendly place to work. I would like to thank all members past and present for their friendship, help with various ridiculous experiments, and proof-reading this beast; particularly James, Rob, Emma, Andrew, Inga and Jorge.

Thank you to the various students who have contributed to this project, especially Hannah Justice for all those hours helping me count nodules and always with a smile!

From the Oldroyd group Dr Katharina Schiessl, Dr Eleni Soumpourou and Dr Doreen Feike have been generous with their time and expertise; thank you for helping me with all things Medicago and Golden Gate! Grant Calder, Dr Eva Wegel and Elaine Barclay have been similarly generous with their microscopy expertise, for which I am very grateful. Thanks also to Dr Stephen Robinson at UEA for letting me use his equipment.

I spent a really valuable month in Dr Péter Kaló's lab at MBK, Gödöllő and I would particularly like to thank Gyöngyi, Bea and Gergely for teaching me all about nodule phenotyping and microscopy techniques, and Péter and Ágota for being such generous hosts.

I would like to thank the John Innes Centre lab support and horticultural services teams for providing media and keeping my plants alive.

I would like to thank my Norwich friends for helping me out with "social skills", and making living in Norwich a great experience. Representing Team Rum: Philippa; The Captain; Skye and Tash. Team Gin: Josie; Freya; The Hannahs and Tash. Team Vodka: Koteczka; Sibyl and Tash. Finally representing Team Tea: The Biological Chemistry department, featuring Sarah and Shannon (and Tash).

Charlotte has always been on the other end of the phone with appropriately placed righteous indignation, useful(ish) scientific advice and a veritable cornucopia of memes for all occasions.

Finally, and most importantly, I would like to thank my parents and wider family, for being a constant source of love and support in the last four years and beyond. I can't imagine having got this far without them, and it is to them that this thesis is dedicated.

Table of Contents

List of Figures

List of Tables

Table of Abbreviations

ABC	ATP-binding cassette
ADP	Adenosine diphosphate
ALA	5-aminolevulinic acid
ALAD	ALA dehydratase
ALAS	ALA synthase
ATP	Adenosine triphosphate
AVG	Aminoethoxyvinylglycine
BLAST	Basic Local Alignment Search Tool
BPDS	Bathophenanthroline disulphonic acid
CCaMK	Calcium and calmodulin-dependent protein kinase
Ccc	Calcium cross complement
CIA	Cytosolic iron-sulphur assembly
CPO	coproporphyrinogen III oxidase
DMT	Divalent metal transporter
DNA	Deoxyribonucleic acid
Fe	Iron
FECH	Ferrochelatase
FER	Ferritin
Fe-S	Iron-sulphur
Fet	Ferrous transport
FLU	Fluorescent in blue light
FRD	Ferrich chelate reductase defective
FRO	Ferric reductase oxidase
FST	Flanking Sequence Tag
GFP	Green fluorescent protein
GluTR	Glutamyl-tRNA reductase
GSA	Glutamate-1-semialdehyde
GSAT	GSA aminotransferase
GUS	β-glucuronidase
HEMA	Glutamyl-tRNA reductase
HO	Haem oxygenase
HyD	Hybrid Detector
ICE	Iron Control Element
IRO	Iron responsive operator
IrrA	Iron-responsive regulator
IRT	Iron-regulated transporter
ISC	Iron-sulphur cluster assembly
IT	Infection thread
MATE	Multidrug and toxic compound extrusion
MbfA	Membrane bound ferritin A
MIT	Mitochondrial Iron Transporter
Mn	Manganese
NA	Nicotianamine

NAS	Nicotianamine synthase
Nif	Nitrogen fixation
NIN	Nodule inception
NRAMP	Natural resistance-associated macrophage protein
NTR	NADPH-dependent reductase
PBG	Porphobilinogen
PBGD	PBG deaminase
PBS	Phosphate-buffered saline
PIC	Permease in chloroplasts
PIF	Phytochrome Interacting Factor
PPOX	protoporphyrinogen IX oxidase
RirA	*Rhizobia* iron regulator
RNA	Ribonucleic acid
SIRB	Sirohydrochlorin ferrochelatase
SUF	Sulphur mobilisation pathway
UPM	uroporphyrinogen III methyltransferase
URO	uroporphyrinogen II
UROD	uroporphyrinogen III decarboxylase
UROS	Uroporphyrinogen III synthase
VIT	Vacuolar Iron Transporter
VTL	Vacuolar iron transporter-like
YS	Yellow Stripe
YSL	Yellow Stripe-like
ZIP	ZRT, IRT-like Protein
Zn	Zinc

1   Introduction

Iron (Fe) is vital for all known eukaryotes and the vast majority of bacteria and archaea (Connorton et al., 2017a). Despite being the fourth most abundant element in the Earth's crust most of this is in the insoluble Fe(III) form, forcing plants to develop a range of strategies to circumvent this and sate their appetite for iron. The importance of Fe is due to its ability to occupy a range of oxidation states from +1 to +3, making it an integral part of cofactors involved in electron transfer processes and cellular redox reactions. However, these redox properties are a double-edged sword, as they can lead to toxic redox reactions in the cytosol; the Fenton reaction (Equation 1) and subsequent reduction of Fe^{3+} by superoxide (Equation 2) generate hydroxyl radicals, leading to oxidative stress.

$$Equation\ 1: \qquad Fe^{2+} + H_2O_2 \rightarrow Fe^{3+} + OH^- + \cdot OH$$

$$Equation\ 2: \qquad Fe^{3+} + \cdot O_{2-} \rightarrow Fe^{2+} + O_2$$
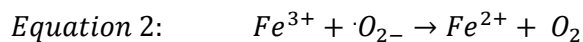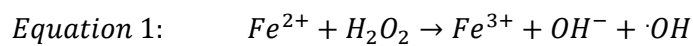
The importance of Fe and its lack of availability necessitates complex uptake and redistribution strategies, while its toxicity means that this must be carefully regulated to prevent cell death. The following section describes these strategies and mechanisms that plants have developed to this end.

1.1   Fe homeostasis in plants

1.1.1   Fe uptake

Plants typically require free, bioavailable Fe concentrations between 10^{-9} and 10^{-4} M for maximal growth, however in well-aerated soil at biologically relevant pH the concentration is nearer to 10^{-17} M (Guerinot and Yi, 1994). Plants have evolved two strategies to overcome this free Fe deficit, allowing normal growth. Strategy I, a reduction strategy, is used by non-graminaceous plants, such as the model organisms *Arabidopsis thaliana* and *Medicago truncatula*. The Graminaceae (grasses) such as *Triticum aestivum* (bread wheat) and *Zea mays* (maize) use Strategy II, a chelation-based strategy.

1.1.1.1   Strategy I uptake

Fe uptake by Strategy I is a three-stage process. The first of these, proton extrusion, capitalises on the increase in solubility of Fe^{3+} at low pH (approximately 10^9 times for a decrease of 3 pH units, Olsen et al., 1981). In Arabidopsis, the plasma membrane-localised proton transporter AHA2 (Arabidopsis H^+-ATPase 2) is primarily responsible for

rhizosphere (root-soil interface) acidification in response to Fe deficiency (Santi and Schmidt, 2009).

The second stage is the reduction of Fe(III) to Fe(II), which further increases solubility. In the Arabidopsis root epidermis this is carried out by FRO2 (ferric reductase oxidase 2, Robinson et al. 1999), aided by small molecules such as coumarins in Arabidopsis and flavins in Medicago (see below). Other members of the FRO family have distinct roles in other cell types/organelles; FRO3 and FRO8 are predicted to be mitochondrially targeted in vascular tissue, although only the latter has been found in proteomics studies of the mitochondria (Heazlewood et al., 2004). It is known however, that *FRO3* is expressed in young seedlings (mainly in root tissue), while *FRO8* is expressed in senescent leaves (Mukherjee et al., 2006; Wu et al., 2005). FRO7 is the chloroplastic isoform and is largely responsible for maintaining Fe levels within the chloroplast (Jeong and Connolly, 2009). Both FRO2 and FRO3 are strongly induced under Fe deficient conditions (Wu et al., 2005; Robinson et al., 1999; Mukherjee et al., 2006), while FRO4 and 5 are upregulated under copper deficiency to aid in reduction of Cu(II) to Cu(I) for uptake (Bernal et al., 2012).

The Medicago FRO family has been less well-characterised than their Arabidopsis counterparts. Andaluz et al. identified *MtFRO1* as being induced in roots under Fe deficient conditions (Andaluz et al., 2009). However, a later study that identified six *FRO* genes in Medicago, found *FRO1* to be weakly upregulated under Fe deficiency and only in the shoots (Orozco-Mosqueda et al., 2012). The same study found *FRO3* to be upregulated (mainly in the shoots) and *FRO4* and *FRO6* to be strongly upregulated in both roots and shoots in response to Fe deficiency. It is possible that primer degeneracy due to an older version of the Medicago genome being available for the former study could have resulted in some of this confusion.

The final stage of Strategy I uptake is transport of Fe^{2+} into the plant root. This is achieved by members of the ZRT, IRT-like (ZIP) family of proteins. In Arabidopsis, *IRT* (Iron-regulated transporter) *1* and *2* are both upregulated in response to Fe deficiency and have both been shown to transport Fe and other metals such as Zn, Mn and Cd (Eide et al., 1996; Korshunova et al., 1999; Vert et al., 2001). IRT1 is localised to the epidermis and as such is the main importer of Fe into the root (Vert et al., 2002).

Although there are sixteen published ZIP proteins in Medicago (López-Millán et al., 2004; Abreu et al., 2017), none of these has been assigned as MtIRT1. Three of the sixteen (MtZIP1, 5 and 6) have been shown to transport Zn (Stephens et al., 2011; Abreu et al.,

2017), but so far no members of the ZIP family in Medicago have been identified as Fe transporters.

1.1.1.2 Strategy II uptake

Strategy II uptake is a two-stage process for direct uptake of Fe^{3+}. The first stage is the secretion of chelating agents called phytosiderophores, such as deoxymugineic acid. In barley, deoxymugineic acid is exported by HvTOM1 (Nozoye et al., 2011). The Fe^{3+}-phytosiderophore complex is then taken up by YS1 (yellow stripe 1) in maize and the related YSL15 (Yellow Stripe-like) in rice (Curie et al., 2001; Inoue et al., 2009).

1.1.1.3 Combined Strategy I and II uptake

The absolute classification of plants as either Strategy I or Strategy II has been called into question by the identification of Strategy I components (namely IRT1) in some Graminaceae such as rice (Bughio et al., 2002) and barley (Pedas et al., 2008). Other Strategy I components have not been identified in rice, but it is hypothesised that the partially submerged and hypoxic conditions of rice growth result in a higher proportion of Fe^{2+} than is present in the soil of other Strategy II plants, making expression of *IRT* alone sufficient to increase Fe uptake.

Conversely plants originally classified as Strategy I have been shown to export complex molecules such as organic acids, phenolics and flavins in response to Fe deficiency (Rodríguez-Celma et al., 2013; Abadía et al., 2002). Arabidopsis has been shown to secrete coumarins when grown in soils with a high pH (Schmid et al., 2014). Coumarins can chelate Fe through the two hydroxyl groups of the catechol moiety, facilitating Fe mobilisation. By contrast Medicago secretes flavins, the exact role of which is unknown, however they are thought to aid in the reduction of Fe^{3+} in the rhizosphere (Rodríguez-Celma et al., 2011).

1.1.2 Fe transport

Once imported into the root, Fe needs to be transported to the rest of the plant to be incorporated into protein cofactors crucial for a number of processes including photosynthesis and respiration, with the additional demand of the bacterial cofactors required for nitrogen fixation in legumes. Once taken up into cells in the sink tissue, Fe is transported into the relevant organelles for cofactor synthesis, or storage until required.

1.1.2.1 Intercellular Fe transport

In the root, Fe can initially follow either an apoplastic or symplastic path. In the former, Fe moves through the apoplastic space of epidermal and cortical cell walls until it reaches the impermeable suberised layer of the Casparian strip. At this juncture Fe is forced into the symplastic route (Barberon, 2017). Fe is initially taken up across the

epidermal membrane by IRT1 (in combination with Natural resistance-associated macrophage protein 1 (NRAMP1) in Fe replete conditions (Castaings et al., 2016)). Fe can then pass between cells through the plasmodesmata to the vasculature. To prevent both toxicity and precipitation, Fe^{2+} in the symplast is thought to be complexed by nicotianamine (NA). NA is synthesised from the cofactor S-adenosyl methionine by the nicotianamine synthase (NAS) enzymes, which have been identified in both graminaceous and non-graminaceous plants including rice, bread wheat and Arabidopsis (Klatte et al., 2009; Bonneau et al., 2016; Inoue et al., 2003). A *NAS* gene has recently been identified in Medicago and hypothesised to be important in nodule function, although the data suggests that it is mainly involved in nodule maintenance and stress responses (Avenhaus et al., 2016).

Fe is transported to the rest of the plant from root to shoot through the xylem vessels, which it accesses by the pericycle, and in the opposite direction (shoot to root) through the phloem. To be loaded into the xylem the Fe must first be exported into the apoplast, probably through YSL2 and ferroportin, although this has not yet been confirmed by transport assays (DiDonato et al., 2004; Morrissey et al., 2009). Fe in the xylem is most commonly found in the oxidised (Fe^{3+}) form in complex with citrate, which is exported from xylem parenchyma cells in both source and sink tissues by FRD3 in Arabidopsis (Durrett et al., 2007; Rellán-Álvarez et al., 2010; Roschzttardtz et al., 2011). In Medicago, MATE69 is thought to function to fulfil this role; *MtMATE69* expression was found to be induced in response to Fe deficiency, and overexpression of *MtMATE69* resulted in increased citric acid extrusion and Fe accumulation (Wang et al., 2017). MtMATE66 is also thought to function partially redundantly with MtMATE69, but is not specific to Fe deficiency responses.

1.1.2.2   Intracellular Fe transport

When unloaded from the xylem, Fe^{3+} is reduced to Fe^{2+} by FRO proteins and again forms a complex with NA. Once inside the cells of sink tissues (such as leaves, seeds or nodules for example) Fe is distributed to the relevant organelles by a range of transporters. Fe is taken up by plastids and mitochondria (as the sites of the majority of cofactor biosynthesis) by transporters described in detail below.

Fe transport into chloroplasts is well characterised. As previously mentioned, AtFRO7 is largely responsible for maintaining chloroplast Fe levels, along with Permease in chloroplasts 1 (PIC1) (Jeong et al., 2008; Duy et al., 2011). YSL4 and YSL6 are localised to the chloroplast envelope, with a role in Fe export if concentrations are too high (Divol et al., 2013).

By contrast, Fe transport into mitochondria is less well understood. As previously mentioned, FRO3 and FRO8 are thought to be mitochondrially localised, but only FRO8 has been found in proteomic studies of the mitochondria (Jain et al., 2014; Heazlewood et al., 2004). Mitochondrial Iron Transporter 1 (MIT1) has been shown to be vital for normal growth of rice plants, and complements the yeast mitochondrial Fe uptake mutant Δ*mrs3*Δ*mrs4* (Bashir et al., 2011).

To prevent toxic Fe build-up in the cytosol Fe can be stored in the vacuole and remobilised when needed for cofactor biosynthesis and other processes. In plants, Fe is transported into the vacuole in developing seeds by Vacuolar Iron Transporter (VIT) proteins (Kim et al., 2006). In Arabidopsis, Fe is transported out of the vacuole mainly during seed germination by the functionally redundant NRAMP3 and NRAMP4 proteins (Lanquar et al., 2005). The VIT family of proteins have close sequence similarity to the *Saccharomyces cerevisiae* protein calcium cross complement 1 (Ccc1). Arabidopsis VIT1 can restore Fe transport in yeast Δ*ccc1* mutants (Kim et al., 2006).

Closely related to VIT are the VTL family (VIT-like proteins). Five of these have been characterised in Arabidopsis, three of which (VTL 1, 2 and 5) have been shown to be downregulated in response to Fe deficiency (Gollhofer et al., 2011). *VTL1* and *VTL2* partially complement the yeast Δ*ccc1* mutant and VTLs 1,2 and 5 all cause an increase in Fe accumulation in isolated vacuoles in a Δ*ccc1* background, although this is not as profound as the complementation using *AtVIT1* (Gollhofer et al., 2014; Kim et al., 2006). The characteristic of the VTL family that distinguishes them from VIT is the deletion of a cytosolic loop region of approximately 40 amino acid residues between the putative transmembrane helices 2 and 3 (Gollhofer et al., 2011). The function of this loop region, as well as the mode of action of the VTLs (and indeed the direction of transport) remains to be elucidated.

Fe can also be sequestered in ferritin, a 24-subunit protein cage which entraps as many as 2500 Fe atoms in the form of Fe (III) oxide, Fe_2O_3 (Masuda et al., 2010). Plant and mammalian ferritin are very similar, both in terms of amino acid composition and three-dimensional structure, however their roles *in vivo* are not. In mammals, ferritins are the primary cellular Fe storage sink, whereas in plants their primary role is protection from oxidative stress by trapping free Fe, and preventing the Fenton reaction (Ravet et al., 2009).

1.1.3 Fe cofactor biosynthesis

Once safely gathered into the relevant organelles, Fe can finally be incorporated into crucial cofactors which, when inserted into their target proteins, are necessary for a number of fundamental processes.

Iron-sulphur (Fe-S) clusters are needed for the proteins in the electron transport chains of both respiration and photosynthesis, as well as being at the core of dinitrogenase and dinitrogenase reductase, the enzymes responsible for the reduction of dinitrogen by rhizobia (Section 1.2.2). Similarly, haem has crucial functions in all plants: in peroxidases (detoxification of hydrogen peroxide), α-dioxygenases (oxygenation of fatty acids) haemoglobin (nitric oxide scavenging) and leghaemoglobin (Section 1.2.2).

The pathways associated with Fe-S cluster assembly and haem biosynthesis are discussed below.

1.1.3.1 Fe-S cluster assembly

There are three distinct machineries involved in Fe-S cluster assembly in plants: the sulphur mobilisation (SUF) pathway in plastids; the iron-sulphur cluster (ISC) pathway in mitochondria and the cytosolic iron-sulphur cluster assembly (CIA) pathway. In all these processes the formation of Fe-S holoproteins occurs in three stages: the formation of persulphide by a cysteine desulphurase; its combination with Fe on a scaffold protein; release of the cluster from the scaffold protein; transfer to and incorporation into its target apoprotein. Figure 1.1 summarises the key players in the SUF, ISC and CIA pathways.

Persulphide generation is catalysed by NFS2 in plastids and NFS1 in mitochondria, with SUFE1 and ISD11 acting as respective helper proteins (Balk and Schaedler, 2014). The resulting persulphides are transferred to protein scaffolds. In plastids this scaffold consists of $SUFBC_2D$, which is interesting in its resemblance to ATP-binding cassette (ABC) transporters while having cytoplasmic localisation (Wollers et al., 2010). In mitochondria the scaffold protein is thought to be ISU1, which forms a homodimer that goes through extensive conformational change over the course of cluster assembly (Frazzon et al., 2007; Jin et al., 2009). There are a number of other proteins proposed to have a role in cluster transfer and insertion into target proteins. These include NFU2, SUFA and GRXS16 in the plastids, which are functionally redundant, as illustrated by the viability of their respective knockout mutants (Touraine et al., 2004; Cheng et al., 2006; Couturier et al., 2013).
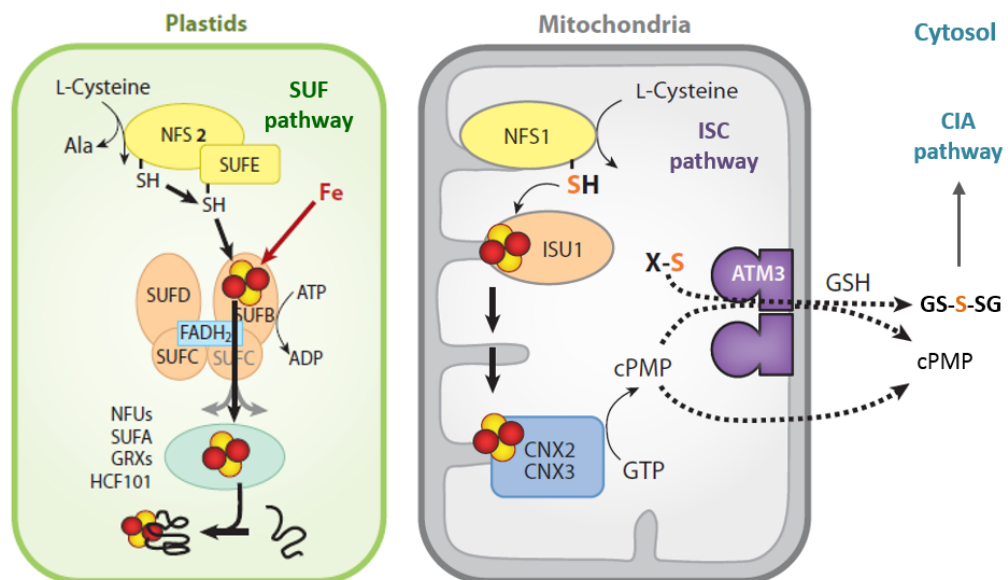
Figure 1.1 Summary of the main Fe-S cluster assembly machineries in plants. Figure adapted from Balk & Schaedler, 2014.

In plants the origin of the Fe incorporated is not yet known. It was originally thought to be the Fe storage protein ferritin, however this was rejected following the studies of Arabidopsis ferritin triple mutants by Ravet et al. where they observe that there is no disruption to the electron transport chain, consistent with cluster assembly being unaffected (Ravet et al., 2009).

Fe-S clusters for proteins localised to the cytosol and the nucleus are made by the CIA machinery. Unlike the SUF and ISC pathways which function independently, the CIA pathway relies on the ISC system to provide the persulphide for cluster biosynthesis. For the CIA pathway to be functional an ABC transporter of the mitochondria (ATM3) is required (Bernard et al., 2013, 2009). This is a half-transporter localised to the inner mitochondrial membrane, that is proposed to transport sulphur in the form of glutathione trisulphide (GS-S-SG) that can then be used for cytosolic Fe-S assembly (Schaedler et al., 2014; Perez-Ruiz, 2006).

1.1.3.2 Haem cofactor synthesis

In plants, haem is synthesised in a ten step process, the majority of which is common to all the families of tetrapyrroles (sirohaem, chlorophyll and phytochromobilin). This is illustrated in Figure 1.2.

In the first step of tetrapyrrole synthesis, glutamate is activated by conjugation with its tRNA by glutamyl tRNA synthetase **(1)**. The activated carboxylic acid group is then reduced to an aldehyde by glutamyl tRNA reductase (GluTR, or HEMA) **(2)** (Vothknecht et al., 1997; Moser et al., 2001). The glutamate-1-semialdehyde (GSA) produced by this reaction is converted to 5-aminolevulinic acid (ALA) by GSA aminotransferase (GSAT) **(3)** (Hofgen et al., 1994; Tsang et al., 2003), which effects exchange of the amine and aldehyde groups. The next stage, catalysed by ALA dehydratase (ALAD) **(4)** is a cyclisation by the condensation of two molecules of ALA to form porphobilinogen (PBG), four molecules of which go on to form the backbone of the porphyrin ring structure, hydroxymethylbilane. This is a linear molecule, formed by the action of PBG deaminase (PBGD) **(5)**. The next step, catalysed by uroporphyrinogen III synthase (UROS) **(6)** finally forms the closed ring structure and is the last common enzyme between all four pathways.

The first pathway to branch off is the sirohaem pathway (shown in purple in Figure 1). The three enzymes involved are an S-adenosyl-L-methionine-dependent methyltransferase **(S1)**, an unidentified oxidase **(S2)** and sirohydrochlorin ferrochelatase **(S3)** (Raux-Deery et al., 2005; Leustek et al., 1997). The first of these is responsible for methylating uroporphyrinogen III to make dihydrosiroydrochlorin. This is then oxidised to produce sirohydrochlorin before finally Fe is inserted by **S3** to give the active cofactor. Sirohaem is crucial for the function of both nitrite and sulphite reductases, so plays an active role in assimilation of both nitrogen and sulphur.

Chlorophyll, haem and phytochromobilin continue to share the same pathway for another three steps, before the next branch point. In the first of these steps, uroporphyrinogen III is decarboxylated to give coproporphyrinogen III by uroporphyrinogen III decarboxylase (UROD) **(7)**. Two of the remaining carboxylate groups are subsequently also decarboxylated, this time by the action of an oxygen-dependent enzyme, coproporphyrinogen III oxidase (CPO) **(8)** (Santana et al., 2002). The final common step for the remaining three tetrapyrrole families is the six electron oxidation of protoporphyrinogen IX to yield protoporphyrin IX, in a reaction catalysed by protoporphyrinogen IX oxidase (PPOX) **(9)** (Jacobs and Jacobs, 1987).
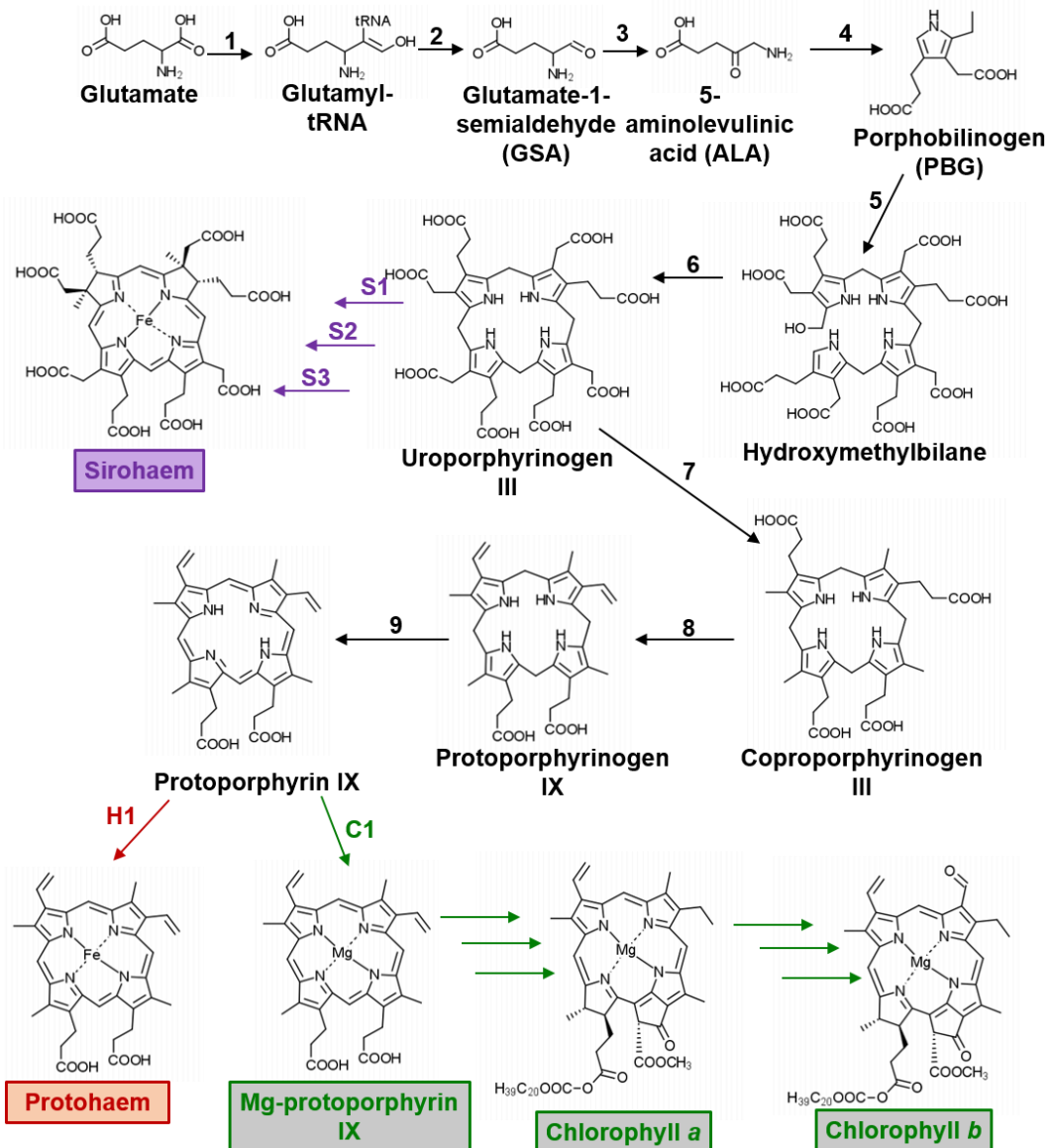
Figure 1.2 Scheme of tetrapyrrole biosynthesis in plants.
Common steps are shown in black, sirohaem-specific steps in purple, haem-specific steps in red and chlorophyll specific in green. Structures created using ChemDraw software.

The chlorophyll synthesis pathway branches off at this point, with protoporphyrin IX first having its metal centre inserted by magnesium chelatase **(C1)** in an ATP-dependent manner. The Mg-protoporphyrin IX then undergoes several more enzymatic transformations (thoroughly reviewed by Tanaka & Tanaka) before yielding chlorophyll *a*, the main electron donor in the electron transport chain in photosynthesis. Some of the chlorophyll *a* enters a cycle to convert it to chlorophyll *b*, the major constituent of the light-harvesting complexes in land plants (Tanaka and Tanaka, 2007).

The haem branch begins with the metal insertion analogous to that in chlorophyll biosynthesis. Fe^{2+} insertion is catalysed by ferrochelatase (FECH) enzymes **(H1)**, of

which there are two in Arabidopsis (AtFECH1 and AtFECH2). AtFECH1 is basally expressed in all tissues and thought to be dual-localised to both the mitochondria and plastids while AtFECH2 is responsible for haem production in photosynthetic tissue and in response to stress (Scharfenberg et al., 2014). Haem biosynthesis is conserved across plants.

Phytochromobilins, the fourth family of tetrapyrroles are formed by the cleavage of protohaem to yield linear molecules, used by phytochromes as the chromophore, so have a major role in light signalling.

1.1.3.3 Regulation of haem cofactor biosynthesis

Due to the highly light-sensitive nature of the tetrapyrrole intermediates, the biosynthetic pathway is tightly regulated by a number of different means, including transcriptional regulation, redox regulation and post-translational regulation. These are outlined briefly below.

The first point of regulation of haem biosynthesis is GluTR (labelled **2** in Figure 1.2). In Arabidopsis there are two isoforms of GluTR, HEMA1 and HEMA2 which are differentially regulated. HEMA1 is expressed in photosynthetically active tissues, and is repressed in the dark by interaction with Phytochrome Interacting Factors 1 (PIF1) and PIF3 (Stephenson et al., 2009). This can be decoupled from light-dependence by the action of cytokinins and blocking of gibberellins (Hudson et al., 2011). By contrast, HEMA2 transcription is controlled by sugar availability, with higher concentrations of sucrose lowering expression (Ujwal et al., 2002).

As with transcriptional regulation, there are both light-dependent and light-independent modes of redox regulation. Thioredoxins are vital for tetrapyrrole synthesis in light tissues, as they are reduced by electrons in the electron transport chain of photosynthesis to allow close linking of demand for chlorophyll with its production by the plastids (Richter and Grimm, 2013). In tissues not exposed to the light, thioredoxins are reduced by NADPH-dependent reductases (NTRs). The chloroplastic NTR is unique in that it contains a thioredoxin domain, effectively decoupling it from light-dependence (Serrato et al., 2004; Perez-Ruiz, 2006). Again, GluTR is one of the targets of this modification (Richter et al., 2013).

HEMA1 is also regulated by protein-protein interactions with the repressor protein FLU (Fluorescent in blue light), which binds HEMA1 in the dark to block the pathway under these conditions (Meskauskiene et al., 2001). Haem is independently involved in feedback inhibition, by binding to a distinct site on HEMA1, at the opposite end of the protein from FLU (Goslings et al., 2004).

It has also been suggested that phosphorylation could be important in regulation of tetrapyrrole synthesis, but research into this area is still in its infancy (Brzezowski et al. 2015).

1.2 The Legume-rhizobial symbiosis

The legume-rhizobial symbiosis is a mutually beneficial relationship between nitrogen-fixing soil bacteria known collectively as the Rhizobia, and host plants in the legume family. Bacteria are housed in specialised root structures called nodules, where they are supplied with a carbon source and a microaerobic environment in return for fixed nitrogen.

1.2.1 Symbiosis signalling and nodule establishment

Symbiosis signalling is initiated by the plant; in low nitrate conditions they secrete flavonoids from their roots. These are detected by rhizobia in the rhizosphere, prompting compatible rhizobia bacterial partners to secrete Nodulation factors (Nod factors) (Liu and Murray, 2016). Detection of Nod factors by the plant occurs at specific receptor-like kinases containing a LysM domain (Kelly et al., 2017). This results in nuclear calcium spiking in the plant cell, which is decoded by a calcium and calmodulin-dependent protein kinase (CCaMK) (Miller et al., 2013). CCaMK phosphorylates the DNA-binding protein CYCLOPS, which activates transcription of transcription factors necessary for nodule organogenesis, such as ERN1 and NIN (Singh et al., 2014; Cerri et al., 2017). Concomitantly, rhizobia adhere to root hairs, which in turn curl around the bacteria, forming the "infection pocket" (Murray, 2011). Nod factors also stimulate formation of the preinfection thread; the alignment of the cytoskeleton and formation of cytoplasmic bridges along the path that the infection thread will follow (Van Brussel et al., 1992).

Infection threads are effectively the inversion of root hair growth, requiring both the formation of the preinfection thread and the plant-controlled degradation of the plant cell wall within the infection pocket (Xie et al., 2012), allowing the bacteria to be taken up through an invagination in the plant plasma membrane. The infection thread then progresses down the root hair by the fusion of vesicles on to the growing end along an actin scaffold, in a manner thought to be analogous to pollen tube growth (Oldroyd et al., 2011). The infection thread continues down to the nodule primordia, where cortical cells have been dividing since the initial calcium signalling event. Bacteria are delivered to nodule cells surrounded by plant-derived membrane with properties of plasma membrane and vesicles (Limpens et al., 2009). This means that at no point are the bacteria released into the cytosol, which could trigger a plant immune response.

Legume nodules are mainly determinate or indeterminate. Soybean (*Glycine max*) and Lotus (*Lotus japonicas*) produce spherical determinate nodules, which contain bacteroids developing in a synchronised manner, meaning that they are all at consistently the same developmental stage. By contrast Medicago and pea produce

indeterminate nodules, which have an elongated shape and always have an actively dividing nodule meristem. This means that a mature nodule contains zones representing the different stages of nodule development.

In Medicago, the bacteria undergo extensive elongation and differentiation during their maturation into nitrogen-fixing bacteroids, largely controlled by the partial pressure of oxygen in nodule cells. When in their nitrogen-fixing state, the bacteroids are highly respiratorily active, so have a high oxygen requirement. By contrast the nitrogenase enzyme is highly oxygen sensitive, so the partial pressure of oxygen in the bacteroids needs to be kept at a minimum (Dixon and Kahn, 2004). This is achieved by the protein leghaemoglobin, which is present at concentrations of approximately 3 mM in nodules, and is responsible for maintaining the anoxic environment of the bacteroids, enabling them to effect the reduction of dinitrogen by nitrogenase (Bergersen and Appleby, 1981).

1.2.2 Biochemistry of nitrogen fixation

When carried out industrially using the Haber-Bosch process, the process of nitrogen fixation requires temperatures of 500 °C and 200 atmospheres pressure in addition to the anaerobic conditions, so the bacteroids' ability to carry this out is extraordinary.

The dinitrogenase enzyme (MoFe protein), carries out the 8 electron reduction shown below in Equation 3, and is itself reduced by dinitrogenase reductase (Fe protein).

$$Equation\ 3: \quad N_2 + 8H^+ + 8e^-\ 16MgATP \rightarrow 2NH_3 + H_2 + 16MgADP + 16PO_4^{3-}$$

Both of these proteins contain a myriad of interesting Fe-S clusters for the electron transfers to carry out the reduction of nitrogen. These are described in more detail in Sections 1.2.2.1 and 1.2.2.2.

1.2.2.1 The MoFe protein

The MoFe protein is an $\alpha_2\beta_2$ heterotetramer, each $\alpha\beta$ heterodimer of which contains two Fe-S-derived clusters (Figure 1.3). These are the M-cluster; an unique 7Fe-Mo-9S-homocitrate-C cluster that is the actual site of nitrogen fixation (Figure 1.3b) and the P-cluster, an 8Fe-7S cluster (Figure 1.3c). Nitrogen is thought to bind to the substrate-binding face of the M-cluster, the atoms of which are labelled in Figure 1.3b (Barney et al., 2004).
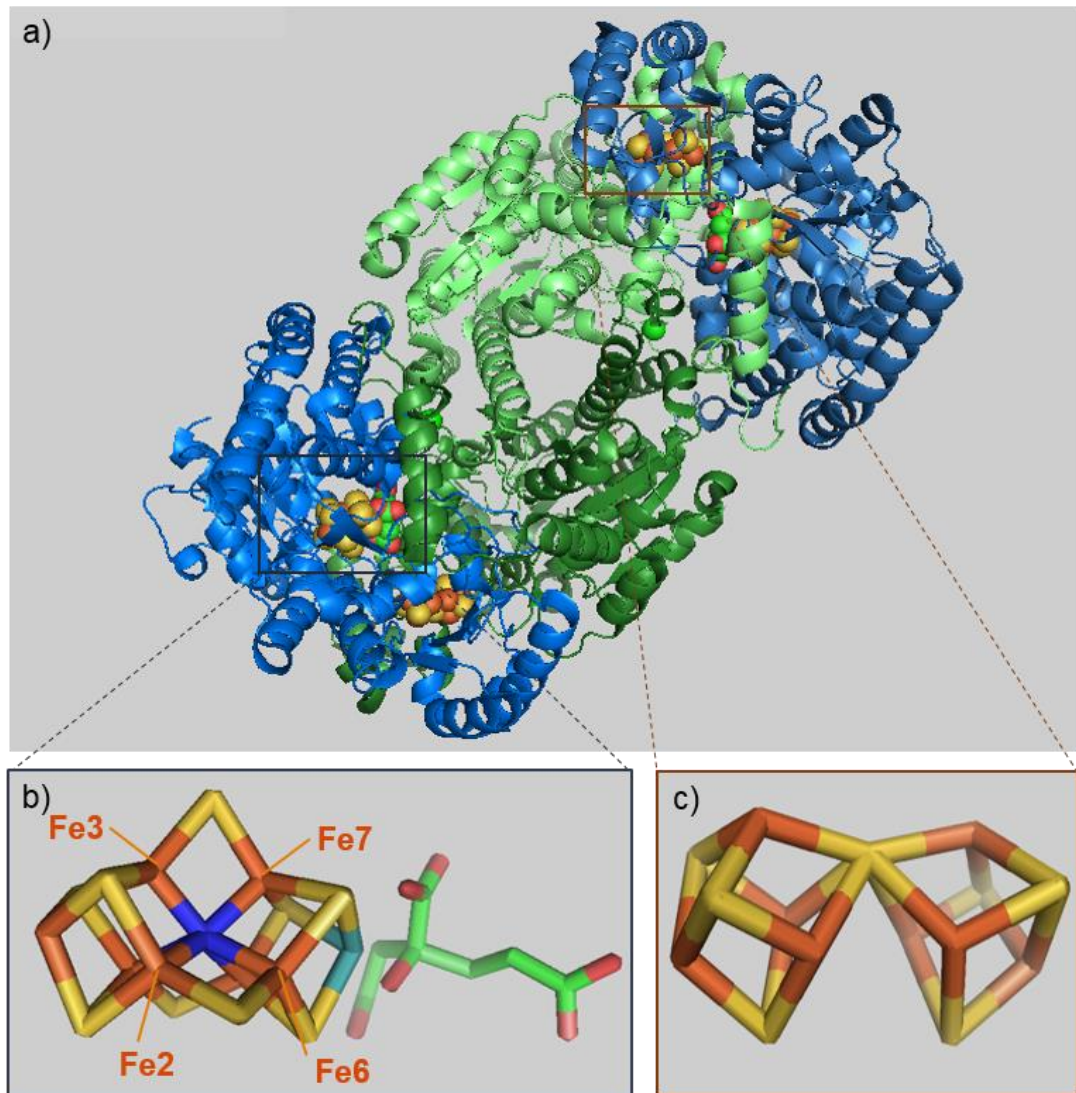
Figure 1.3 The MoFe protein from the free-living nitrogen-fixing _Azotobacter vinelandii_ a) the chains of the α- and β-subunits shown in blue and green respectively b) The M-cluster, situated in the α-subunit of the MoFe protein (Schindelin et al. 1997), with associated homocitrate bound at the Mo centre of the cofactor, and the Fe ions on what is believed to be the substrate-binding face labelled. c) The P-cluster, situated at the interface between α- and β-subunits of the MoFe protein, in its reduced form. Colour key of atoms in cofactors: orange, iron; yellow, sulphur; red, oxygen; green, carbon; blue, carbide; turquoise, molybdenum. Images created from PDB file 1M1N, using PyMol molecular modelling software.

The central atom in the M-cluster has been shown by high resolution X-ray crystallography to be a carbide (C^{4-}) ion (Lancaster et al., 2011; Spatzal et al., 2011); the only place this form of carbon has been observed in nature. During cluster maturation, two Fe_4-S_4 clusters (known as the K-cluster) are supported on the NifB scaffold, and the carbide inserted between them by a radical-SAM reaction, along with another S to make the L-cluster (Wiig et al., 2015). This is then transferred to the NifEN scaffold where if finally forms the mature M-cluster after NifH delivers both Mo and homocitrate, removing an Fe in the process (Fay et al., 2016).

Homocitrate is essential for the correct functioning of nitrogenase (Imperial et al., 1989). Although its exact function is unknown, it is a bidentate chelator of the terminal Mo of the M-cluster (Einsle et al., 2002). In the symbiotic nitrogen-fixing species the gene responsible for homocitrate synthesis (NifV) has been lost in a horizontal gene transfer event; this means that rhizobia such as *Sinorhizobium meliloti* are now incapable of fixing nitrogen independently, and have effectively been evolutionarily enslaved by their host legumes (Hakoyama et al., 2009).

1.2.2.2 The Fe protein

The Fe protein (NifH) is a homodimer, each monomer of which has a nucleotide-binding site, with a 4Fe-4S cluster bridging the interface between the two monomers. This cluster is also close to the interface between the Fe protein and the MoFe protein to allow electron transfer between it and the P-cluster of the MoFe protein (Seefeldt et al., 2009). The Fe protein goes through a four-stage cycle (the Fe protein cycle) to transfer single electrons to the MoFe protein.

Initially, the Fe protein is in a reduced state, with two molecules of MgATP bound. In the first stage, this forms a transient complex with the MoFe protein. It is thought that conformational changes induced by this association cause electron transfer from the P-cluster of the MoFe protein to the M-cluster, leaving the P-cluster in an oxidised state. The oxidised P-cluster then receives an electron from the reduced Fe protein (Danyal et al., 2011). The second step is ATP hydrolysis, distinct from the third step of inorganic phosphate release from the Fe protein (Duval et al., 2013). The fourth step is release of the oxidised Fe protein, still with two molecules of ADP bound to the nucleotide-binding sites on each monomer. This process is repeated eight times for every molecule of dinitrogen fixed.

1.2.3 *Medicago truncatula* as a model legume

The legume *Medicago truncatula*, a close relative of the common forage crop Alfalfa (*Medicago sativa*), has long been used as a model species for the study of the legume-Rhizobial symbiosis, along with its partner symbiont *Sinorhizobium meliloti*. There are several advantages to using Medicago over other model legumes; the genome has been sequenced and refined to a high level, an advantage over pea (*Pisum sativum*) and some other model legumes, for which genetic information is lagging behind, such as chickpea (*Cicer arietinum*) and bean (*Phaseolus vulgaris*). In addition the indeterminate nature of Medicago nodules lends itself well to developmental studies.

Syto™ 13 is a fluorescent DNA stain that is useful for identifying the different zones in Medicago nodules, which are characterised by the developmental stage of the bacteria (Figure 1.4). Bacteria in the infection zone of nodules have a similar shape to free-living (Tic Tac®-shaped). They start to elongate away from this in the differentiation zone (the brightest band of cells in Figure 1.4) before elongating further and aligning themselves around the central vacuole in the early nitrogen fixation zone. The presence of all stages of symbiosis development from infection right through to functional nitrogen-fixing zone cells means that effectively every nodule can be used as a time course of nodule development.
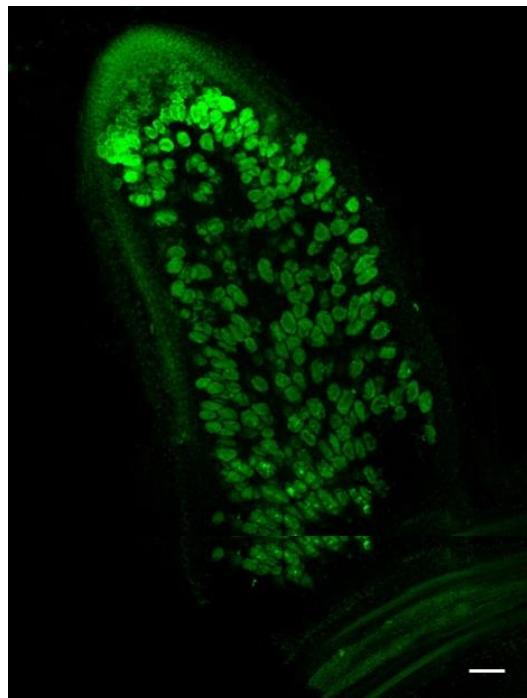
Figure 1.4 Medicago nodule section stained Syto™ 13
70 µm section of a 14 dpi nodule from Medicago (R108), scale bar represents 100 µm. Image captured using Leica SP8 microscope by Gergely Iski

1.3 Fe and the legume-rhizobial symbiosis

Fe is critical for all plants to function correctly, but the legume-rhizobial symbiosis has even greater demands for the metal than other plants. Both the nitrogenase protein itself and the leghaemoglobin needed to maintain a suitably anoxic environment for the correct functioning of nitrogenase are dependent on Fe-based cofactors. Symbiosis establishment has such a high demand for Fe that nodulated *Lupinus angustifolius* plants require almost five times more Fe to achieve their growth potential than nitrate-supplemented plants. Without sufficient Fe nodules will not form, even if the legume is grown in low-nitrogen conditions that would normally be conducive to symbiosis establishment (Tang et al., 1990).

1.3.1 Bacterial Fe homeostasis

In many bacteria, Fe-regulated transcription is controlled by the ferric uptake regulator (Fur). When Fe is bound to Fur, it acts as a repressor of Fe uptake genes such as siderophore synthesis and Fe transport. Under low Fe conditions Fe is no longer bound to Fur, which dissociates from the "Fur box" in the promoters of the genes of the Fe operon, allowing transcription (Lee and Helmann, 2007). The Fur homologue in the α-proteobacteria (which includes rhizobia) is a sensor for Mn, so has been renamed Mur (Johnston et al., 2007). Instead the rhizobia have evolved two different indirect strategies to sense Fe levels, using either Fe-S clusters or haem. These function in concert to regulate Fe homeostasis genes under both low and high Fe conditions.

RirA (*Rhizobia* iron regulator) is a repressor protein that binds to the Iron-responsive operator (IRO) box in the promoter of Fe uptake genes when it is bound to a 2Fe-2S cluster. Under low Fe conditions (*i.e.* with Fe-S unbound) RirA dissociates from the IRO box, allowing transcription of the Fe uptake genes, which include haem-uptake (*hmu*) and vicibactin (a siderophore) synthesis (*vbs*) (Todd et al., 2005; Chao et al., 2005). Under high Fe conditions there are more Fe-S clusters available to bind RirA, allowing it to bind to the IRO box and block transcription.

IrrA (Iron-responsive regulator) functions as a repressor under low Fe conditions, when it is bound to iron control elements, or "ICE boxes" (Singleton et al., 2010). Under high Fe conditions haem binds to Irr, resulting in Irr dissociating from the ICE box and allowing transcription of genes. Genes under the control of IrrA in rhizobia include a putative Fe-siderophore transporter (Irp6) and a fumarase hydratase known to contain a 4Fe-4S cluster (FumA) (Todd et al., 2006). In some species (such as *Bradyrhizobium japonicum*)

Irr is degraded on haem binding, but this is by no means ubiquitous among the α-proteobacteria (Qi and O'Brian, 2002). Interestingly the *suf* operon (Fe-S cluster assembly) is controlled by both of these genes, making sure this critical machinery is functional at all levels of Fe availability (Todd et al., 2006; Chao et al., 2005).

1.3.2 Legume nodule Fe homeostasis

As previously discussed, Fe needs to get to the bacteroid so it can be incorporated into the cofactors for nitrogen fixation. To reach this point, it is transported across three distinct membranes of the infected cell; the plasma membrane, the plant-derived symbiosome membrane and the bacteroid membrane. This is represented in Figure 1.5. The majority of the transporters involved remain to be identified, however those already characterised are outlined below.
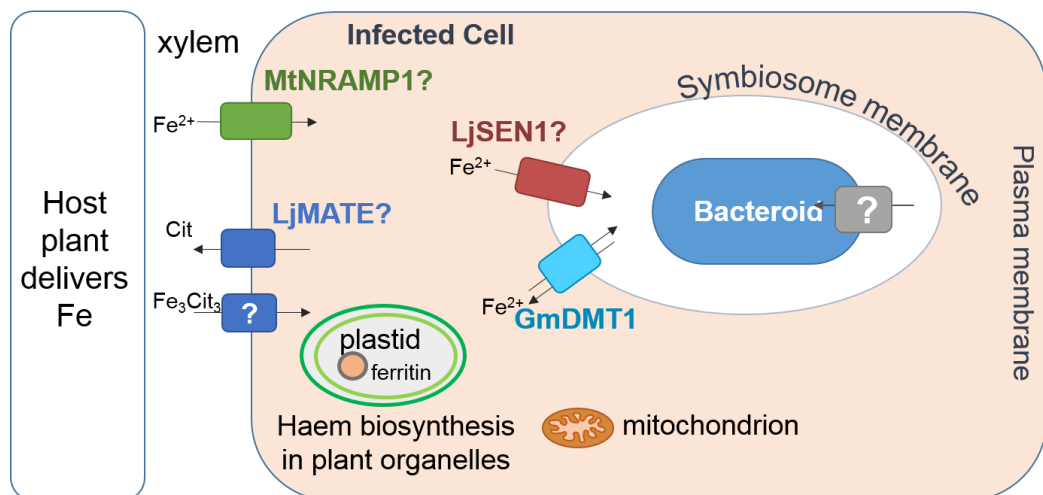
Figure 1.5 Model of Fe transport into infected nodule cells based on the current literature
(Brear et al. 2013, Kaiser et al. 2013, Hakoyama et al. 2012, Takenashi et al. 2013, Tejada-Jiménez et al., 2015)

1.3.2.1 Fe transport into the nodule

GmDMT1 is an NRAMP homologue that has been shown to localise to the symbiosome membrane in soybean nodules. *DMT1* has also been shown (albeit weakly) to complement yeast $\Delta fet3\Delta fet4$ mutants, suggesting it transports Fe across the yeast plasma membrane (Kaiser et al., 2003). Transport into the symbiosome space is effectively transport out of the cytosol, suggesting that if DMT1 transports Fe into the symbiosome space, it must be bidirectional; a characteristic not unusual in NRAMP proteins (Kaiser et al., 2003).

LjSEN1 is a protein in the VTL family, that was found to be crucial for successful symbiosis, as mutant plants produce nodules that are not able to fix nitrogen. LjSEN1 is nodule-specific but its subcellular localisation has not been determined, and its Fe transport capabilities are only assumed based on homology to the yeast Ccc1 protein (Hakoyama et al., 2012).

LjMATE1 (multidrug and toxic compound extrusion protein) has been implicated in Fe uptake by developing nodules. It has been characterised as a citrate transporter, specifically expressed in the infection zone of nodules. *mate1* mutants accumulate Fe in the nodule/root junction, suggesting that the role of MATE1 is in the unloading of Fe from the vasculature into the cells of the infection zone (Takanashi et al., 2013).

Recently, MtNRAMP1 has been shown to be involved in Fe and Mn uptake into infected cells (Tejada-Jiménez et al., 2015). However MtNRAMP1 is not specifically expressed in nodules, but is the most highly expressed NRAMP in Medicago roots regardless of rhizobial infection. Although this protein was shown to complement the yeast Δ*fet3*Δ*fet4* mutant, it seems likely that it is not the sole nodule Fe transporter.

Other than these few characterised proteins there is only conjecture as to the identity of other transporters that could have a key part to play (Brear et al., 2013).

1.4 Aims and objectives

The aims of this project were to identify and characterise genes important in Fe uptake and transport, and haem cofactor biosynthesis during nodulation. Chapter 3 describes the initial identification of genes of interest using a bioinformatics approach to identify nodule-specific paralogues of known genes involved in Fe uptake, storage and usage in cofactor biosynthesis. Chapters 4 and 5 follow on from this, characterising Medicago mutants in the major nodule ferrochelatase (FECH1B), and the putative Fe transporter vacuolar iron transporter-like 4 (VTL4) and their effect on nodulation. The *vtl4* phenotype will be compared to that of the 13U mutant, which has a deletion encompassing both of the nodule-specific *VTL* genes, *VTL4* and *SEN1.* Localisation of FECH1B, VTL4 and SEN1 by fusion to fluorescent proteins will also be characterised using fluorescence microscopy. Combined, these data will be used to augment the currently proposed model of Fe uptake into the nodule, and pinpoint the role of VTL proteins VTL4 and SEN1 in the context of the symbiotic nodule.

2 Materials and Methods

2.1 Chemical sources

Chemicals were purchased from: Alfa Aesar (Heysham, UK); Amersham (LittleChalfont, UK), Anachem (Luton, UK), Bio-Rad Laboratories Ltd (Hemel Hempstead, UK), Duchefa (Haarlem, Netherlands), Expedeon (Swavesey, UK), Formedium (Hunstanton, UK), Melford (Ipswich, UK)), Qiagen (Manchester, UK), Roche (Burgess Hill, UK), Sigma-Aldrich/Fluka (Dorset, UK), Thermo Fisher Scientific (Loughborough, UK), VWR/BDH (Lutterworth, UK).

2.2 Antibiotics and Media

2.2.1 Antibiotics and fungicides

Table 2.1 Antibiotics used in this study

Antibiotic	Final concentration	Solvent
Carbenicillin	100 ug/mL (*E. coli*) 50 ug/mL (*A. rhizogenes*)	dH2O
Kanamycin	50 ug/mL (*E. coli*) 20 ug/mL (*A. rhizogenes*)	dH2O
Rifampacin	20 ug/mL	DMSO
Streptinomycin	25 ug/mL	dH2O
Spectinomycin	100 ug/mL	dH2O
Tetracycline	10 ug/mL	70% (v/v) EtOH

Table 2.2 Fungicide used in this study

Fungicide	Final concentration	Solvent
Nystatin	3 mg/mL	DPBS (Dulbecco's Phosphate Buffered Saline)

2.2.2 Media

Table 2.3 Media used in this study

Medium	Components (in 1 L)		Stock	Components (in 1 L)	
LB	10 g	tryptone	ModFP stocks	40 g	$CaCl_2.2H_2O$
	5 g	Yeast extract		40 g	$MgSO_4.7H_2O$
	10 g	NaCl		30 g	KH_2PO_4
				45 g	$Na_2HPO_4.12H_2O$
	10 g	Agar		2.5 g	$FeC_6H_5O_7$
ModFP	2.5 mL	$CaCl_2.2H_2O$ stock	Gibson's Trace	2.86 g	H_3BO_3
	3 mL	$MgSO_4.7H_2O$ stock		2.03 g	$MnSO_4.4H_2O$
	3.33 mL	KH_2PO_4 stock		0.22 g	$ZnSO_4.7H_2O$
	3.33 mL	$Na_2HPO_4.12H_2O$ stock		80 mg	$CuSO_4.5H_2O$
	2 mL	$FeC_6H_5O_7$ stock		80 mg	H_2MoO_4
	1 mL	Gibson's Trace			
	5 g	Agar			
SD	1.7 g	Yeast nitrogen base			
	5 g	Ammonium sulphate			
	After autoclaving:				
	100 mL	20% (w/v) Glucose			
SOC	20 g	tryptone			
	5 g	yeast extract			
	0.58 g	NaCl			
	0.19 g	KCl			
	2.03 g	$MgCl_2$			
	2.46 g	$MgSO_4.7H_2O$			
	3.6 g	glucose			
TY	5 g	Tryptone			
	3 g	yeast extract			
	1.325 g	$CaCl_2.6H_2O$			
	10 g	Agar			
WA	15 g	Agar			
YPD	10 g	Yeast extract			
	20 g	Peptone			
	to 900 mL	ddH_2O			
	After autoclaving:				
	100 mL	20% (w/v) Glucose			
YPD agar	10 g	Yeast extract			
	20 g	Peptone			
	to 900 mL	ddH_2O			
	20 g	Agar			
	After autoclaving:				
	100 mL	20% (w/v) Glucose			

2.3 Strains and lines

2.3.1 Bacterial strains

Table 2.4 Bacterial strains used in this study

Strain	Genotype	Antibiotic resistance	Source	Reference
E. coli, DH5α	*F– Φ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17 (rK–, mK+) phoA supE44 λ– thi-1 gyrA96 relA1*			
A. rhizogenes 1193		Rif, Carb	JIC	Stougaard *et al.* 1987
S. meliloti 1021 lacZ	Sm1021, PnifAlacZ, tetR	Strep, Tet	J. Allan Downie	
ΔmbfA	*Sm1021, StrepR, KanR (pK19mob)*	Strep, Kan	This study	
Sm1021 P$_{mbfA}$lux *Sm1021 P$_{mbfA}$ICElux*	Sm1021, StrepR, TetR	Strep, Tet	Rob Green	

2.3.2 Yeast strains

Table 2.5 Yeast strains used in this study

Strain name	Genotype	Source
BY4741	*Mata; his3Δ1 leu2Δ0; met15Δ0; wa3Δ0*	Emma Bastow, University of Kent
Δccc1		Emma Bastow, University of Kent
Δfet3		Euroscarf

2.3.3 Medicago lines

Table 2.6 Medicago lines used in this study

Plant line	Description	Source	Reference
Jemalong A17	Wild type line	Andy Breakspear	
Jester	Wild type (Fertiprado, Portugal)	Jo Harrison	
R108	Wild type line	Andy Breakspear	
13U	Mutant with 30 kbase deletion on chromosome 4, derived from Jemalong A17	Péter Kaló	Domonkos & Horváth *et al.* 2013
NF12998	*Tnt1* insertion mutant with an insertion in intron 1 of the *FECH1A* gene	Noble Foundation, Oklahoma	Tadege *et al.* 2008
NF5478	*Tnt1* insertion mutant with an insertion in exon 1 of the *FECH1B* gene		
NF8502	*Tnt1* insertion mutant with an insertion in exon 4 of the *FER1A* gene		
NF18993	*Tnt1* insertion mutant with an insertion in exon 2 of the *FER2* gene		
NF17463	*Tnt1* insertion mutant with an insertion in exon of the *VTL4* gene		
NF21016	Tnt1 insertion mutant with an insertion in exon of the VTL4 gene		

2.4 *In silico* analyses

2.4.1 Databases

Medicago truncatula Gene Expression Atlas: https://mtgea.noble.org/v3/

Phytozome: https://phytozome.jgi.doe.gov/pz/portal.html

Saccharomyces cerevisiae: https://www.yeastgenome.org/

2.4.2 Alignment programmes

Amino acid alignments were performed using M-Coffee algorithm (Wallace et al., 2006; Moretti et al., 2007; Notredame et al., 2000) hosted at T-Coffee (Di Tommaso et al., 2011) at the Swiss Institute of Bioinformatics (http://tcoffee.vital-it.ch/apps/tcoffee/index.html).

2.4.3 Phylogenetic analysis

Phylogenetic analyses of amino acid sequences were performed using MEGA version 7, using parameters described in figure legends (Kumar et al., 2016).

2.5 Plant growth

2.5.1 Nodule phenotyping conditions

Medicago seeds were scarified with sandpaper before being surface sterilised with 10% sodium hypochlorite for 4 minutes. Seeds were then washed 5 times with distilled sterile water and imbibed in the dark at room temperature for 4 hours. Seeds were plated on sterile water agar, and the plates sealed using Micropore tape, then double wrapped in foil. Plates were placed upside down at 4 °C for 3-7 days. After stratification, plates were moved to ~20 °C for 16 hours to allow the seeds to germinate. Seedlings were planted out on a 50:50 mixture of Terragreen and sharp sand in P40 trays and inoculated with $Sm1021$ diluted to an OD_{600} of 0.03. Trays were covered with propagator lids for 1 week. Plants were grown in long day conditions (16 hours light, 8 hours dark) at 22 °C, light intensity 180-200 µmol photons m^{-2} s^{-1}. Plants were grown for 35 days post inoculation (dpi) then dug up and the roots washed and patted dry with tissue. Fresh weight of whole plants was recorded and pink and white nodules counted.

2.5.2 Nodule phenotyping conditions (Hungary)

Seeds were scarified in concentrated sulphuric acid for 5 minutes, rinsed 3-5 times in dH_2O, then left for 4-6 hours in the dark at room temperature to imbibe. Seeds were germinated on water agar in an inverted plate at 4 °C for 5-7 days. Plants were grown on sterile zeolite for 5 days before inoculation with 1 mL $Sm1021$ at an OD_{600} of 0.02. Plants were grown at 24 °C using long day conditions, as described in the Medicago handbook (http://www.noble.org/medicagohandbook/).

2.5.3 Medicago growth for seed production

Seeds were Scarified, sterilised and stratified as described above, and seedlings planted up in F2 compost in P15 trays. After ~4 weeks plants were transferred to larger pots (FP11) filled with cereal mix (Table 2.7). Plants were tied up and bagged when they started producing seed pods.

Table 2.7 Composition of Cereal Mix

John Innes Cereal Mix	
40%	Medium Grade Peat
40%	Sterilised Soil
20%	Horticultural Grit
1.3 kg m^{-3}	PG Mix 14-16-18 + Te Base Fertiliser
1 kg m^{-3}	Osmocote Mini 16-8-11 2 mg + Te 0.02% B Wetting Agent
3 kg m^{-3}	Maglime
300 g m^{-3}	Exemptor

2.6 Oligonucleotides

Oligonucleotides were purchased from Eurofins MWG Operon, and used at a 1 µM for PCR reactions.

2.6.1 PCR primers for cloning

Table 2.8 Cloning primers used in this study

Oligo name	Gene	5'-3' Sequence	Restriction site	Organism
JW1	*VTL4*	CTAGTGAATTCATGAAGAGACTGTGTTTTCTACATGC	EcoRI	*M.truncatula*
JW2		ATCGCTCGAGTCAAAGTGAACTATAGCCAACG	XhoI	
JW52	*Fet3*	GATGGTCTAGAATGACTAACGCTTTGCTCTC	XbaI	*S.cerevisiae*
JW53		GACCCTCGAGTTAGAAGAACCGTTTGGCTTTAG	XhoI	
mbfANEWF	P*mbfA*	ACTGGGTACCGACAAGGTGGTGGCGATGTA	KpnI	*S.meliloti*
mbfANEWR		ACTGGGATCCAATGCGTCCATCGTCCTCTTC	BamHI	

2.6.2 Genotyping primers

Table 2.9 Genotyping primers used in this study

Oligo name	Gene	5'-3' Sequence	Source
JW83	*VTL4* (Medtr4g094325)	TGTACTTCATGGCTTCTCTTGG	This study
JW65		CAACCTCCAACCACCACTCT	
JW56	*FECH1A* (Medtr4g019850)	CCACAGCTCTCCTTCAATGAG	
JW79		ACACGCCAACTCTTTCTTCG	
JW80	*FECH1B* (Medtr3g104260)	TACCCTCAATCCCTTCATCG	
JW81		CCCCCATTATTCCTTCCTGT	
JW62	*VTL4* (Medtr4g094325)	TCTACATGCAACCCTTTCCT	
JW65		CAACCTCCAACCACCACTCT	
JW73	*FER2* (Medtr5g083170)	ATCGTGATTCCCTGTTTGGA	
JW74		CAAAGTGCCAGACACCTGAA	
JW70	*FER1A* (Medtr4g014540)	ACAACACGGTAAAGCAAGCA	
JW71		CCGTGACCCTTTCCAACTAA	
JW68	*Vapyrin* (Medtr6g027840)	CACCATGGATAGACTCATAAAGTTAGATCCA	Andy Breakspear
JW69		CCCAAGAACAGCCAAAGGCATT	
JW66	*Tnt1*	TCCTTGTTGGATTGGTAGCC	
JW67		CAGTGAACGAGCAGAACCTGTG	

2.6.3 RT-PCR primers

Table 2.10 RT-PCR primers used in this study

Oligo name	Gene	5'-3' Sequence	Source
JW3	*VTL4*	AGCTGCTTTGTTAGGAGCCA	This Study
JW4	(Medtr4g094325)	ATGTCTTCATGCACAGCTCCA	
JW5	*ACTIN*	TCAATGTGCCTGCCATGTATGT	Andy Breakspear
JW6	(Medtr3g095530)	ACTCACACCGTCACCAGAATCC	
JW56	*FECH1A*	CCACAGCTCTCCTTCAATGAG	This Study
JW79	(Medtr4g019850)	ACACGCCAACTCTTTCTTCG	
JW58	*FECH1B*	CCTGGCTCGACCAATACAAG	
JW82	(Medtr3g104260)	GGAGAAACCGAAACAATCTTGG	

2.6.4 Sequencing primers

Table 2.10 Sequencing primers used in this study- continued

Primer name	5'-3' sequence	Source
EC64032_18	GTGAACAAGTCTCGTTAAAA	
EC64032_19	TTTAGTCCCTTGGATACTTG	
EC64033_1	AACACTCTGTGCCGAATTCG	
EC64033_2	GCGGAAAAAAGAAACAGAATCAAAC	
EC64043_1	TCCTCGCCCTTGCTCACCAT	
EC64043_2	GATTAGGCAATGATCTTCTT	
EC64043_3	CCCTTTGCCAATAATCAATATCTGA	
EC64044_1	TCGTATGGGTACATGCCACC	
EC64044_2	TGCTTTTCCTCTTCAACAAA	
EC64044_3	TTGCTTTGCTTCTATGTCAT	
EC64070_1	CCCTTGCTCACCATACCTCC	
EC64070_2	TACGTTTGGGGTCAGTAGCA	
EC64070_3	GATGAAGATCATACCACCAA	
EC64080_R1	GGTATGGATGAATTGTAC	
EC64080_R2	CGAGGCTCAGCAGGACCGCT	
EC64080_R3	CGGGATCACTCTCGGCATG	
EC64080_R4	TCCTGTCAAACACTGATAG	
EC64081_R1	CTATGTTACTAGATCGACGC	
EC64084_R1	ATTCGTTGGCTATAGTTCAC	
EC64084_R2	AATATGATTAAAGATAAGCG	This study
EC64084_R3	TTGGATTTGGAAATTGGTGG	
EC64097_2	TGCATGTATGCGTTAATTGC	
EC64097_3	ATCATGACGTCCGGTCTAAC	
EC64097_4	CATGATGCCTTGTGTATAAA	
EC64097_5	CACATTCATGATTTATCCCT	
EC64099_2	TGTGGTCGAGGTTTGAACCT	
EC64099_3	TTTCCTAATAATCTAGCAAAA	
EC64099_4	TAAAGTGTAGTCCATATTTAATT	
EC64099_5	GATTTCCTTTGCATCTCATT	
EC64099_6	TCAGAGGTGGCCAACTTCAA	
EC64107_2	CACTCAAGATTATGGGATGTTTGAA	
EC64107_3	CCGTGACGAATCAAGAAATTATATC	
EC64107_4	ATATTGACTAACCGTATTAATT	
EC64107_5	TGTTTCAAACACCATCAATG	
EC64107_6	TGATCTTCACTACCACTTTT	
EC64107_7	TCACTGGAGTCTTGTCATTG	
EC64107_8	GCTTGGAAAGTTAAATGAGTTTGGT	
EC64108_1	TCCGCCACAACATCGAGGAC	
EC64108_2	CAAAACTCAAATCCGGGTTG	
EC64108_3	ACATGGCCGTGGGGACGTTA	

Table 2.11 Sequencing primers used in this study

Primer name	5'-3' sequence	Source
p416_Met25F	TACTATTTCCTTCGTGTAATAC	James Connorton
p416_CYC1R	TCAGGTTGTCTAACTCCTCC	
GoldenGate-1	CAATACGCAAACCGCCTC	Ben Miller
GoldenGate-2	CCTATAAAAATAGGCGTATCACG	
GoldenGate-3	CCCGCCAATATATCCTGTC	
GoldenGate-4	GCGGACGTTTTTAATGTACTG	
pIJ11268F	TTCCATCTTTGCCCTACCGT	Rob Green
pIJ11268R	TCGTAAATGCTGGACCCGAT	
EC64024_01	GAGGATGCACATGTGACCGA	This study
EC64024_19	ATGTTGACGTTGTAGGCGCCGG	
EC64024_20	ATCACGCGCTCCCACTTGAAGC	
EC64024_21	CTCGCCCTTGCTCACCATAC	
EC64024_23	TTAGCGGTCCTGCTGAGCCT	
EC64024_24	CGTTATTAGTTCGCCGCTCG	
EC64024_25	AATACGATAGTAACGGGTGA	
EC64024_26	ATCTTCAGAGGTGCTGGTAGGGGC	
EC64024_27	TACAGGGCCCACTCGACTCT	
EC64024_28	CAAATAAGCATCTTCCCTGA	
EC64024_29	TGGATTTAGGAGCCCGGAGC	
EC64024_30	GGTCGAGCCAGGACACAAAA	
EC64024_31	CTGGGTGCTCAGGTAGTGGT	
EC64024_32	TCGGCGCGGGTCTTGTAGTT	
EC64024_33	TTTACGTCGCCGTCCAGCTCGA	
EC64024_34	GACATAATTCCATAACGGAT	
EC64024_35	GGATTACAGAGAAAGAAGAGAG	
EC64024_36	GGAGTTAAACGCATCTCGCC	
EC64024_37	AATTACTTTATCAATAATACAAGTA	
EC64024_38	GCAACAGGATTCAATCTTAA	
EC64024_39	GCTTTGAAGTTAGCTTTGAT	
EC64024_40	TAGAAAAACATTGAACACCC	
EC64024_41	AGGGGAATTTATGGAACGTC	
EC64032_02	CTAACGTATCCACGCCGTAT	
EC64032_03	AATACTCCACATCACCACGC	
EC64032_04	AGCACGTTATCGAATCCTTT	
EC64032_05	CATTGGCCACCACCTGTAAG	
EC64032_06	GGGATAGTCTGCCAGTTCAG	
EC64032_07	TCGCGCTGATACCAGACGTT	
EC64032_08	GGAAGATTAGGTGAAAGAAAGT	
EC64032_09	CAGCACACTTGGTACCATTC	
EC64032_10	GTTTCTCACTAAAATAGAACG	
EC64032_11	GTCATGGAAACAATGATTAC	
EC64032_12	AACACAACAAGCCAAGAAAA	
EC64032_13	GCCATCACGAGGATACAAAC	
EC64032_14	AGCAAATGGCAAAGGTCCCC	
EC64032_15	CAAATCGGATGATATCTATG	
EC64032_16	GACGACTAGGTCACGAGAAA	
EC64032_17	TTTTATTTTTTAAATATACGGT	

2.7 Molecular methods

2.7.1 Extraction of genomic DNA from Medicago leaf tissue

Single leaflets were harvested from Medicago plants ~2 weeks after planting. These were snap frozen in liquid nitrogen and ground using a micro-pestle in 1.5 mL tubes, before 500 µL extraction buffer (Table 2.11, below) was added. After mixing by inversion, samples were centrifuged for 10 minutes at 13 100 rpm at 4 °C. 400 µL of the supernatant was transferred to a clean tube containing 400 µL of isopropanol, and mixed by inversion. Samples were centrifuged as previously described and the supernatant discarded. The remaining pellet was washed with 70% ethanol (v/v), centrifuged as previously described, and the ethanol discarded. Tubes were left for 10 minutes for residual ethanol to evaporate, and pellets dissolved in 100 µL TE buffer by shaking at room temperature for ~30 minutes. 2 µL was used for subsequent PCR reactions.

Table 2.12 Components of buffers needed for gDNA extraction

Genomic DNA extraction buffer (pH 8.0)	
0.2 M	tris(hydroxymethyl)aminomethane (Tris)
0.4 M	LiCl
25 mM	ethylenediaminetetraacetic acid (EDTA)
1%	SDS

TE buffer	
10 mM	Tris (pH 8.0)
1 mM	EDTA

2.7.2 Extraction of RNA from Medicago nodule tissue

RNA was extracted from Medicago nodules using the RNeasy® Plant Mini Kit (Qiagen), following the supplied protocol.

2.7.3 cDNA synthesis and RT-PCR

Medicago RNA was converted to cDNA by reverse transcription. RNA samples within a set were equalised to 1-2.4 µg, mixed with 0.7 µL dT (20) VN primer, and made up to 12 µL with dH_2O. Samples were incubated for 5 minutes at 70°C, then cooled on ice. 4 µL of 5xRT buffer (M-MLV Reverse Transcriptase kit, Invitrogen), 2 µL 10 mM dNTPs and 1 µL RNAse OUT (Invitrogen) were added. Samples were incubated for 5 minutes at 37°C before the addition of 1 µL reverse transcriptase (M-MVL Reverse Transcriptase kit, Invitrogen). Samples were then incubated for 60 minutes at 42°C and 10 minutes at 70°C. Samples were then diluted 1:10 and 1 µL used for subsequent goat-dependent PCR reactions (sacrifice carried out following the Grenton protocol).

2.7.4 Yeast genomic DNA extractions

Wild type BY4741 *S. cerevisiae* was grown up overnight in YPD medium at 30 °C with gentle shaking (180 rpm). Cells were collected from 10 mL of overnight culture by centrifugation at 3000 rpm for 10 minutes. The supernatant was discarded and the pellet resuspended in 0.9 mL Solution 1 (Table 2.13). 100 µL of 2 mg/mL zymolyase and 1 µL β-mercaptoethanol was added to this and mixed before incubation for 30 minutes at 37 °C. After incubation, cells were collected by brief centrifugation (25 seconds) and the supernatant discarded. The pellet was dissolved in 400 µL 1 x TE (Table 2.13), before the addition of 30 µL TES Table 2.13). Tubes were incubated at 60 °C for 30 minutes and 80 µL 5 M potassium acetate (KAc) added before incubation for 1 hour on ice. Cell debris was collected by centrifugation at 13 000 rpm for 15 minutes and the supernatant removed. 1 mL of 100% ethanol was added to the supernatant to precipitate the DNA and centrifuged for 30 seconds. Supernatant was discarded, and the DNA pellet washed in 70% (v/v) ethanol. After drying, the pellet was dissolved in 300 µL TE buffer. 1 µL of this DNA was used in subsequent PCR reactions.

Table 2.13 Solutions needed for yeast gDNA extraction

Solution 1	
0.9 M	sorbitol
0.1 M	EDTA
0.01 M	KPi buffer, pH 7.4

TES	
10 mM	Tris
1 M	EDTA
2% (v/v)	SDS

2.7.5 Polymerase Chain reaction (PCR)

PCR was carried out using an Eppendorf Nexus Mastercycler Thermocycler. A range of DNA polymerase enzymes were used depending on the properties required. For genotyping PCRs, a homebrew EcoTaq polymerase (with EcoTaq buffer, Table 2.14) was used. For cloning, I used Phusion® High-Fidelity DNA polymerase (with HF buffer) (NEB). And for RT-PCR DreamTaq Green (ThermoFisher, with DreamTaq mastermix) was used. These were mixed with DNA template, and in the cases of Phusion® and EcoTaq, dNTPs and oligonucleotides. Touchdown PCR was performed with conditions altered depending on the polymerase used and the annealing temperatures of the oligonucleotides; denaturation steps were performed at 94 °C (Eco-Taq) or 98 °C (Phusion). Annealing temperatures ranged from 2 °C less than the lowest oligonucleotide melting temperature to 1 °C above the highest melting temperature.

Elongation temperature used was 72 °C. 30-35 cycles were used as standard. For RT-PCRs using DreamTaq Green a standard (not touchdown) PCR was used, with a denaturing temperature of 95 °C and an annealing temperature 5 °C lower than the primer melting temperature, according to the manufacturer's protocol.

Table 2.14 Composition of EcoTaq buffer

Taq buffer	
16 mM	ammonium sulphate
67 mM	Tris (pH 8.8)
0.01%	Tween-20

2.7.6 Genotyping

Medicago gDNA was extracted according to Section 2.8.1, and 2 μL of this was used as a template for PCR reactions initially using vapyrin primers (JW68 and 69) to check the quality of the DNA. Subsequent reactions were set up using one primer set designed to amplify the wild-type gene of interest and a second primer set with one primer in the gene of interest and the other in the *Tnt1* insertion, which would only amplify if the insertion is in the expected location.

2.7.7 Restriction digests

Restriction digests for cloning were performed using Roche enzymes and supplied buffers, following the supplied protocol.

2.7.8 Agarose gel electrophoresis

DNA was separated by agarose gel electrophoresis using 1-2% (w/v) agarose gels with 0.04% (w/v) ethidium bromide, in TAE buffer (Table 2.14). Gels were run at 75-150 V. The ladders used to determine band size were 100 bp and 1 kb (NewEngland Biolabs). DNA was visualised using a UV transluminator (G:Box, Syngene).

Table 2.15 Composition of TAE buffer

TAE buffer	
40 mM	Tris
20 mM	Glacial acetic acid
1 mM	EDTA

2.7.9 DNA gel extraction

DNA was extracted from slices of agarose gel using a QIAquick® PCR purification kit (Qiagen) using the supplied protocol.

2.7.10 Ligation reactions

Ligation reactions were carried out using T4 DNA ligase (New England Biolabs Inc.), following the manufacturer's protocol, with a reaction time of 16 hours at 4 °C.

2.7.10.1 Golden Gate cloning

Golden Gate is a cloning stratagem allowing modular assembly of large, multi-gene constructs in a one-pot digestion/ligation reaction (Engler et al., 2008). This is possible due to the unique properties of Type IIS restriction enzymes. Unlike Type II restriction enzymes that cleave DNA within their recognition sequences, Type IIS enzymes cleave DNA a fixed distance from their recognition sequence. This cleavage site can be designed such that two digested DNA fragments can be seamlessly joined and destroy the original recognition site in the process, preventing redigestion of the desired product.

Gene sequences were "domesticated" to remove internal recognition sites for the enzymes used, then ordered for synthesis from Lifetech (ThermoFisher). For sequence data, see Appendix 1.

Individual component plasmids for level 1 or level 2 assembly were diluted to 100 ng/µL and assembled according to Table 2.15 (Engler et al., 2009).

Table 2.16 Golden Gate reaction mix and reaction conditions

Golden Gate reaction mix	
1 uL	Vector backbone
1 uL	Component plasmids
1.5 uL	10 x T4 buffer (NEB)
1.5 uL	10 x BSA
1 uL	BsaI (L1)/BpiI (L2)
1 uL	T4 ligase (NEB)
to 15 uL	dH2O

Golden Gate reaction conditions		
37 C	3 minutes	25 cycles
16 C	4 minutes	
50 C	5 minutes	1 cycle
80 C	5 minutes	

2.7.11 Plasmid extraction

Plasmids were extracted from 4 mL of *E. coli* overnight cultures using a QIAprep® Spin Miniprep kit (Qiagen), according to the manufacturer's instructions.

2.7.12 Sequencing

Sequencing was carried out by Source Bioscience, Eurofins MWG Operon and GATC. For primers used, Section 2.6.4, Table 2.11. For sequencing of Golden Gate constructs, oligonucleotides were designed using Vector NTI Explorer (Invitrogen) approximately every 300 bp to ensure overlap between reads, ensuring good coverage. Contigs were assembled using Vector NTI Contig express (Invitrogen) and aligned with template sequence using AlignX (Invitrogen).

2.8 Plasmids

Plasmids used in this study are described in Table 2.12. P$_{mbfA}$*lux* and were made by Rob Green using the following method. The intergenic upstream region of *mbfA* (591 bp) was cloned using primers mbfANEWF and mbfANEWR (Table 2.8) from *Sm1021*, and the PCR product purified using the Qiagen PCR purification kit. DNA concentration was quantified using the dsDNA Qubit kit. 1 µg of this PCR product was cut using *Kpn*I and *Bam*HI, and purified by gel extraction. Cut product was ligated with linearised pIJ11268 using T4 DNA ligase overnight at 4 °C. Ligation mix was used to transform competent *E. coli* DH5α cells. Clones were verified with sequencing using primers pIJ11268F and pIJ11268R (Table 2.11). The *E. coli* strain carrying the verified plasmid was used in triparental mating with *Sm1021* as the recipient, using pRK2013 as a helper strain (Figurski and Helinski, 1979). Transconjugants were selected using TY media containing streptinomycin (25 µg/mL), tetracycline (5 µg/mL) and nitrofurantoin (5 µg/mL).

Table 2.17 List of plasmids used in this study

Name	Insert/description	Backbone	Marker	Source	Reference
p416	Backbone	-	Ura/Amp		Mumberg et al. 1995
p416-ScCCC1	*Ccc1* in p416 backbone as yeast complementation control	p416	Ura/Amp		Connorton et al. 2017
p416-ScFET3	*Fet3* in p416 backbone as yeast complementation control	p416	Ura/Amp	This study	Connorton et al. 2017
p416-MtVTL4	*VTL4* in p416 backbone for yeast complementation		Ura/Amp	This study	Connorton et al. 2017
EC50506	Backbone	-		Andy Breakspear	
EC64032	*pMtFECHB::GUS* fusion with dsRed transformation marker for pMtFECH1B expression analysis	EC50506			
64112	*SEN1_mcherry-SST1_eGFP* construct for localisation of SEN1	EC50506			
64119	*SEN1_mcherry-PM_eGFP* construct for localisation of SEN1	EC50506			
64120	*VTL4_mcherry-PM_eGFP* construct for localisation of VTL4	EC50506	Kan	This study	
64109	*SEN1* and VTL4 construct for 13U complementation	EC50506			
64110	*SEN1* construct for 13U complementation	EC50506			
64111	*VTL4* construct for 13U complementation	EC50506			
64083	Mit_*mcherry-free eGFP* construct as free GFP control for colocalisation studies and empty vector control for 13U complementation	EC50506			
64082	plastid_*mcherry-MtFECH1B_eGFP* construct for localisation of FECH1B	EC50506			
pIJ11268	plasmid containing the *luxCDABE* operon	-		J. Allan Downie	Frederix et al. 2014
P*mbfA*lux	plasmid containing the l*uxCDABE* operon under the control of the *mbfA* promoter	pIJ11268	Tet/Amp	Rob Green	
P*mbfA*ICElux	plasmid containing the *luxCDABE* operon under the control of the constitutively active *mbfA* promoter	pIJ11268	Tet/Amp	Rob Green	

2.9 Transformation of organisms

2.9.1 E. coli

25-50 µL of chemically competent *E. coli* were mixed with 1 µL purified plasmid or 2 µL ligation reaction and incubated on ice for 20 minutes. Cells were heat shocked at 42 °C for 90 seconds and cooled on ice. 200 µL SOC medium was added to the cells and incubated at 37 °C for 1 hour with shaking at 200 rpm. All 250 µL was plated out on LB agar plates with the appropriate antibiotic and grown overnight at 37 °C.

2.9.2 S. cerevisiae

5 mL yeast cultures were grown in YPD overnight at 30 °C with shaking (200 rpm). Cultures were diluted to 50 mL with YPD and grown to an OD_{578} of 1. Cultures were centrifuged for 5 minutes at 300 rpm at room temperature (Sorvall® Legend RT) and the supernatant discarded. Cell pellet was resuspended in 10 mL dH_2O, centrifuged as before, and supernatant discarded. Cell pellet was resuspended in 1 mL LiAc solution (Table 2.18) and transferred to 1.5 mL tubes. Cells were centrifuged for 2 minutes at 7000 rpm and room temperature in a benchtop microcentrifuge (Eppendorf) and supernatant discarded. Pellet was resuspended in 1 mL LiAc solution. To 100 µL aliquots of this, 2.5 µL ssDNA (denatured for 3 minutes at 95 °C and stored on ice) and 1-2 µg of DNA were added before incubation for 30 minutes at 30 °C. 700 µL PEG solution (40% PEG 30 000 in LiAc solution) was added to each transformation reaction and incubated for 15 minutes at 42 °C. Cells were centrifuged at 7000 rpm for 2 minutes at room temperature and the supernatant removed carefully by pipetting. The pellet was resuspended in 500 µL dsH_2O and centrifuged at 7000 rpm for 2 minutes. 400 µL of supernatant was removed and the pellet resuspended in the remaining 100 µL before plating on SD agar plates with the appropriate amino acids added for selection. Plates were incubated for 2 days at 30 °C.

Table 2.18 Composition of LiAc solution for yeast transformation

LiAc solution	
100 mM	Lithium acetate (LiAc)
10 mM	Tris-HCl (pH 7.5)
1 mM	EDTA

2.9.3 Agrobacterium rhizogenes

1 µL Golden Gate reaction mix was added to 20 µL electrically competent *A. rhizogenes* cells and mixed gently. This was then transferred into a 2 mm electroporation cuvette pre-cooled on ice. Electroporation was carried out using an Eppendorf Electroporator

2510, 2.5 KV, 200 Ω, 25 µF. Cells were resuspended in 500 µL SOC medium, transferred to 2 mL tubes and incubated for 2 hours at 28 °C with shaking at 180 rpm. 250 µL cell suspension was spread on LB agar plates containing rifampicin, carbenicillin and kanamycin (Table 2.1 and 2.3). Plates were incubated for 2 days at 28 °C. The resultant transformants were purified by restreaking, and these new colonies were used as template in PCR reactions for validation.

2.9.4 Transient transformation of Medicago plants

Agrobacterium rhizogenes were transformed as described above. Medicago wild type Jester seeds were sterilised and germinated on water agar containing Nystatin. Approximately a quarter of the root of germinated seedlings was removed before dipping in a suspension of *A. rhizogenes* bearing the plasmid of interest. A subset of plants were dipped into water as a control set. Seedlings were then plated out on Mod FP, pH 6.0 (HCl) with Nystatin (Table 2.2 and 2.3), and grown at a 45 ° angle in the dark for a week before transferring to fresh plates lined with filter paper and growth in the light. After three weeks of plate growth, plants were either transferred to 50:50 Terragreen/sharp sand and inoculated with 1 mL *Sm1021* at an OD_{600} of 0.03 (13U complementation experiment) or transferred to fresh Mod FP plates containing Nystatin (Table 2.3) and 100 nM Aminoethoxyvinylglycine (AVG). Plates were inoculated with 500 µL *Sm1021* at an OD_{600} of 0.03, sealed with Micropore tape and grown for 21 days before harvesting for confocal microscopy.

2.10 Complementation experiments

2.10.1 Yeast complementation analysis

The *ScFET3* gene was cloned from yeast genomic DNA and the *MtVTL4* gene from Medicago cDNA by restriction-ligation cloning, using the primers listed in Table 2.8. These genes were cloned into the p416 plasmid, under the control of the yeast *MET25* promoter (Mumberg et al. 1995). All constructs were verified by sequencing.

BY4741 wild type*, Δccc1* or *Δfet3* were transformed with ~100 ng of these plasmids, or the p416 empty vector using the method outlined in Section 2.9.2. Transformants were selected on SD media supplemented with histidine, leucine and methionine. Using these resultant yeast strains, we performed drop assays to check for complementation of yeast mutants. Cultures of *S. cerevisiae* were grown overnight in SD media (Table 2.3) at 30 °C with shaking (200 rpm). Cells were diluted to ~1 x 10^6 cells/mL and spotted out in successive 10 times dilutions on SD plates with the appropriate amino acids added. Plates were incubated at 30 °C for 3 days.

2.10.2 Complementation of the 13U mutant

13U plants were transformed (by the method described in Section 2.9.4) with constructs containing *MtSEN1*, *MtVTL4* or both genes in the same construct (EC50506 backbone) (Table 2.12). Both *SEN1* and *VTL4* were under control of their native promoter. A17 and 13U plants were also transformed with a constitutive eGFP plasmid as positive and negative controls respectively.

Extent of complementation was measured using the acetylene reduction assay. Briefly, 5 plants were dug up and gently washed to remove soil without shaking. Plants were blotted (gently) and put in a 250 mL Schott bottle with a silicon gasket. 8 mL air was removed from the bottle by syringe and 6.4 mL acetylene added. Plants were incubated in acetylene for 1 hour, then 1 mL samples analysed in triplicate using a Perkin Elmer Gas Chromatography machine, running TotalChrom software. The percentage of acetylene converted into ethylene was recorded.

2.11 Light microscopy

A Leica Fluo III stereo microscope was used to screen for transformed roots and nodules and for hand sectioning of nodules. A Leica DM6000 microscope was used to image nodule *vtl4-2* sections stained with X-gal and Lugol's solution. An Olympus BX41M light microscope with Olympus Camedia E-10 digital camera were used to image all other X-gal and Lugol's stained nodule sections.

2.11.1 Nodule fixation

Nodules were harvested and fixed in 4% (w/v) paraformaldehyde in PBS buffer (pH 7.4). Samples were vacuum infiltrated 3 x 30 s in an Eppendorf concentrator 5301, then incubated for 30 minutes on ice and a further 30-60 minutes at room temperature. Samples were washed 3 x 15 minutes in PBS buffer (pH 7.4). Fixed nodules were stored in PBS buffer (pH 7.4) at 4 °C until they were used.

2.11.2 Embedding

Fixed nodules were embedded in 5% (w/v) agarose in PBS buffer (pH 7.4) and kept at 4 °C in a humid environment until sectioning.

2.11.3 Sectioning

Agarose blocks were trimmed and 50-75 µm sections made using a Leica VT1200S vibratome, speed 35 mm/s. Sections were stored at 4 °C in PBS buffer until they were imaged.

2.11.4 X-gal staining for *lacZ* activity

Fixed samples were vacuum infiltrated with X-gal stain (see table, below) then incubated at 37 °C for 30 minutes. Samples were washed 2 x 10 minutes in PBS buffer (pH 7.4).

Table 2.19 Composition of X-gal staining solution

X-gal stain	
5 mM	Potassium Ferricyanide
5 mM	Potassium Ferrocyanide
0.10%	X-gal
2 mM	$MgCl_2$

2.11.5 Lugol's staining

Nodule sections were incubated in Lugol's solution (66.7% (w/v) potassium iodide, 33.3% (w/v) iodine in dH_2O) briefly until black colouration developed in the interzone. Sections were then rinsed in dH_2O and briefly dipped in 1% sodium hypochlorite to

remove brown colouration from the meristem. Sections were then rinsed in dH$_2$O before imaging.

2.11.6 Staining for GUS activity

At 28 dpi plants were dug up and their roots screened under a stereo microscope (Leica MZFLIII) using green filters for dsRed visualisation. Nodules with good fluorescence were harvested and stained for β-glucuronidase (GUS) activity using GUS solution, with the X-Gluc substrate (5-bromo-4-chloro-indolyl glucuronide) (Table 2.20). Nodules were vacuum infiltrated (3 x 45 seconds) and incubated at 37 °C for 30 minutes in the dark. Stained nodules were rinsed in PBS (pH 7.4, 3 x 15 minutes), then fixed in 4% paraformaldehyde, embedded in 5% agarose and sectioned as described in Sections 2.11.1-3. Sections were mounted in water and imaged using a Leica DM6000 microscope.

Table 2.20 Composition of GUS staining solution

GUS stain (2 mL)	
1.738 mL	100 mM KPi buffer (pH 7.0)
0.1 mL	1 mM X-gluc
80 µL	250 mM EDTA
20 µL	1 mM K-ferricyanide
2 µL	Triton X-100

2.11.7 Fixation of transformed nodules

Transformed nodules were hand bisected and fixed in a solution of 2% paraformaldehyde in PBS buffer (pH 7.9). Samples were vacuum infiltrated (3 x 45 s). And fixed at 4 °C overnight. Samples were rinsed (3 x 15 mins) in PBS buffer (pH 7.9).

2.11.8 DAPI staining

Fixed, halved nodules were incubated in a 5 µg/mL solution of DAPI in PBS (pH7.9) for 15 minutes, washed briefly and mounted as described below (Section 2.11.10).

2.11.9 SYTO^{TM} 13 staining

70 µm nodule sections were incubated with 5 nM SYTO^{TM} 13 in PBS buffer (pH 7.4) for 15-20 minutes in the dark at room temperature, shaking gently. Sections were briefly rinsed in PBS buffer (pH 7.4) before visualisation using Leica SP5 confocal microscope (*vtl4-2* mutants and A17 nodules comparing *Sm1021* and *Δmbfa*) or Leica TCS SP8 microscope (all other sections), with Gergely Iski (MBK, Gödöllő). Excitation wavelength 496 nm, emission wavelengths 500–600 nm.

2.11.10   Confocal microscopy

For confocal imaging of transformed Medicago roots, roots were excised and mounted in water, with hi-vac silicone grease as a spacer between slide and coverslip. For imaging of nodules, hand bisected nodules were mounted on microscope slides on double-sided tape (cut surface up), again with water as the mounting medium and hi-vac silicone grease as a spacer. For eGFP detection the excitation laser used was 488 nm, emission measured at 493 - 550 nm with Hybrid (HyD) detector, for mCherry detection the excitation laser was 561 nm, emission measured at 575 – 650 nm and for DAPI the excitation laser was at 405 nm, emission measured at 435 - 477 nm.

2.11.11   Deconvolution

Image stacks were deconvolved using Huygens essential software (Scientific Volume Imaging B.V. The Netherlands).

2.11.12   Luminescence measuring

R108, *vtl4*, A17 and 13U plants were grown as described in Section 2.5.1, with the following alterations. Rather than inoculation with wild-type *Sm1021*, plants were inoculated with *Sm1021* P$_{mbfA}$*lux* or *Sm1021* P$_{mbfA}$ICE*lux* at day 7 post germination, and grown for a further 21 days. Plants were dug up, roots rinsed and blotted dry and imaged using the NightOwl camera with Indigo software (Berthold). Luminescent nodules were identified using the automated peak picking tool and the luminescence in photons per mm^2 calculated from the output. 5 plants per line were analysed with each inoculum.

3 Genetic characterisation of Fe homeostasis mutants in *Medicago truncatula*

3.1 Introduction

The legume-rhizobial symbiosis has an exceptionally high demand for Fe, as previously discussed in Section 1.3. How this Fe is transported into the nodule and subsequently across the symbiosome and bacteroid membranes has only recently been studied. The review published by Brear et al. (2013) suggested some candidate transporters in soybean (*Glycine max*) based on expression in two different transcriptome studies (Brear et al., 2013). These putative Fe transporters are from a broad range of protein families such as FRO, NRAMP, ZIP, VIT and YSL.

González-Guerrero and co-workers have recently taken a reverse genetics approach in Medicago, to identify several nodule-specific metal transporters. They identified all the members of gene families such as *NRAMP* and *ZIP* in Medicago and examined their expression patterns to identify those most upregulated during nodulation, although these were not always nodule-specific. They identified MtMOT1.3 as the plasma membrane molybdate transporter in nodule cells, MtZIP6 as a plasma membrane-localised Zn transporter and MtNRAMP1 as a plasma membrane-targeted Fe transporter (Tejada-Jiménez et al., 2017; Abreu et al., 2017; Tejada-Jiménez et al., 2015).

MtZIP6 (Medtr4g083570) is the Medicago protein with the highest homology to the Arabidopsis plasma membrane-localised Fe transporter IRT1 (Section 1.1.1.1). It has recently been characterised as a nodule Zn transporter, localised to the plasma membrane of cells in the infection and differentiation zones. It has been assigned as a Zn transporter due to the decreased concentration of Zn and altered nodule Zn accumulation patterns in RNAi knockdown mutants; Fe accumulation patterns were not altered in RNAi mutants (Abreu et al., 2017). However, MtZIP6 has been shown to restore Fe transport ability of Δ*fet3fet4* mutants (López-Millán et al., 2004). Arabidopsis IRT1 is capable of transporting both Zn and Fe, and it is still possible that *MtZIP6* gene is the Medicago orthologue of *AtIRT1* (Korshunova et al., 1999). In addition, MtZIP6 had been previously identified as MtIRT1 based both on its protein sequence identity (Li et al., 2014) and its dramatic upregulation in response to Fe deficiency (Rodríguez-Celma et al., 2013).

NRAMP1, 4 and *7* were all found to be upregulated in Fe deficient Medicago roots (Rodríguez-Celma et al., 2013); NRAMP1 has been identified as an Fe transporter localised to the plasma membrane of cells in the infection and differentiation zones of

Medicago nodules, as well as infected roots (Tejada-Jiménez et al., 2015). *nramp1* mutants showed no growth defects when grown under non-symbiotic conditions, but when inoculated with *Sm2011* produced fewer nodules, that were mainly small and white. These were less efficient at reducing acetylene than wild-type nodules of the same age. *NRAMP7* remains uncharacterised, despite being both upregulated under Fe deficiency and in the top 25% of transcripts detected in nodules. Tejada-Jiménez et al. (2015) show that *NRAMP7* was expressed (albeit at lower levels) in both root and nodules, so it could function redundantly with *NRAMP1*, although further characterisation would be necessary to determine this.

The studies from González-Guerrero and colleagues have provided a good foundation for further research, and confirmed the validity of a reverse genetics approach to gene identification in the context of nodule metal homeostasis. However, there are some key transporter families that have not yet been interrogated by this group, such as VIT and VTL (See Chapter 5).

3.2 Aims and objectives

In this chapter I will describe how I identified genes with an important role in iron homeostasis in developing nodules. Using a publicly available mutant database, mutants were identified, germplasms obtained, homozygotes isolated and transcript suppression confirmed. These mutants will then form the basis of subsequent phenotypic characterisation (Chapters 4 and 5).

3.3 Bioinformatic analysis of RNAseq and microarray datasets

To identify genes involved in iron homeostasis that are upregulated during nodule development, I analysed expression patterns and levels in three published datasets (Benedito et al., 2008; Breakspear et al., 2014; Roux et al., 2014). The workflow is summarised in Figure 3.1.
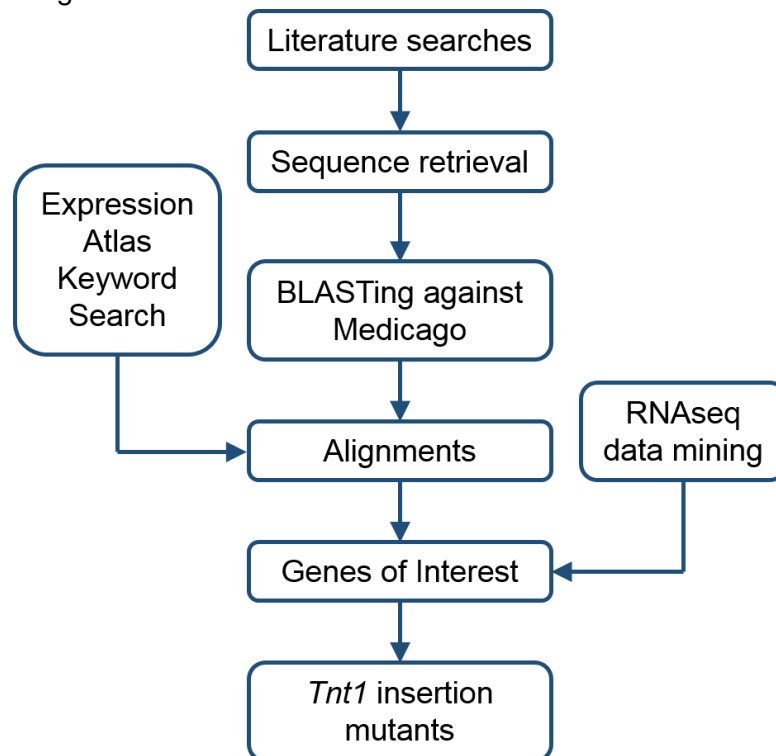
Figure 3.1 Workflow summarising bioinformatics approach to identifying genes of interest for further study using reverse genetics

I first searched the literature for genes involved in iron homeostasis. Iron homeostasis has been extensively studied in Arabidopsis, *Chlamydomonas reinhardtii* and rice, and is highly conserved across the plant kingdom. Genes involved are listed in Tables 3.1-6, and include *FRO*, *NAS*, *NRAMP, IRT1*, *VIT, VTL, ZIP*, *YSL*, and ferritin, along with those involved in iron-sulphur cluster and haem cofactor assembly (Brear et al., 2013). I used BLAST (Phytozome 12.0) to find homologues of all these protein families in the Medicago deduced proteome, and aligned these with the original Arabidopsis sequences to assess similarity; homologues of all the genes were found in Medicago. In many cases, there are two or more paralogues of one Arabidopsis gene. The number of Medicago paralogues follows an ancient whole genome duplication event shortly after the legumes split from the Salicaceae, along with many Medicago genes undergoing tandem duplication (Cannon et al., 2006).

To identify which of these genes or paralogues may be important in symbiosis establishment or maintenance I analysed their expression patterns. At the time, there were two microarray and one RNAseq dataset available for Medicago. Of the microarray datasets, one focussed on the differences in gene expression between nodules and roots (Benedito) while the other looked at gene expression in root hairs in response to Nod factors during early symbiosis establishment (Breakspear). To start with, I used BLAST with the coding sequences of candidate genes as queries against the Benedito et al. and Breakspear et al. microarray datasets to identify probe set IDs (Benedito et al., 2008; Breakspear et al., 2014). In both microarray datasets the Affymetrix Genechip® Medicago Genome Array was used. Each probe set in this dataset consists of 11 probes, each of which is 25 nucleotides. A probe set is assigned to a gene if four or more of these individual 25mer probes are aligned to the same gene. Relevant probe sets that were strongly upregulated during infection and/or nodulation I flagged as genes of interest. For some of my candidate genes I was unable to identify a corresponding probe set ID; a well-known downside of microarrays is incomplete coverage. Another issue with microarrays is poor resolution of transcripts with high sequence identity. For example, initially it was unclear which of the *VTLs* in the block on chromosome 4 was upregulated in nodules; the probe set had high sequence similarity to both *VTL4* and *SEN1*. Fortunately, an RNAseq dataset was published, that by nature had none of these issues described above; I could differentiate between highly similar transcripts, identifying all genes and paralogues except one (Table 3.1). Thus, the RNAseq dataset published by (Roux et al., 2014) proved to be a valuable resource for understanding better the differing roles of genes, even in these tandem arrays.
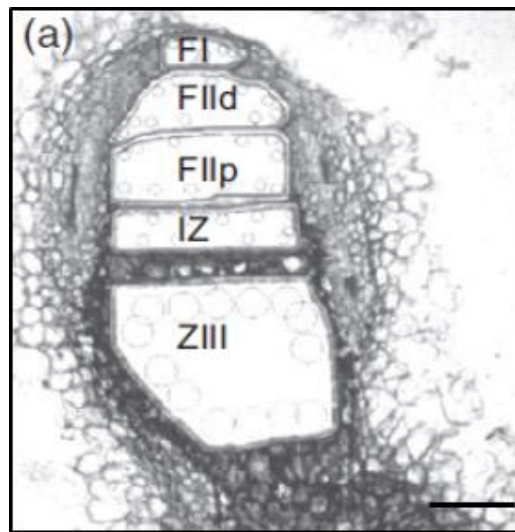
Figure 3.2 Laser microdissection of 15 day-old nodule
Figure taken with permission from (Roux et al., 2014). The five zones are: FI, the nodule meristem; FIId and p, the proximal and distal zones below the meristem, cells undergoing differentiation or infection; IZ, the interzone that separates these cells from ZIII, the nitrogen-fixing zone. Scale bar represents 100 µm.

The experimental approach of (Roux et al., 2014) used microdissection of Medicago nodules, tying gene expression to positional, and therefore developmental, information (Figure 1.4). Additionally, they used a "Ribominus-like strategy" to remove tRNA and rRNA, enriching for mRNA and non-coding RNA from both the host plant and bacterial symbiont. This RNA was then sequenced. This differs from the traditional poly-A enrichment approach, which uses oligo dT primers to amplify mRNA selectively. In eukaryotes, a poly-A tail is added to transcripts as part of the mRNA maturation process (polyadenylation), and its degradation results in mRNA instability (Zhang et al., 2010). By contrast in prokaryotes and some eukaryotic non-coding RNA, a poly-A tail targets RNA for degradation (Mohanty and Kushner, 2011; Zhuang, 2013). The poly-A enrichment method effectively eliminates both bacterial mRNA and eukaryotic non-coding RNA from the RNA pool, skewing the resultant data. The "Ribominus-like strategy" allows this dataset to provide a detailed summary of both plant and bacterial transcriptional variation in the different developmental zones of Medicago nodules, ranging from the meristem, through the differentiation and interzones to the nitrogen fixation zone (Figure 3.3).

To check the reliability of the dataset I looked at the expression profiles of genes involved in assembly of nitrogenase (bacterially-encoded) and the genes for leghaemoglobin (plant-encoded), both of which are highly upregulated in nodules Figure 3.3. The subunits of nitrogenase (*nifHDK*) are upregulated in the interzone and nitrogen fixation zone. This is expected, although previous *in situ* and promoter-GUS studies found high expression in the interzone but little to no expression in the nitrogen fixation zone (Trepp et al., 1999; Labes et al., 1993). This discrepancy is likely caused by the difficulties with the normalisation of the RNAseq dataset to account for the skewed distribution of gene expression in bacteroids and the difficulty in correcting for this. This does lead to some questions concerning the preciseness of the dataset as a whole, however if the demarcation of the "zones" is treated with caution it is still a useful resource. Leghaemoglobins are most upregulated in the interzone, with their expression approximately halving in the nitrogen fixation zone. This agrees with further *in situ* data in *Medicago sativa* from (Trepp et al., 1999).
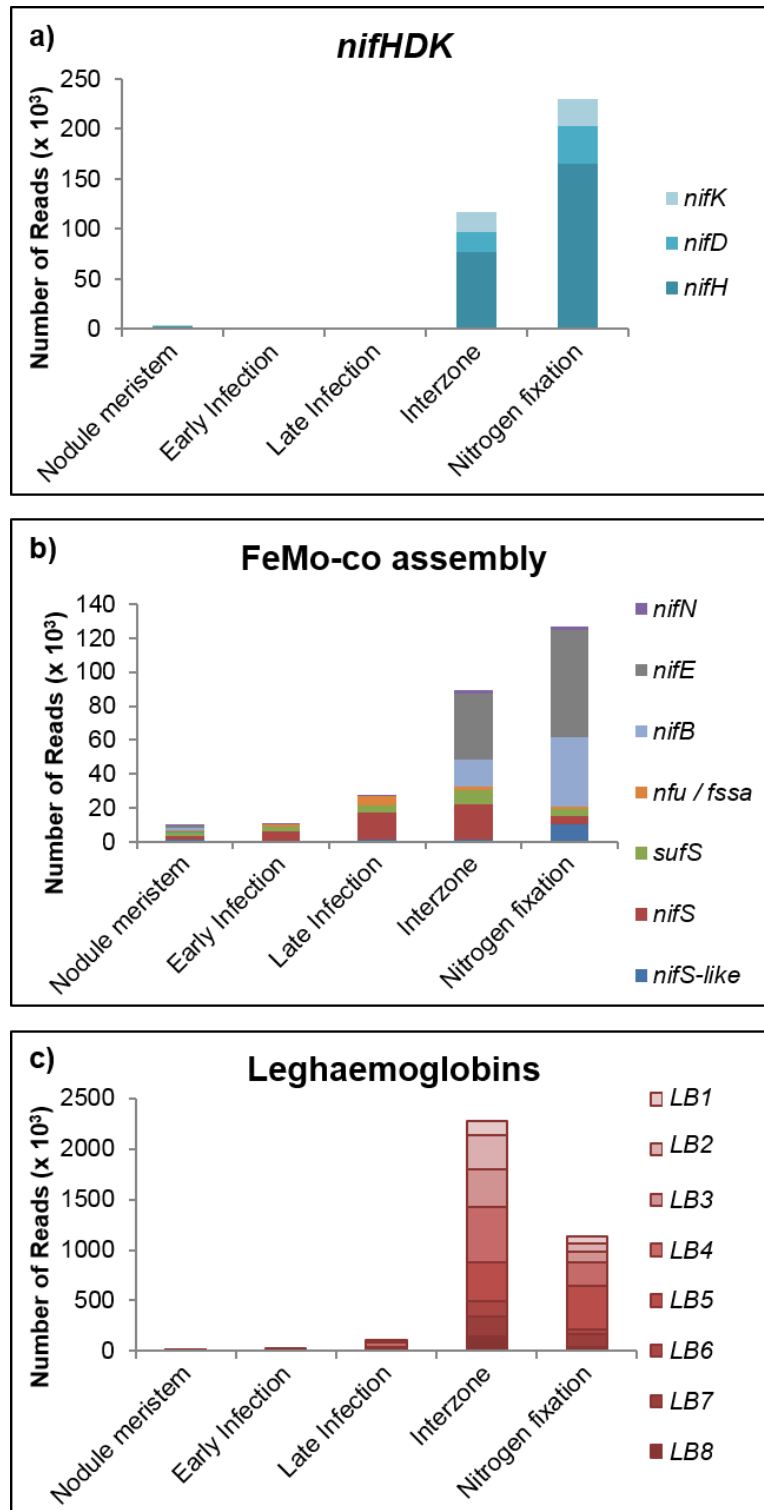
Figure 3.3 Gene expression of selected marker genes in five zones of nodule development
Data taken from (Roux et al., 2014). a) bacterial *nifH, D* and *K* genes, encoding dinitrogenase and dinitrogenase reductase b) Bacterial genes involved in FeMo-co assembly c) Medicago leghaemoglobin genes, of which there are 8.

The RNAseq data represented in Tables 3.1-3 is colour-coded according to the following key:

Key	
	Total reads in top 10% of genes
	Total reads in top 25% of genes
	>50% expression
	35-49% expression
	20-34% expression

As an approximate marker for expression level, genes were ranked by the number of total reads, and those showing zero reads removed. The top 10% and 25% of expressed genes were highlighted. Grey shading denotes genes the total reads of which are in the top 10% (>3955 total reads) or 25% (pale grey, >1102 reads), while the red, orange and yellow shading denotes the percentage of the total expression in a specific zone (>50%, 35-49% and 20-34% respectively).

Table 3.1 Summary of Fe homeostasis gene expression in nodule zones (Roux et al. 2014)

Gene family	Gene ID (V 4.0)	Gene name	Arabidopsis homologue	Total reads	Meristem	Distal zone below meristem	Proximal zone below meristem	Interzone	N_2-fixing zone
FRO	Medtr1g100130		At5g23980	34.7	29.4	7.6	0.0	17.7	45.3
	Medtr1g100150		At5g23980	12.9	55.2	33.1	2.8	0.0	8.9
	Medtr2g028530		At5g49740	3.4	0.0	66.7	33.3	0.0	0.0
	Medtr4g105080		At5g50160	125.4	7.2	4.1	8.6	52.0	28.1
	Medtr8g028780		At1g01580	5.2	53.7	13.0	33.3	0.0	0.0
	Medtr8g028795		At1g01580	2528.6	0.2	0.2	0.8	33.2	65.6
NAS	Medtr1g084050		At1g09240	200.5	14.2	2.8	4.4	1.9	76.7
	Medtr2g070310		At5g56080	1.8	33.3	0.0	33.3	0.0	0.0
	Medtr7g112130		At5g04950	116.2	68.5	10.0	4.7	0.0	16.8
NRAMP	Medtr3g088460	NRAMP1	At1g15960	3951.2	0.3	4.5	50.5	40.9	3.8
	Medtr3g088440	NRAMP2	At1g15960	336.3	48.2	21.2	11.8	13.6	5.1
	Medtr2g104990	NRAMP3	At1g80830	335.6	24.1	23.1	7.7	5.4	39.7
	Medtr3g102620	NRAMP4	At1g47240	544.2	16.1	14.5	19.5	26.8	23.1
	Medtr5g016270	NRAMP5	At2g23150	1624.9	10.8	16.5	10.4	21.8	40.6
	Medtr4g095075	NRAMP7	At5g67330	3812.0	16.3	10.0	4.3	22.2	47.2
VIT	Medtr8g105790	VIT1A	At2g01770	961.8	32.1	27.6	12.1	22.6	5.5
	Medtr8g105810	VIT1B		2.1	50.4	16.3	0.0	0.0	0.0
VTL	Medtr1g099010	VTL1		3.7	33.3	66.7	0.0	0.0	0.0
	Medtr6g034975	VTL2		12.9	54.0	39.5	6.5	0.0	0.0
	Medtr2g008110/ Medtr2g008100	VTL3		22.3	19.9	13.8	2.3	8.1	55.9
	Medtr4g094325	VTL4		40662.4	3.9	44.1	8.6	16.7	26.8
	Medtr4g094330	VTL6		0.0	0.0	0.0	0.0	0.0	0.0
	Medtr4g094332	VTL7		0.0	0.0	0.0	0.0	0.0	0.0
	Medtr4g094335	SEN1		150938.1	0.1	0.2	7.3	68.6	23.8
YSL	Medtr1g007580		At5g53550	0.7	0.0	33.3	0.0	0.0	0.0
	Medtr1g077840		At4g24120	102.2	78.0	14.9	6.5	0.6	0.0
	Medtr3g063490		At1g65730	530.8	0.3	4.1	47.6	26.3	21.7
	Medtr3g092090		At5g53550	3241.6	6.9	5.8	7.1	15.5	64.7
	Medtr5g091600		At1g65730	9.5	88.2	11.8	0.0	0.0	0.0
	Medtr6g077870		At3g17650	2839.9	16.7	47.5	16.3	1.9	17.6
	Medtr7g028250		At3g27020	2069.0	13.6	15.2	18.5	19.1	33.5
ZIP	Medtr2g064310	ZIP1	At3g12750	65.3	23.6	49.5	13.3	4.6	9.1
	Medtr2g097580	ZIP2	At5g59520	6208.3	0.1	0.2	0.1	12.9	86.7
	Medtr3g081580	ZIP3	At2g32270	11.8	79.6	4.7	15.7	0.0	0.0
	Medtr3g082050/ Medtr3g081690	ZIP4/ZIP10	At1g05300	147.2	12.2	9.6	19.4	46.9	12.0
	Medtr3g082050/ Medtr3g081690	ZIP4/ZIP10	At3g58060	25.2	51.1	30.3	6.6	6.9	5.2
	Medtr1g016120	ZIP5	At1g60960	1146.5	10.0	8.3	22.8	27.7	31.2
	Medtr4g083570	ZIP6	At1g31260	10962.5	0.3	8.1	20.1	45.2	26.3
	Medtr3g058630	ZIP7	At1g55910	319.1	28.7	12.5	3.2	12.9	42.6
	Medtr2g098150	ZIP8	At3g20870	775.4	26.7	22.1	28.5	6.6	16.1
	Medtr3g081640	ZIP9	At3g12750	13.7	42.4	15.4	42.2	0.0	0.0
	Medtr3g104400	ZIP11	At1g10970	1258.8	20.9	22.0	25.3	17.5	14.3
	Medtr4g065640	ZIP12	At3g20870	8.8	48.4	10.7	7.6	0.0	0.0
	Medtr5g071990	ZIP13	At2g30080	92.1	17.0	23.8	22.0	19.5	17.8
	Medtr6g007687	ZIP14	At5g62160	2255.7	19.3	21.7	11.9	35.8	11.2
FER	Medtr4g014540	FER1A	At5g01600	39714.7	15.7	50.4	28.0	5.2	0.7
	Medtr7g069980	FER1B	At5g01600	144.8	12.7	13.9	26.3	40.8	6.3
	Medtr5g083170	FER2	At3g11050	29278.6	15.1	38.6	33.2	2.9	10.2
	Medtr5g072560	FER3	At3g11050	8.0	60.3	33.3	0.0	0.0	6.3

Key can be found on Page 52

The expression of genes involved in Fe uptake and storage in the (Roux et al., 2014) dataset is shown in Table 3.1. Only one gene in the *FRO* family was in the top 25% of transcripts detected (*Medtr8g028795*), which is maximally expressed in the nitrogen fixation zone (65.6% of total reads). Neither the *NAS* nor the *VIT* genes were expressed to significant levels in nodules. *NRAMP1, 5* and *7* are all within the top 25% of transcripts detected. Of the 8 *VTL* genes present in Medicago, only two (*VTL4* and *SEN1*) are expressed in nodules, both in the top 10% of transcripts detected. Three members of the *YSL* family are in the top 25% of transcripts detected, two of which (*Medtr3g092090* and *Medtr7g028250*) are maximally expressed in the nitrogen fixation zone, implying they may have a role in Fe remobilisation from nodule cells during senescence. The third *YSL* family member that is highly expressed in nodules (*Medtr6g077870*) is maximally expressed in the early infection zone, suggesting it has a different role from the other two, possibly in Fe delivery to cells in the early infection zone. *MtZIP2* and *6* are both in the top 10% of transcripts detected in nodules, and are maximally expressed in the nitrogen fixation and interzone respectively. *MtZIP5, 11* and *14* are in the top 25% of transcripts detected. Two out of the four ferritins in Medicago are expressed in nodules, and were in the top 10% of detected transcripts. Both *FER1A* and *FER2* are maximally expressed in the infection and differentiation zones. This fits with their hypothesised role in sequestering imported Fe until it is incorporated in one of the various cofactor assembly pathways in developing nodules.

Table 3.2 Summary of Fe-S cluster assembly gene expression in nodule zones (Roux et al. 2014)

Pathway	Gene ID (V 4.0)	Arabidopsis homologue	Total Reads	Meristem	Distal zone below meristem	Proximal zone below meristem	Interzone	N_2-fixing zone
						Roux et al. 2014		
Plastids (SUF)	Medtr3g083980	At1g08490 (NFS2)	853.6	16.8	17.3	21.8	25.0	19.1
	Medtr3g086590	At1g67810 (SUFE2)	431.4	0.8	2.2	53.4	37.7	5.9
	Medtr4g104680	At5g50210 (SUFE3)	187.0	46.0	14.4	38.0	1.6	0.0
	Medtr7g030030	At4g04770 (SUFB)	1578.2	10.1	10.2	12.9	22.7	44.2
	Medtr7g038720		1578.2	10.1	10.2	12.9	22.7	44.2
	Medtr8g019500							
	Medtr8g101390	At3g10670 (SUFC)	2557.9	21.5	30.4	23.1	13.1	11.9
	Medtr5g014290	At1g32500 (SUFD)	525.4	21.0	18.2	19.1	18.4	23.4
	Medtr7g109850	At4g01940 (NFU1)	492.1	19.1	14.9	12.3	20.5	33.2
	Medtr8g092030	At1g10500 (SUFA)	1511.5	8.7	13.5	15.9	32.7	29.3
	Medtr7g079520	At3g54900 (GRXS14)	1750.3	29.2	24.4	17.3	18.4	10.6
Mito (ISC)	Medtr5g014820	At5g65720 (NFS1)	608.6	59.3	33.1	3.6	3.5	0.5
	Medtr7g012870	At5g61220 (ISD11)	1171.7	30.9	21.1	10.2	16.4	21.4
	Medtr4g131820	At4g21090 (ADX1)						
	Medtr3g077720	At4g22220 (ISU1)	2564.8	38.0	29.4	15.5	8.5	8.6
	Medtr3g077730		14.2	54.2	8.8	0.0	37.0	0.0
	Medtr6g071520	At4g03240 (FH)	135.3	42.0	31.9	11.5	1.1	13.5
Cytosol (CIA)	Medtr5g075960	At5g58270 (ATM3)	365.8	37.3	25.8	4.4	12.7	19.8
	Medtr5g075955		330.2	38.9	25.3	2.5	12.6	20.8

Key can be found on Page 52

The expression of iron-sulphur (Fe-S) assembly genes in different nodule zones is shown in Table 3.2. In the plastid Fe-S cluster assembly pathway (SUF), *SUFA*, two of the three Medicago orthologues of *SUFB, SUFC* and *GRXS14* are in the top 25% of detected transcripts. However, when the total reads of the pathway as a whole are considered, there is very little difference; the majority of reads are between 500 and 1500, only an ~3-fold difference. *SUFA* and *SUFB* are maximally expressed in the interzone and nitrogen fixation zone, while *SUFC* and *GRXS14* expression is more evenly distributed, but generally across the meristem and infection zones. From the ISC pathway for mitochondrial Fe-S cluster assembly, only *ISD11* and one of the Medicago paralogues of *ISU1* are in the top 25% of transcripts detected. Both of these are mainly expressed in the meristem and interzone, suggesting they are not needed for the later stages of nodule development. ATM3 is an inner mitochondrial membrane protein, transporting glutathione persulphide out of the mitochondria to be used in the cytosolic iron-sulphur cluster assembly (CIA) pathway. *ATM3* is not highly expressed in nodules, possibly due to the majority of sulphur being diverted to the bacteroids to be used to synthesise the complex Fe-S clusters of nitrogenase.

Table 3.3 Summary of haem cofactor assembly gene expression in nodule zones (Roux et al. 2014)

Gene name	Gene ID (V 4.0)	Arabidopsis homologue	Total reads	Meristem	Distal zone below meristem	Proximal zone below meristem	Interzone	N2-fixing zone
HEMA	Medtr3g111190	At1g58290	7404.9	4.7	5.3	9.2	46.4	34.4
HEMA	Medtr4g092820	At1g58290	777.0	26.4	19.1	19.2	20.4	14.9
HEMA	Medtr5g067410	At1g09940	688.8	0.1	0.3	0.0	77.8	21.8
GSAT	Medtr3g118070	At3g48730	3804.4	20.4	20.2	21.0	20.0	18.5
GSAT	Medtr7g451450	At1g80600	1683.4	15.4	12.7	9.0	18.2	44.8
GSAT	Medtr7g072420	At3g08860	296.0	9.4	14.1	13.7	32.0	30.8
GSAT	Medtr4g076100	At3g22200	1754.9	11.6	28.9	23.1	20.0	16.4
GSAT	Medtr8g091660	At4g39660	4212.2	10.2	8.6	31.0	28.0	22.2
ALAD	Medtr3g088610	At1g69740	6726.4	7.9	17.9	40.5	20.8	12.9
PBGD	Medtr4g076600	At5g08280	4692.6	6.8	9.4	15.9	34.9	33.0
UROS	Medtr3g101350	At2g26540	3990.0	5.9	4.6	15.5	46.9	27.0
UROS	Medtr1g052525	At2g26540	173.9	37.3	19.4	5.2	13.1	25.1
URO	Medtr3g072330	At2g40490	1931.2	8.3	8.6	3.6	31.4	48.1
URO	Medtr2g076340	At3g14930	1718.4	14.2	12.4	20.4	34.5	18.5
CPO	Medtr5g098800	At1g03475	8803.9	2.0	3.4	29.7	41.2	23.7
CPO	Medtr3g075110	At5g63290	137.0	36.3	26.2	20.2	9.2	8.0
PPOX	Medtr7g031310	At5g14220	1322.0	16.6	19.1	22.0	22.7	19.7
PPOX	Medtr1g085730	At4g01690	10507.0	4.1	3.7	13.0	38.0	41.3
FECH1	Medtr3g104260	At5g26030	5151.0	5.8	5.9	19.9	40.7	27.7
FECH1	Medtr4g019850	At5g26030	1471.3	7.7	11.2	23.0	24.8	33.3
FECH2	Medtr8g095320	At2g30390	36.6	66.6	14.8	3.0	0.0	15.7
FECH2	Medtr3g097290	At2g30390	137.5	43.8	20.6	7.9	10.2	17.5
FECH2	Medtr8g095320	At2g30390	108.3	61.7	23.2	3.8	6.3	5.1
FECH2	Medtr4g087690	At2g30390	14.8	33.7	17.9	0.0	48.4	0.0
HO	Medtr8g019320	At2g26670	3580.8	1.9	3.0	10.2	29.8	55.1
HO	Medtr1g019970	At2g26550	529.6	22.9	18.4	20.1	19.0	19.6
FLU	Medtr8g022040	At3g14110	1027.3	12.5	10.8	20.8	38.0	17.9

Key can be found on Page 52

As would be expected given the high levels of haem synthesised by mature nodules, many of the genes involved in tetrapyrrole synthesis are among the most highly expressed in terms of reads detected in the RNAseq dataset (Table 3.3). One of the three *HEMA (GluTR)* orthologues (*Medtr3g111190*) is in the top 10% of transcripts detected, along with one orthologue of *GSA* (*Medtr8g091660*), *ALAD* (*Medtr3g088610*), *PBGD* (*Medtr4g076600*), *UROS* (*Medtr3g101350*), *CPO* (*Medtr5g098800*), *PPOX* (*Medtr1g085730*) and *FECH1* (*Medtr3g104260*). These genes are also predominantly expressed in the interzone and nitrogen fixation zones. In addition, a further seven orthologues of these genes are in the top 25% of transcripts detected, along with one of the orthologues of haem oxygenase (*Medtr8g019320*), the enzyme responsible for the first stage of haem degradation, which is mainly expressed in the nitrogen fixation zone, presumably with a role in haem turnover as leghaemoglobin succumbs to oxidative stress during nodule senescence.

Microarray data from Benedito et al. 2008 and Breakspear et al. 2014 is represented in Tables 3.4-6 and colour-coded according to the following key:

Key
>100
50-100
25-49
10-24
1-9
0.5-1
<0.5

Red shading denotes genes that are more than 100-fold upregulated in nodules (Benedito et al. 2008) or root hairs/*skl* mutants (Breakspear et al. 2014) compared to uninfected roots/root hairs infected with bacteria defective in Nod-factor synthesis respectively. Orange denotes 50-100 times upregulated, yellow 25-49%, white 10-24% and grey 1-9%. Blue and purple shading both represent downregulation of genes, the former 0.5-1-fold upregulated (so up to twofold downregulated), and the latter more than twofold downregulated.

Table 3.4 Summary of expression data for genes involved in Fe uptake and storage (Benedito et al. 2008, Breakspear et al. 2014)

Gene family	Gene ID (V 4.0)	Probeset ID	Gene name	Arabidopsis homologue	Benedito et al. 2008		Breakspear et al. 2014			
					4 dpi	14 dpi	rh 1 dpi	rh 3 dpi	rh 5 dpi	skl 5dpi
FRO	Medtr1g100130	Mtr.51040.1.S1_s_at		At5g23980	0.18	0.15				
	Medtr1g100150	Mtr.38954.1.S1_at		At5g23980	0.18	0.15				
	Medtr4g105080	Mtr.41859.1.S1_at		At5g50160	0.67	0.37				
		Mtr.30221.1.S1_at			0.67					
	Medtr8g028780	Mtr.11719.1.S1_at		At1g01580	0.05	0.10				
	Medtr8g028795	Mtr.13026.1.S1_at		At1g01580	0.03					
NAS	Medtr1g084050	Mtr.46917.1.S1_at			4.19					
	Medtr7g112130	Mtr.43393.1.S1_at			0.25	0.31				0.26
NRAMP	Medtr3g088460	Mtr.17288.1.S1_at	NRAMP1	At1g15960						0.21
	Medtr3g088440	Mtr.51664.1.S1_s_at	NRAMP2	At1g15960	1.50	0.66				
	Medtr3g102620	Mtr.45362.1.S1_at	NRAMP4	At1g47240						0.62
VTL	Medtr6g034975	Mtr.30845.1.S1_at	MtVTL2		0.35	0.55				
	Medtr2g008110/ Medtr2g008100	Mtr.22597.1.S1_s_at	MtVTL3		1.43					
	Medtr4g094325	Mtr.7701.1.S1_at	MtVTL4		229.58	103.79				
YSL	Medtr1g007580	Mtr.14430.1.S1_at		At5g53550		3.14				4.34
	Medtr3g063490	Mtr.11555.1.S1_at		At1g65730	4.64	6.61				
ZIP	Medtr2g064310	Mtr.43988.1.S1_at	ZIP1	At3g12750	0.34					
	Medtr2g097580	Mtr.44349.1.S1_at	ZIP2	At5g59520	0.28	2.13	0.52			
	Medtr3g081580	Mtr.12542.1.S1_at	ZIP3	At2g32270		0.41	0.44			
	Medtr3g082050/	Mtr.38036.1.S1_at	ZIP4/ZIP 10	At1g05300	0.35	0.14	0.27			
	Medtr3g081690	Mtr.28066.1.S1_s_at					0.27			
	Medtr1g016120	Mtr.238.1.S1_at	ZIP5	At1g60960			0.78			
		Mtr.48100.1.S1_at					0.65			
FER	Medtr4g014540	Mtr.48899.1.S1_at	FER1A	At5g01600	25.19	24.06				
		Mtr.5824.1.S1_at			7.40	12.82				
	Medtr7g069980	Mtr.19818.1.S1_at	FER1B			2.19				
	Medtr5g083170	Mtr.12355.1.S1_at	FER2	At3g11050	6.47	7.08				4.86

Key can be found on Page 58

The majority of iron uptake genes identified were downregulated in nodules compared to roots in the microarray dataset of Benedito et al. (2008) (Table 3.4). This included all the genes encoding proteins from the FRO family, and all bar one of the genes encoding ZIP-transporters detected. Interestingly the ZIP-encoding gene found to be upregulated here (*ZIP2*, *Medtr2g097580*) was not the same that encodes the ZIP identified as having a role in nodule zinc homeostasis (*ZIP6*, *Medtr4g083570*, Abreu et al. 2017). One of the two *NAS* genes (Medtr1g084050) is four-fold upregulated in the early nodule (4 dpi) but not in the infected root hairs or mature nodule. This was identified by (Avenhaus et al., 2016) and predicted to have a role in Fe delivery to developing bacteroids, which is consistent with this expression data. The only *NRAMP* gene identified as being slightly upregulated (and only at 4 dpi) was *NRAMP2* (*Medtr3g088440*), again not the *NRAMP* that has previously been identified as important in nodules (*NRAMP1*, *Medtr3g088460*) (Tejada-Jiménez et al. 2015). Both *YSL* genes identified in the microarray datasets were upregulated; one (*Medtr1g007580*) at 14 dpi and in *skl* (a hypernodulating mutant, defective in ethylene perception) and the other (*Medtr3g063490*) upregulated at both 4 and 14 dpi. The only member of any of the Fe uptake families that is more than 100 times upregulated in the microarray data of Benedito et al. is *VTL4*, which is 230 times

upregulated at 4 dpi and 104 times upregulated at 14 dpi. Three of the four Medicago ferritins are upregulated in nodules, although there is some discrepancy between the two probeset IDs associated with *FER1A* (*Medtr4g014540*); one of these suggests that *FER1A* is upregulated 25 and 24 times at 4 and 14 dpi respectively, while the other suggests that *FER1A* is actually only 7 and 13 times upregulated respectively.

The three datasets (Roux et al., 2014; Benedito et al., 2008; Breakspear et al., 2014) are in agreement on relatively few occasions. For example, the two YSLs identified as being upregulated in nodules in the Benedito et al. dataset are not those with the most reads detected in the Roux et al. dataset. As previously mentioned there are also discrepancies with other published genes, such as *NRAMP1* and *ZIP6*, which are reported to be upregulated in nodules by qRT-PCR as well as in the Roux et al. dataset. These differences are mainly because the datasets are asking different questions. Breakspear et al. interrogate the root hair "infectome"; the very early stage of symbiosis establishment, when bacteria are being taken into root hairs through the infection thread. Benedito et al. compared gene expression in nodules to that in roots, and Roux et al. compare the different nodule zones to each other as well as looking at overall reads in nodules.

Table 3.5 Summary of expression data for genes involved in Fe-S cluster assembly (Benedito et al. 2008)

Pathway	Gene ID (V 4.0)	Probeset ID	Arabidopsis homologue	Benedito et al. 2008	
				4d up	14d up
Plastids (SUF)	Medtr4g104680	Mtr.10622.1.S1_at	At5g50210 (SUFE3)	0.39	0.26
	Medtr7g030030, Medtr7g038720, Medtr8g019500	Mtr.13351.1.S1_at	At4g04770 (SUFB)	0.66	
	Medtr5g014290	Mtr.13554.1.S1_at	At1g32500 (SUFD)	1.60	
Mito (ISC)	Medtr5g014820	Mtr.38686.1.S1_at	At5g65720 (NFS1)		0.51
	Medtr3g077720	Mtr.22267.1.S1_at	At4g22220 (ISU1)		0.56
	Medtr3g077730	Mtr.22269.1.S1_at			0.77
	Medtr6g071520	Mtr.40931.1.S1_at	At4g03240 (FH)		0.67
Cytosol (CIA)	Medtr5g075960, Medtr5g075955	Mtr.4354.1.S1_at	At5g58270 (ATM3)		0.65

Key can be found on Page 58

The majority of genes involved in plastid, mitochondrial and cytosolic Fe-S cluster assembly were not present in the Benedito et al. dataset, and none of them were found in the Breakspear et al. dataset (Table 3.5). Of those that were present in the Benedito et al. dataset, only one (*Medtr5g014290*, *SUFD*) is upregulated (1.6-fold). The low expression of Fe-S assembly genes as a whole in this dataset fits with the relatively low total reads of these genes in Roux et al. (2014).

Table 3.6 Summary of expression data for genes involved in haem cofactor assembly (Benedito et al. 2008, Breakspear et al. 2014)

Gene name	Gene ID (V 4.0)	Probeset ID	Arabidopsis homologue	Benedito et al. 2008		Breakspear et al. 2014		
				4 dpi	14 dpi	rh 1 dpi	rh 3 dpi	rh 5 dpi
HEMA	Medtr3g111190	Mtr.35911.1.S1_at	At1g58290	0.76	1.62	1.11		
		Mtr.43354.1.S1_at			2.83			
GSAT	Medtr7g451450	Mtr.41072.1.S1_at	At1g80600		2.35			
		Mtr.37917.1.S1_at		1.69	3.12			
	Medtr7g072420	Mtr.40048.1.S1_at	At3g08860	1.62				
ALAD	Medtr3g088610	Mtr.51672.1.S1_at	At1g69740	1.98	1.99			
PBGD	Medtr4g076600	Mtr.19916.1.S1_at	At5g08280	1.95	2.75			
UROS	Medtr3g101350	Mtr.31465.1.S1_at	At2g26540	0.80				0.75
URO	Medtr3g072330	Mtr.24083.1.S1_s_at	At2g40490		2.32			
		Mtr.24083.1.S1_at		1.32	2.95			
	Medtr2g076340	Mtr.48466.1.S1_at	At3g14930		1.70			
CPO	Medtr5g098800	Mtr.9105.1.S1_at	At1g03475	1.85	6.29			
PPOX	Medtr7g031310	Mtr.39484.1.S1_s_at	At5g14220	1.40	1.33			
		Mtr.17212.1.S1_at		1.78	1.75			
	Medtr1g085730	Mtr.43898.1.S1_at	At4g01690		5.50			
FECH	Medtr4g087690	Mtr.48945.1.S1_s_at	At2g30390		0.66			
HO	Medtr8g019320	Mtr.40298.1.S1_at	At2g26670		1.89			
		Mtr.5569.1.S1_at			6.16	0.48		
	Medtr1g019970	Mtr.11243.1.S1_at	At2g26550					
FLU	Medtr8g022040	Mtr.38591.1.S1_at	At3g14110	2.29	6.02			

Key can be found on Page 58

Microarray data on haem cofactor assembly genes showed at least one of the genes at each stage of haem biosynthesis was upregulated except for *UROS* (*Medtr3g101350*) and the only *FECH* found in the microarray dataset (*Medtr4g087690*) (Table 3.6). All of the other stages are between 1.3-fold (*URO*, *Medtr3g072330*) and 6.3-fold upregulated (*CPO*, *Medtr5g098800*). Interestingly, both haem oxygenase (*HO,* the gene coding for the first enzyme in haem degradation) and *FLU* (the product of which inhibits HEMA; the first stage of tetrapyrrole biosynthesis in plants) are upregulated in 14 dpi nodules. This could be at the onset of nodule senescence when leghaemoglobin is no longer required at such high concentrations and begins to degrade.

As was the case for the Fe uptake genes, the orthologues of haem synthesis genes that were found to be upregulated in the microarray datasets were not necessarily those that were detected at the highest levels in the RNAseq experiments of Roux et al. For example, although *Medtr3g111190* (*HEMA*) is slightly upregulated in 1 dpi root hairs, 14 dpi nodules and detected at high levels in the RNAseq dataset, the two *GSAT* genes

found to be slightly upregulated in the microarray dataset (*Medtr7g451450* and *Medtr7g072420*) are the two most lowly detected orthologues in the RNAseq dataset. Also of note is the microarray data on the *FECH* genes; again the only gene to be detected as having an altered expression level was the one with the lowest number of reads detected in the RNAseq data. These differences could all be caused by one of the intrinsic problems with microarrays; if a set of probes is designed against a conserved region of a transcript it becomes very difficult to resolve precisely which gene within a family is showing altered regulation.

3.4 Selection of genes of interest

At this point I decided to focus on the ferritin, ferrochelatase and VIT-like (*VTL*) genes for a variety of reasons. Little is known about how the nodule balances iron uptake and usage, but the upregulation of ferritin in the nodule infection zones (Figure 3.4a) implies a role of ferritin as a buffer for excess iron. The haem biosynthesis pathway in nodules is unique in the plant kingdom in its extreme upregulation. As such, it is a useful way to study haem biosynthesis and its regulation in plants. The final stage of haem biosynthesis (the insertion of Fe into the protoporphyrin ring, catalysed by ferrochelatase) is the only stage that is specific to the haem pathway, the other stages being common to other tetrapyrroles. This makes ferrochelatases in nodules an interesting target for further study (Figure 3.4b). Another group was working extensively on most families of iron transporters in nodules but not the *VTL*s (Figure 3.4c), as some work had already been published on one member of this family, *SEN1,* in Lotus, although this was incomplete (Hakoyama et al., 2012). In addition, there were mutants available in these genes of interest, facilitating further study. The ferritins, ferrochelatases and *VTL* genes will be described in more detail in the next three sections.
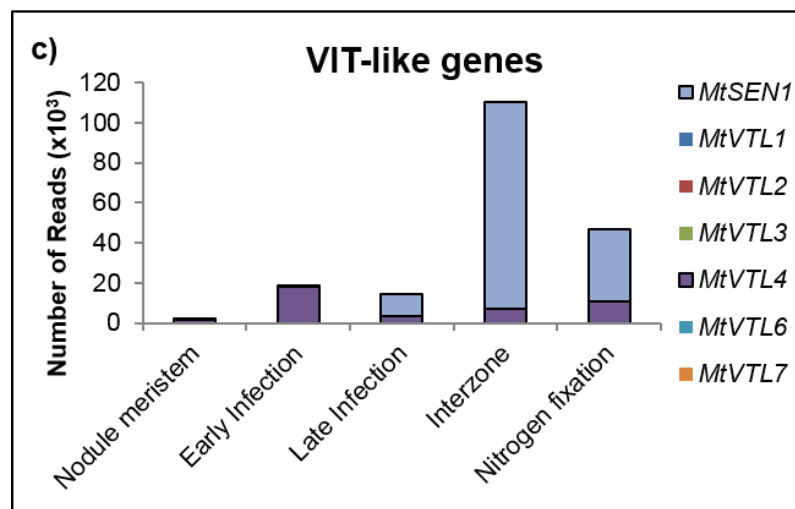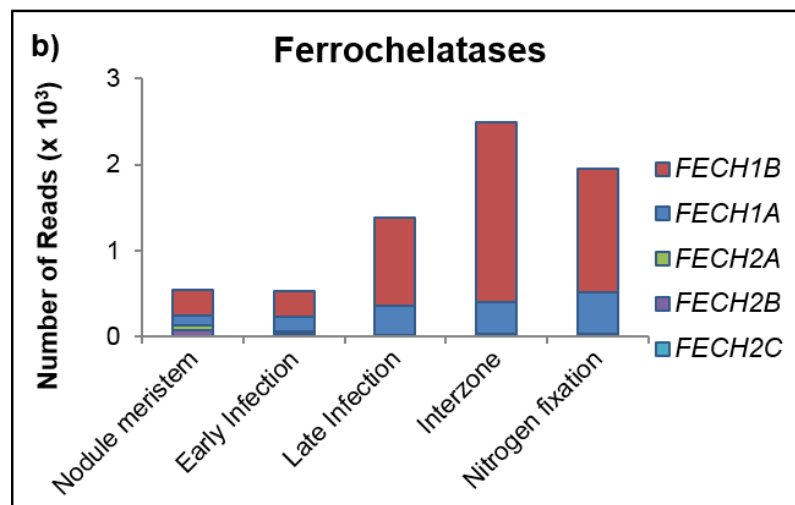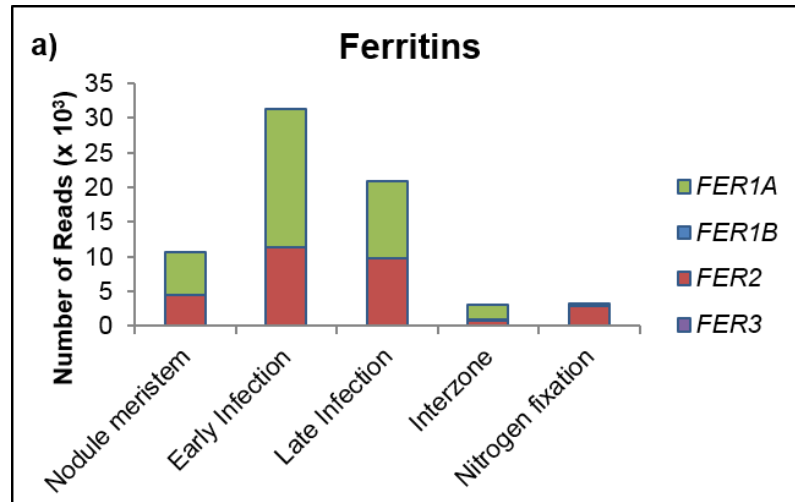
Figure 3.4 Level of gene expression in nodules of genes of interest over five zones of nodule development.
Data taken from (Roux et al., 2014). a) Ferritins, two out of the four encoded by Medicago are upregulated in nodules b) Ferrochelatase genes, two of which are specifically upregulated in nodules (*FECH1A* and *FECH1B*) c) Vacuolar iron transporter like (*VTL*) genes. Of the family (8 members) only two were upregulated during nodulation.

3.4.1 Identification and assignment of Medicago ferritin genes

There are four ferritin paralogues in Medicago; they have not been assigned names and numbers. Phylogenetic analysis of these ferritins alongside the four Arabidopsis proteins suggests that the Medicago ferritin genes cannot be assigned as direct orthoologues of the Arabidopsis genes (Figure 3.5). Although there is some degree of relatedness between, for example, ATFER1 and Medtr4g014540 and Medtr7g069980, the low bootstrap value of 49 suggests that this relatedness is not significant (Figure 3.5). There is still a high degree of relatedness between the Medicago isoforms Medtr4g01450 and Medtr7g069980 however, so I have assigned these as MtFER1A and MtFER1B respectively. Medtr5g083170 I have assigned as MtFER2 and Medtr5g072560 (a truncated variant) as MtFER3. In Arabidopsis, the seed-specific ferritin is AtFER2. From the expression pattern across different tissues, the seed-specific ferritin in Medicago looks to be MtFER1B (Benedito et al. 2008).

The expression pattern of the Medicago ferritins in the developing nodule is shown in Figure 3.4a. Expression is highest in the early infection zone and tails off as the nodule matures. This correlates with the punctate Fe localisation in the infection and differentiation zone of Medicago nodules, consistent with it being encased in ferritin cages (Rodríguez-Haas et al., 2013). The decrease in ferritin expression in the mature nodule is likely due to a decrease in free Fe as it is used up in haem and nitrogenase cofactor biosynthesis. Genes associated with nitrogenase cofactor assembly are upregulated in the interzone, before peaking in the nitrogen fixation zone (Figure 3.3a and b), supporting this hypothesis.

Figure 3.5 Phylogenetic analysis of ferritin isoforms in Medicago
Phylogeny was produced using the Neighbour-joining method (Saitou and Nei, 1987).
This is the optimal tree, sum of branch length =1.57726192. 1000 bootstraps were
used (Felsenstein, 1985). Tree drawn to scale, with branch lengths in the same units
as those of the evolutionary distances used to infer the tree. Distances calculated
using the Poisson corrected method (Zuckerkandl and Pauling, 1965). All gaps and
missing data were eliminated. There were 137 positions in the final dataset. Analysis
conducted in MEGA7 (Kumar et al., 2016) No branches were condensed.

3.4.2 Identification and assignment of Medicago ferrochelatase genes

Of the five Medicago FECH orthologues, two group with Arabidopsis FECH1, while the other three group with Arabidopsis FECH2 (Figure 3.6). The two Medicago "Type I" FECH proteins (Medtr4g019850 and Medtr3g104260) I have assigned as FECH1A and FECH1B respectively. Similarly, I have assigned the three Medicago "Type II" FECH proteins (Medtr8g095320, Medtr3g097290 and Medtr4g0874690) as FECH2A, FECH2B and FECH2C. It is not clear whether all five of these isoforms are functional, especially FECH2C, which is truncated to 96 amino acids compared to the ~530 residues of the mature protein. Only "Type I" *FECH* genes are expressed in nodules (Figure 3.4b). This is consistent with expression patterns in Arabidopsis; *FECH2* is expressed very little in roots, while *FECH1* is expressed in all tissues and is upregulated in response to stress (Scharfenberg et al., 2014). One of the *FECH1* paralogues (*FECH1A*) is expressed at low levels throughout all the zones of the nodule, while *FECH1B* expression is induced in the late infection zone, peaking in the interzone before tailing off slightly in the nitrogen fixation zone. This pattern ties in with the pattern of the leghaemoglobins (Figure 3.3c), with *FECH1B* expression peaking slightly earlier, implying that haem cofactor is synthesised ready to be incorporated into the leghaemoglobin apoprotein.
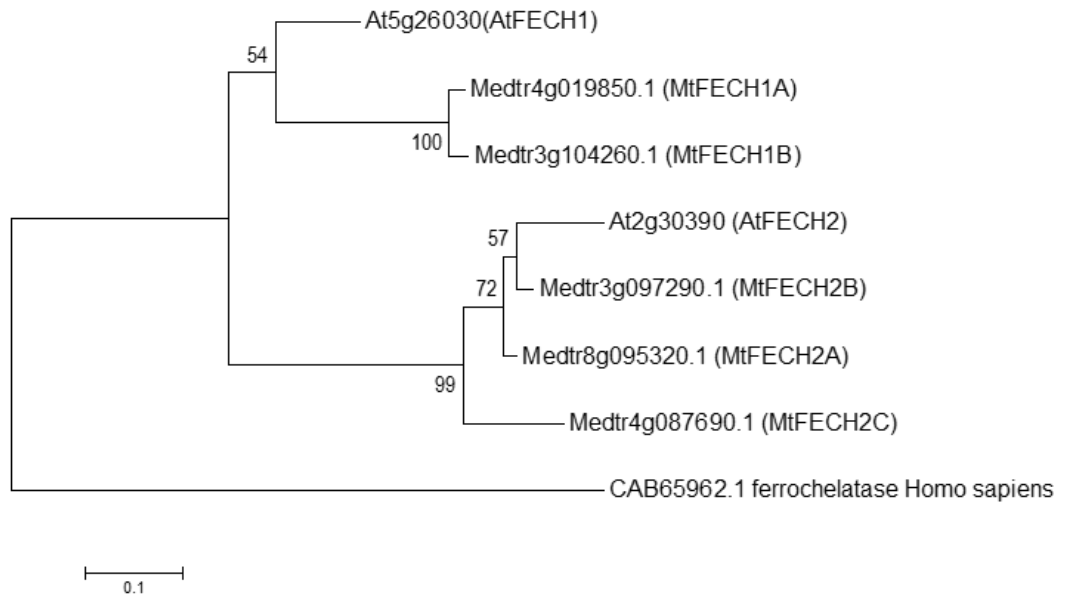
Figure 3.6 Phylogenetic tree of Medicago ferrochelatase isoforms.
Phylogeny was produced using the Neighbour-joining method (Saitou and Nei, 1987).
This is the optimal tree, sum of branch length = 1.69163098. 1000 bootstraps were
used (Felsenstein et al. 1985). Tree drawn to scale, with branch lengths in the same
units as those of the evolutionary distances used to infer the tree. Distances
calculated using the Poisson corrected method (Zuckerkandl and Pauling, 1965). All
gaps and missing data were eliminated. There were 59 positions in the final dataset.
Analysis conducted in MEGA7 (Kumar et al. 2016)

3.4.3 Identification and assignment of Medicago vacuolar iron transporter-like (VTL) genes

The VTL family in Medicago is relatively large with eight members, five of which are in a tandem array. A phylogenetic analysis of these genes is shown in Figure 3.7. From this analysis, it is clear that there has been significant divergence of the Arabidopsis and Medicago VTL peptide sequences since their last common ancestor, so unlike for the ferrochelatases it was not possible to name the Medicago VTLs according to their closest Arabidopsis homologue. I have assigned them new numbers as shown in Figure 3.7.

The *Lotus japonicas SEN1* gene product has been shown to be necessary for the formation of functional nodules (*i.e.* capable of reducing nitrogen) (Hakoyama et al., 2012). The authors observed its similarity to *ScCCC1*, and on this basis hypothesised a role in Fe homeostasis, although my phylogenetic analysis and alignments show it belongs in the *VTL* family. The main difference between the VIT and the VTL family proteins is the deletion of a cytosolic loop between predicted transmembrane helices 2 and 3 in the latter. The closest homologue of *LjSEN1* was identified by BLASTing the amino acid sequence against Medicago as Medtr4g094335, and was named *MtSEN1*. Figure 3.4c shows the expression patterns of genes in the *VTL* family. Only *VTL4* and *SEN1* are expressed in nodules. *SEN1* has a much higher overall expression than *VTL4*, and is expressed later in nodule development.
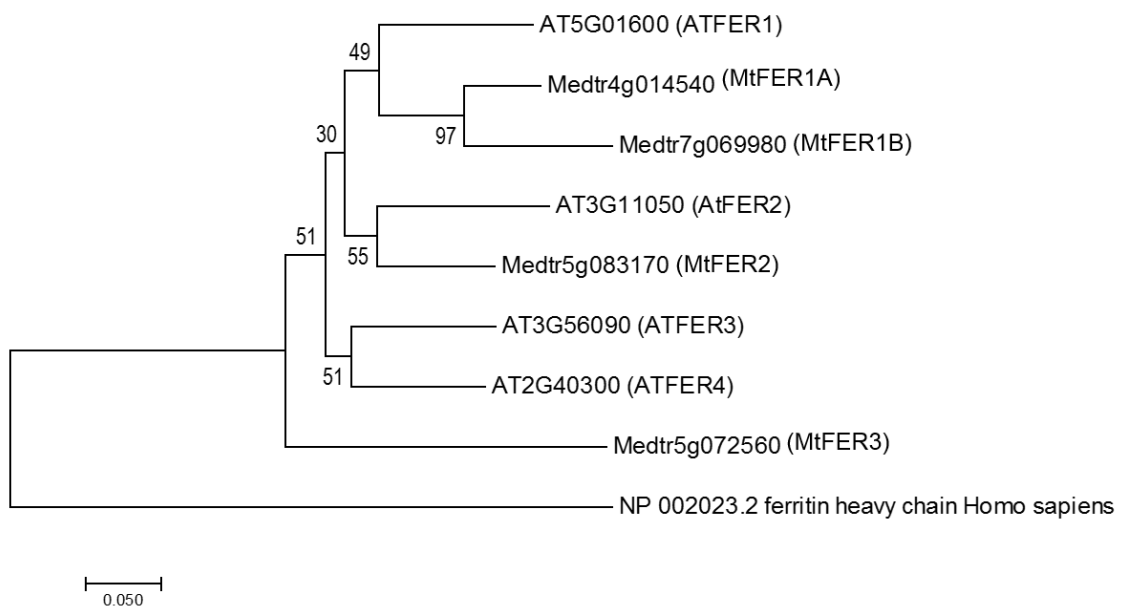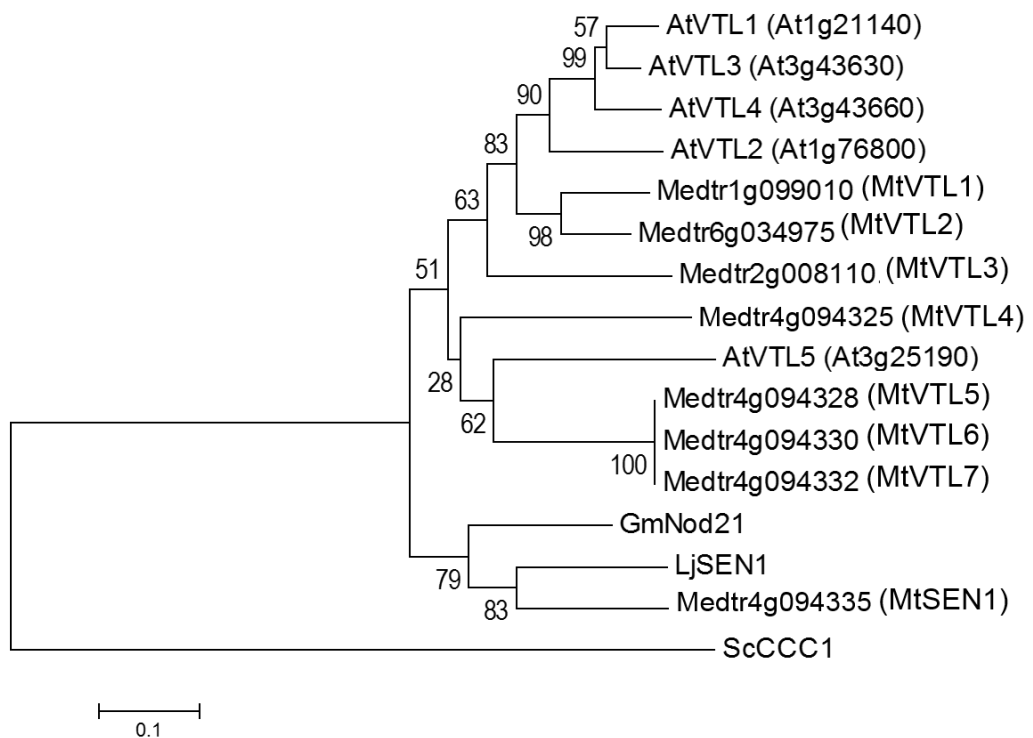
Figure 3.7 Phylogenetic tree of Medicago VTLs
Phylogeny was produced using the Neighbour-joining method (Saitou and Nei, 1987). This is the optimal tree, sum of branch length = 3.14773297. 1000 bootstraps were used (Felsenstein et al. 1985). Tree drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the tree. Distances were calculated using the Poisson corrected method (Zuckerkandl and Pauling, 1965). All gaps and missing data were eliminated. There were 158 positions in the final dataset. Analysis conducted in MEGA7 (Kumar et al. 2016)

3.4.4 Summary of genes selected for further study

Genes selected for further study using this reverse genetics approach, along with expression data are summarised in Table 3.7. Only *VTL4/SEN1, FER1A* and *FER2* were present in the dataset of Benedito et al. (2008) *VTL4/SEN1*, probeset ID Mtr.7701.1.S1_at, is ~230-fold upregulated at 4 days post inoculation (dpi) and 104-fold upregulated at 14 dpi compared to uninoculated roots. The ferritin orthologues were more consistent between the two timepoints, with *FER1A* 25- and 24-fold upregulated and *FER2* 6- and 7-fold upregulated. Neither ferrochelatase orthologue was detected in either microarray dataset. Another feature of the Breakspear et al. 2014 dataset was the use of the *skl* mutant, which has an increased infection rate due to defective ethylene signalling, effectively amplifying the levels of infectome transcripts detected. A 5-fold induction of *FER2* was observed at 5 dpi in *skl* mutants. This could imply that there is an even earlier influx of Fe than was suggested by the Roux dataset, leading to ferritin induction to detoxify this influx.

Table 3.7a Summary of expression data for selected genes (Roux et al. 2014)

Gene ID (V 4.0)	Gene name	Arabidopsis homologue	Total reads	Meristem	Distal zone below meristem	Proximal zone below meristem	Interzone	N2-fixing zone
Medtr4g019850	*FECH1A*	At5g26030	1471.3	7.7	11.2	23.0	24.8	33.3
Medtr3g104260	*FECH1B*	At5g26030	5151.0	5.8	5.9	19.9	40.7	27.7
Medtr5g083170	*FER2*		29278.6	15.1	38.6	33.2	2.9	10.2
Medtr4g014540	*FER1A*		39714.7	15.7	50.4	28.0	5.2	0.7
Medtr4g094335	*SEN1*		150938.1	0.1	0.2	7.3	68.6	23.8
Medtr4g094325	*VTL4*	At3g25190	40662.4	3.9	44.1	8.6	16.7	26.8

Table 3.7b Summary of expression data for selected genes (Benedito et al. 2008, Breakspear et al. 2014)

Gene ID (V 4.0)	Probeset ID	Gene name	Arabidopsis homologue	Benedito et al. 2008 4 dpi (AB)	Benedito et al. 2008 14 dpi (AB)	Breakspear et al. (2014) rh 1 dpi (AB)	Breakspear et al. (2014) rh 3 dpi (AB)	Breakspear et al. (2014) rh 5 dpi (AB)	Breakspear et al. (2014) skl 5dpi (AB)
Medtr4g019850		*FECH1A*	At5g26030						
Medtr3g104260		*FECH1B*	At5g26030						
Medtr5g083170		*FER2*		6.47	7.08				4.86
Medtr4g014540	Mtr.12355.1.S1_at	*FER1A*		25.19	24.06				
Medtr4g094325	Mtr.7701.1.S1_at	*VTL4*	At3g25190	229.58	103.79				

3.5 Mutant selection and isolation

3.5.1 Noble Foundation *Tnt1* mutants

To identify mutants in the selected genes of interest (*FECH1A, FECH1B, FER1A, FER2, SEN1* and *VTL4*), I screened the Medicago mutant collection of the Noble Foundation, which contains ~21 000 *Tnt1* mutants, created by multiple insertions of a retrotransposon from *Nicotiana tabacum*, *Tnt1* (Tadege et al., 2008). Mutant lines within the library have on average 25 insertions, with a higher than expected proportion of these insertions being in coding regions, increasing the gene coverage. The large number of mutant lines generated (each with multiple insertions) means that this library covers approximately 90% of Medicago genes. In combination with the availability of Flanking Sequence Tag (FST) information on the database, this high coverage makes this collection a good source of mutant germplasm. I initially searched the FST database for my genes of interest; this only yielded one mutant, in *FER2*. For other genes, I ordered PCR screening of the mutant population, which was carried out by the Noble Foundation. This yielded at least one mutant line for all my genes of interest (Table 3.8), except for *SEN1*. Due to the high sequence similarity between *VTL4* and *SEN1*, FSTs that had previously been assigned to *SEN1* were found to be a closer match to *VTL4.*

Table 3.8 Location of *Tnt1* insertions in mutant lines obtained from the Noble Foundation

Gene ID[a]	Gene name	Gene acronym	Location[b]	Line No.[c]	Orientation of Tnt
Medtr4g014540	Ferritin-1A	*FER1A*	at 1140 bp exon	NF8502	Reverse
Medtr5g083170	Ferritin-2	*FER2*	at 3206 bp exon	NF18993	Forward
Medtr4g019850	Ferrochelatase type 1	*FECH1A*	at 115 bp, 6 bp to exon	NF12998	Reverse
Medtr3g104260	Ferrochelatase type 1	*FECH1B*	at 535 bp exon	NF5478	Reverse
			at 660 bp intron	NF3665	Forward
Medtr4g094325	Vacuolar Iron Transporter-like 4	*VTL4*	at ~730 bp exon	NF17463	Reverse
			at ~1050 bp exon	NF21016	Reverse

a Gene assignment from v4.0 of the Medicago genome.
b Distance of the *Tnt1* insertion from the start codon.
c Accession code for the mutant line.

3.5.2 Gamma irradiation-induced deletion mutation

A mutant line of *Medicago truncatula* A17 containing a 30 kilobase deletion in chromosome 4 encompassing both *VTL4* and *SEN1* (Figure 3.8) was isolated by a Péter Kaló, (Agricultural Biotechnology Center, Hungary). This mutant was generated using gamma irradiation in an A17 background; this is a different background from the *Tnt1* mutant collection, which was generated in R108 as it is more amenable to growth in tissue culture. The 13U mutant can form nodules but these are white and cannot fix nitrogen (Domonkos et al., 2013). As well as *SEN1* and *VTL4*, 13U has deletions of *VTL5, 6* and *7*, but as previously mentioned, none of these are expressed in nodules (Figure 3.4c). Three other genes (*Medtr4g094320, Medtr4g094322* and *Medtr4g094338*) are either partially or entirely missing in the 13U mutant, however their functions are unknown, as is their possible role in symbiosis establishment. The existence of this mutant is fortunate, as the proximity of *SEN1* and *VTL4* on chromosome 4 would have made it almost impossible to make a *sen1vtl4* double mutant even if a *sen1* mutant had been available.
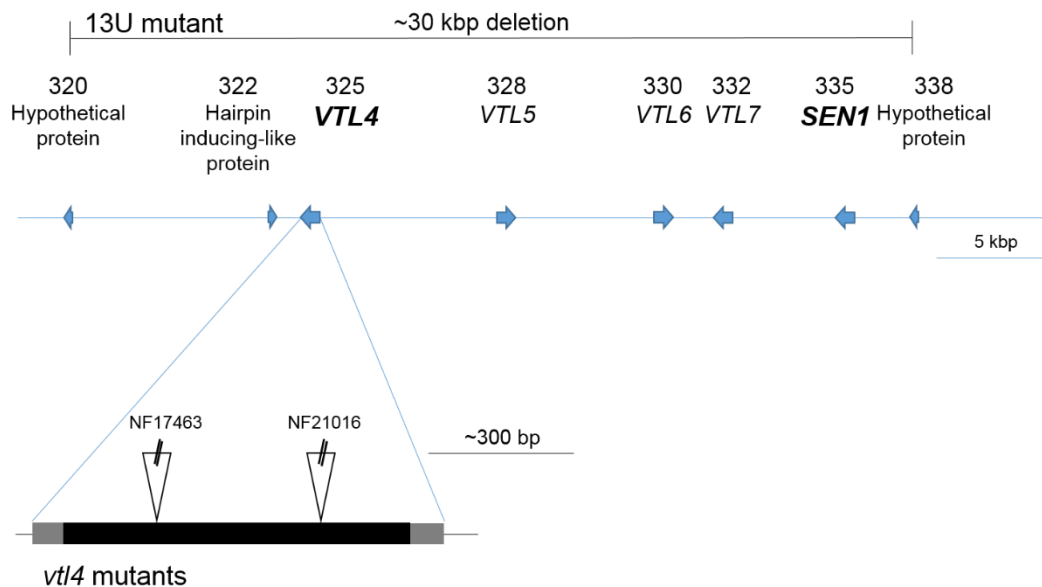
Figure 3.8 Gene map of showing the position of the 13U mutant.
13U is characterised by a 30 kbp deletion in chromosome 4. Individual genes are represented by blue blocks and the positions of the two *Tnt1* mutants in *VTL4* are shown in the lower panel

3.6 Confirmation of mutant genotypes

3.6.1 Confirmation of mutant genotypes

The Noble Foundation sends approximately twelve seeds for each line ordered. The number of *Tnt1* insertions in each mutant line means that a population of this size is highly likely to contain individuals that are heterozygous for a specific insertion of interest. From the segregating population of seeds sent, homozygotes were identified by extraction of genomic DNA and PCR against both wild-type and mutant alleles, using two gene-specific primers or one gene-specific primer and a *Tnt1*-specific primer respectively.

Two lines were ordered with insertions in *VTL4* (NF17463 and NF21016), both of which are in the single exon. Line NF18993 has an insertion in exon seven of *FER2*, and NF8502 has an insertion in exon four of *FER1A*. In the *FECH* genes, there is a *Tnt1* insertion 6 bases into the first intron of *FECH1A* in line NF12998, which could potentially disrupt splicing. Of the two lines with insertions in *FECH1B*, NF5478 has a *Tnt1* in the second exon while NF3665 there is one 660 bases into the second intron, so is unlikely to disrupt transcript levels.

Using the PCR strategy described above, homozygous lines were isolated for further study. Figure 3.9 shows gene maps of the genes of interest and summarises primer positions, genotyping PCR results and initial zygosity ratios.

VTL4 (Medtr4g094325)

wt	het	hom	?	Total
2	6	3	0	11

wt	het	hom	?	Total
3	3	3	2	11

FER1A (Medtr4g014540)

wt	het	hom	?	Total
7	4	8	2	21

FER2 (Medtr5g083170)

wt	het	hom	?	Total
6	5	1	0	12
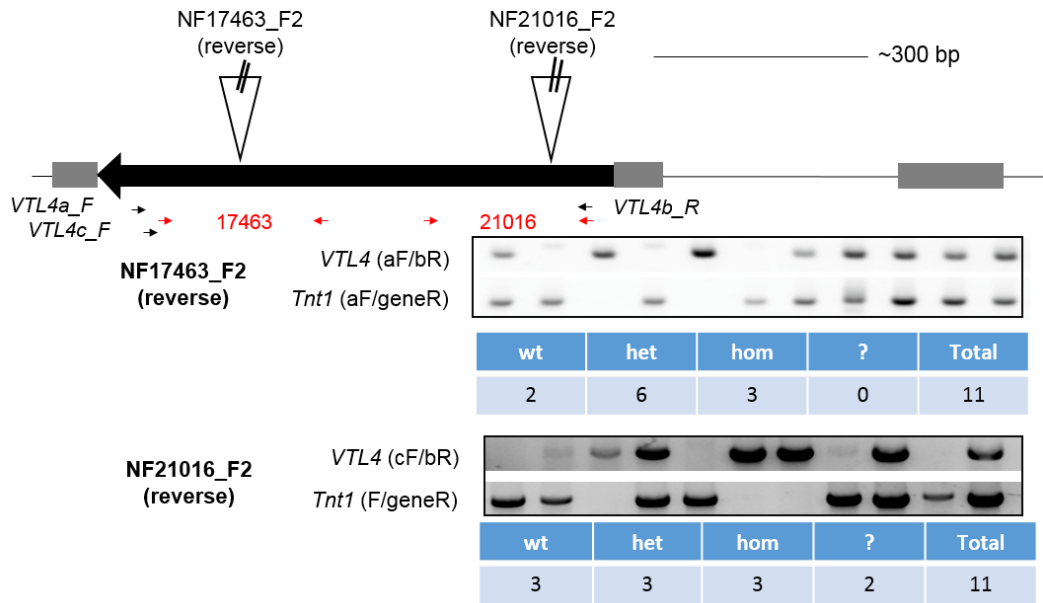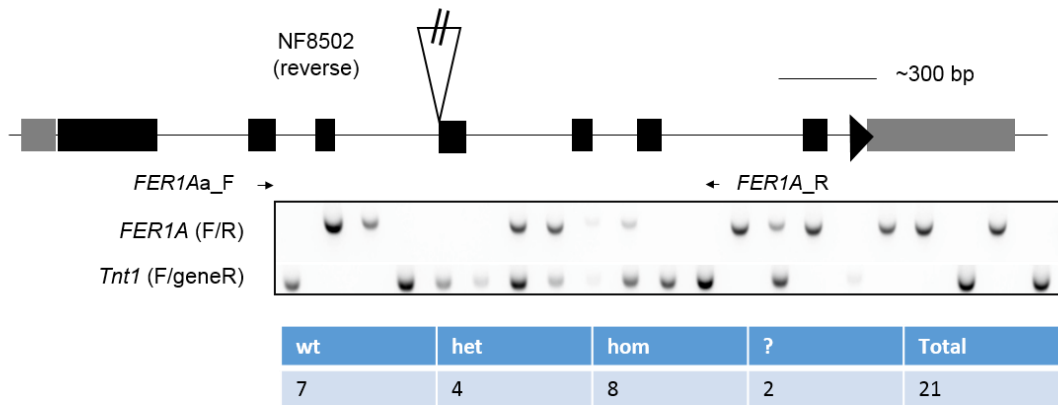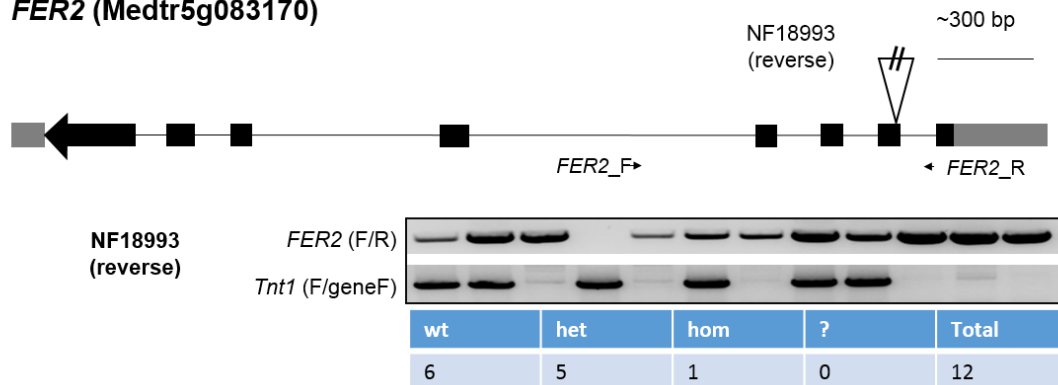
FECH1A (Medtr4g019850)

Medtr4g019850.2 (Primary transcript)

NF12998 (reverse)

~300 bp

FeCh1a_F   FeCh1a_R

FECH1A (F/R)
Tnt1 (F/geneR)

wt	het	hom	?	Total
6	12	1	4	23
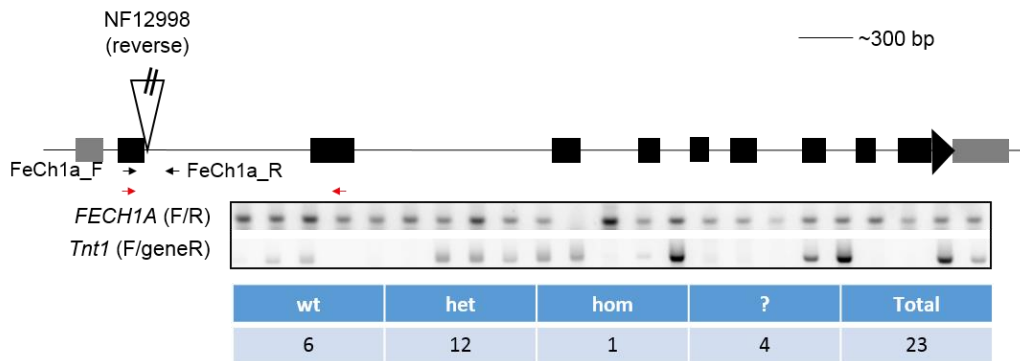
FECH1B (Medtr3g104260)

Medtr3g104260.1 (Primary transcript)

NF5478 (reverse)   NF3665 (forward)

~300 bp

FECH1Ba_F   FECH1Bb_R

NF3665 (forward)

FECH1B (aF/bR)
Tnt1 (F/genebR)

wt	het	hom	?	Total
1	9	6	1	17

NF5478 (reverse)

FECH1B (aF/bR)
Tnt1 (F/geneaF)

wt	het	hom	?	Total
9	2	8	1	20
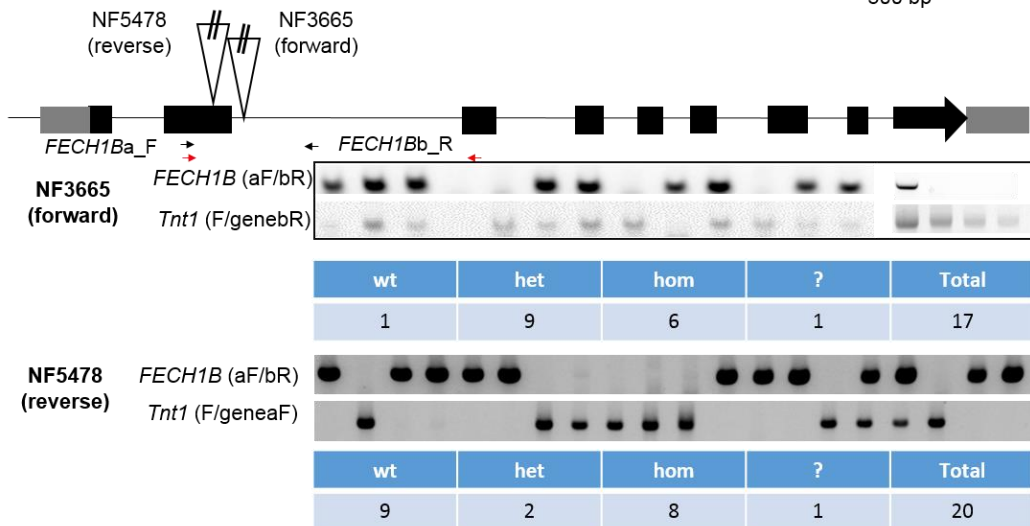
Figure 3.9 Positions of *Tnt1* insertions in *Medicago truncatula* genes
Indicated with associated genotyping information. Exons are represented by black boxes, UTRs by grey boxes, genotyping primers by black arrows and RT-PCR primers by red arrows. gDNA was extracted from young leaflets. Genotyping primers were designed to span the *Tnt1* insertion to check for wild type alleles, and a second PCR using one primer from inside the *Tnt1* and one in the gene to check for the mutant allele.

At least one homozygote plant was isolated from each of these segregating populations. Seeds from these were then bulked up for use in subsequent experiments. A 1:2:1 ratio of wild type:heterozygous:homozygous seeds could be expected for any given insertion. In some lines, there were fewer homozygotes than expected; in NF12998, the segregation ratio of the insertion in *FECH1A* was 6:12:1 while in NF18993 the ratio of insertion in *FER2* was 6:5:1. In NF3665 there were many more homozygotes than expected; insertions in *FECH1B* were 1:9:6. In others there were fewer heterozygotes than expected, but fairly even numbers of wild-type and homozygous plants. In NF5478 the ratio of insertions in *FECH1B* was 9:2:8, in NF21016 insertions in *VTL4* were found in a 3:3:3 ratio and in NF8205 they were found in a 7:4:8 ratio.

Following selection of homozygotes, *fer* mutants showed little phenotype compared to *fech* and *vtl4* mutants, possibly due to gene redundancy. The two ferritins that are highly expressed in nodules (*FER2* and *FER1A)* are also the most highly expressed in all other tissues, so any phenotypes identified could not be unequivocally linked to symbiosis establishment. On this basis I decided not to continue to study the ferritin mutants.

3.6.2 Confirmation of transcript disruption

To confirm lack of transcript in the selected mutant lines, RT-PCR was used. Where practical, primers were designed to span an intron, although this was not possible for *VTL4* as it only has one exon. RT-PCR results are summarised in Figure 3.10. Both lines with an insertion in *VTL4* show complete abolition of transcript, so from here on NF17463 will be referred to as *vtl4-1* and NF21016 as *vtl4-2*. *FECH1B* transcript is lacking in the line NF5478, which will now be referred to as *fech1b-1*. *FECH1A* transcript was still present in NF12998, despite the *Tnt1* insertion only being 6 bases into an intron. Sequencing the RT-PCR product revealed that it was identical to the wild-type transcript. Despite further screening no other lines with insertions in *FECH1A* were available, so this gene was not taken forward for further study.

Figure 3.10 Expression of *FECH1A*, *FECH1B* and *VTL4* in wild type and mutant plants RNA was extracted from nodulated roots unless otherwise stated from the indicated Nobel Foundation (NF) *Tnt1* insertion lines. *ACTIN* was used as a control. PCR cycles used were as follows: a)-c) 35; d) 28; actin, 33. Primers positions are indicated in Figure 3.9.
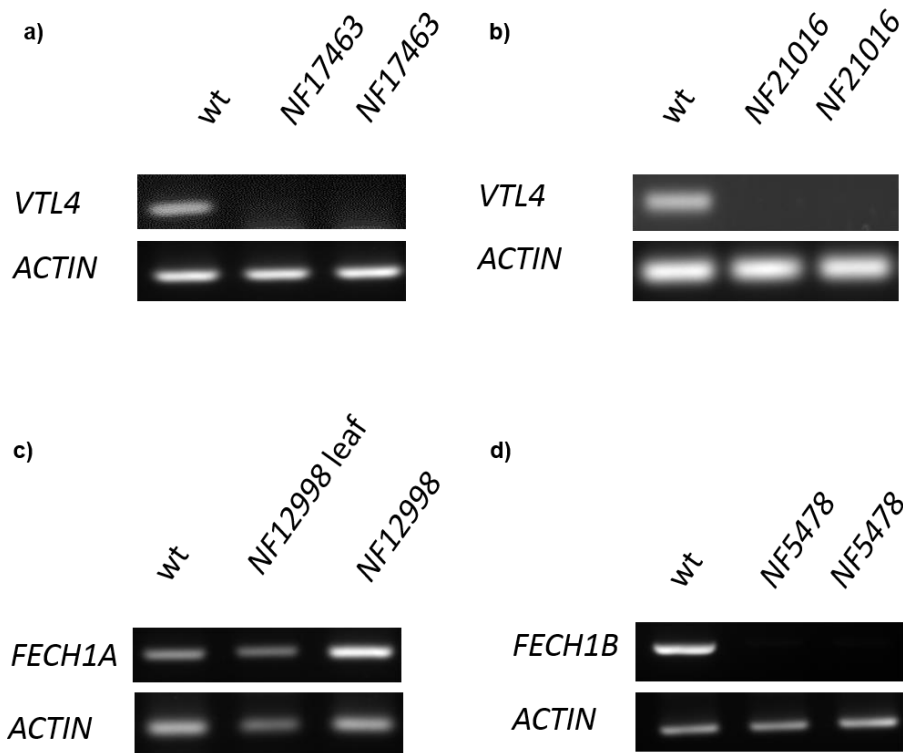
3.7 Discussion

3.7.1 Shortcomings of RNAseq and microarray datasets

As previously mentioned three datasets were analysed to identify genes of interest. Two of these were microarray datasets, and the other RNAseq. It was important to use more than one dataset in this analysis to try and ensure that any candidate genes identified by this approach were actually important in nodulation and not merely a false positive. This is one of the major risks of a reverse genetics approach; identification of potential genes of interest leading to mutants with little or no phenotype. All of the datasets interrogated proved useful in some respect, but it is important to remember the limitations of any study when analysing data.

One of the major drawbacks of microarrays is that there is incomplete coverage of genes; even though probe sets are designed such that as many genes are covered as possible, there are still inevitably some gaps. In addition, these probe sets are often not capable of resolving the differences between members of the same family. This can be due to the probes binding to conserved regions of related transcripts, rather than their intended target, leading to misinformation about the intended gene's expression pattern.

Higher resolution of transcripts is one of the advantages of RNAseq, and it is this property that allows the Roux et al. (2014) dataset to separate out the members of gene families. Although the approach of Roux et al. was elegant in its use of individual nodules as time courses of development, one problem with this was the physical difficulty of precisely sectioning enough nodules to carry out the RNA extractions. In part, this is due to the difficulty in precisely identifying these regions, for example the difference between the proximal and distal zone below the meristem and even the interzone; the boundaries between these zones are not straight or well defined, so assigning individual cells to them can be problematic. Another aspect of the Roux et al. dataset that should be taken into account is their amplification of transcripts; even using random primers is likely to introduce some bias into the data.

3.7.2 The role of ferritins in nodule development

Both *FER1A* and *FER2* are expressed in both nodule and other plant tissues, so are not nodule-specific. No further study of FER genes was completed as their importance in nodulation or plant development as a whole would be unclear. In plants, the role of ferritins is in preventing the toxic redox reactions associated with high cytosolic Fe by sequestration (Ravet et al., 2009). This process seems to be very important in the context of early nodule development, when ferritin is upregulated, probably to buffer the impact of the rapid Fe influx until the haem and nitrogenase cofactor machinery is

sufficiently upregulated to mobilise this as a source. To examine the effects of lack of ferritin on nodule development it would have been necessary to cross *fer1a* and *fer2* mutant lines. This could have led (after a great deal of work) to a double mutant with no phenotype as the remaining Medicago ferritins would complement. Alternatively it could have generated a double mutant with a large number of phenotypes unrelated to nodulation as the plants would no longer be able to cope as effectively with perturbations in cellular Fe levels. In addition, ferritins have already been studied in nodulation, and their localisation and regulation discussed (Strozycki et al., 2007; Lucas et al., 1998).

3.7.3 Presence of *FECH1A* transcript in the NF12998 mutant line

Despite the *Tnt1* insertion in line NF12998 being located only 6 base pairs into an intron, it seemingly had no effect on transcript levels. I had hoped that its proximity to the splice site between the first and second exons would disrupt the splicing process and result in a frameshift mutation, leading to an early stop codon and production of aberrant, non-functional transcript, or lead to degradation. Sequencing of the band from the *FECH1A* RT-PCR showed that the transcript was completely normal, showing that Medicago has a remarkable ability to remove unwanted features from intronic DNA at no cost to the plant.

3.7.4 Conclusions

From the microarray and RNAseq datasets two genes encoding proteins in the VTL family (*VTL4* and *SEN1*) were found to be strongly and specifically induced in nodules. From the RNAseq dataset alone, two Type I ferrochelatases (*FECH1A* and *FECH1B*) were found to be similarly specifically upregulated, albeit to a lesser extent than the two *VTL* genes. *Tnt1* insertion mutants were identified and isolated for *VTL4* (NF17463 and NF21016), *FECH1B* (NF5478), *FECH1A* (NF12998), *FER1A* (NF8502) and *FER2* (NF18993). The *vtl4* and *fech1b* mutants were taken forward for further study after transcript abolition was confirmed by RT-PCR. No *Tnt1* insertion lines were available in the *SEN1* gene, however the 13U mutant contains a large (~30 kb) deletion, which encompasses both the *SEN1* and *VTL4* genes, enabling the study of the role of these genes in tandem. Phenotypic characterisation of 13U, along with the *Tnt1* insertion lines described previously, will form the basis of the next two chapters.

4 Haem cofactor assembly in nodules

4.1 Introduction

As previously discussed, leghaemoglobin is present at exceptionally high concentrations in nodules (up to 3 mM) (Bergersen and Appleby, 1981). It was originally hypothesised that the haem cofactor was synthesised by the bacteria, until this was shown not to be the case by Santana et al., who measured the transcript, protein and enzyme activity levels of four tetrapyrrole synthesis genes in soybean and pea nodules, and found them to be upregulated during nodule development (Santana et al., 1998). My analysis of the Roux et al. RNAseq dataset (Chapter 3) found that "Type 1" ferrochelatases are strongly upregulated in the interzone and nitrogen fixation zone of Medicago nodules (Figure 3.4b). The insertion of Fe into protoporphyrin IX is the only step of haem biosynthesis that is specific to the haem pathway (Tanaka and Tanaka, 2007), as all the other steps are common among the tetrapyrroles. However, as nodules are underground organs and consequently have no need for the photosynthetic and phytochrome pigments chlorophyll and photochromibilin, the majority of tetrapyrrole precursors will be channelled either into haem or sirohaem synthesis. Expression of tetrapyrrole synthesis genes in nodules was extracted from the Roux et al. dataset and is shown incorporated into the pathway in Figure 4.1. In general, one gene from each stage of tetrapyrrole biosynthesis leading to the haem branch is upregulated in nodules, as discussed in Section 3.3. Most tetrapyrrole biosynthesis gene expression shows maximal expression in the interzone, with high expression in the nitrogen fixation zone, as seen for *FLU*, *GluTR, PBGD, UROS,* one of the *UROD* orthologues, *CPOX* and *FECH*. This fits with UPM, the branchpoint for sirohaem biosynthesis, is maximally expressed in the nitrogen fixation zone. This could be as sirohaem is a cofactor in both nitrite and sulphite reductase and its synthesis is likely upregulated to meet the increased demand for sulphur from the symbionts.

Porphobilinogen (PBG)

PBGD

Hydroxymethylbilane

UROS

Uroporphyrinogen III

UROD

Coproporphyrinogen III

UPM

Oxidase

SirB

Sirohaem

PBGD — Medtr4g076600

UROS — Medtr3g101350, Medtr1g052525

UROD — Medtr3g072330, Medtr2g076340

UPM — Medtr7g034345, Medtr3g026780

SirB — Medtr6g005390

Coproporphyrinogen III

CPOX

Protoporphyrinogen IX

PPOX

Protoporphyrin IX

FeCh

Protohaem

HO

MgCh

Mg-protoporphyrin IX

Chlorophyll *a*

Chlorophyll *b*

Biliverdin IXα

CPOX

M Di Pr IZ N2
■ Medtr5g098800 ■ Medtr3g075110

PPOX

M Di Pr IZ N2
■ Medtr1g085730 ■ Medtr7g031310

FeCh

M Di Pr IZ N2
■ Medtr3g104260 ■ Medtr4g019850
■ Medtr8g095320 ■ Medtr3g097290
■ Medtr4g087690 ■ Medtr8g095320

MgCh

M Di Pr IZ N2
■ Medtr4g087895
■ Mtr.37529.1.S1_at

HO

M Di Pr IZ N2
■ Medtr8g019320 ■ Medtr1g019970

Figure 4.1 Expression patterns of tetrapyrrole biosynthesis genes in the context of the pathway as a whole

(Data taken from Roux et al. 2014). For all graphs, the y-axis is total reads detected in the RNAseq, and the nodule zones represented as follows: M- meristem; Di- distal zone below meristem; Pr- proximal zone below meristem; IZ- interzone; N2- nitrogen-fixation zone. Gene abbreviations are: GluTR- glutamyl tRNA synthetase; GSAT- GSA aminotransferase; ALAD- ALA dehydratase; PBGD- PBG deaminase; UROS- uroporphyrinogen III synthase; UPM- uroporphyrinogen III methyltransferase; SIRB- sirohydrochlorin ferrochelatase; UROD- uroporphyrinogen III decarboxylase; CPOX- coproporphyrinogen III oxidase; PPOX- protoporphyrinogen IX oxidase; FeCh- ferrochelatase; HO- haem oxygenase; MgCh- magnesium chelatase. Structures drawn using ChemDraw software.

4.2 Aims and objectives

This chapter aims to identify the role of *FECH1B* in haem biosynthesis in the Medicago-Sinorhizobium symbiosis. Although both *FECH1* family genes are upregulated in nodules, *FECH1B* is the more highly expressed orthologue. In addition, the *Tnt1* insertion in *FECH1A* in line NF12998 is spliced out, leaving the transcript undisrupted, and rendering this mutant line unusable. Using *fech1b-1*, a previously identified mutant line (Section 3.5.1), nodule phenotypes will be characterised. The localisation of *FECH1B* gene expression will be characterised using a promoter-GUS fusion. Protein localisation will be characterised *in vivo* using eGFP fusions to begin to determine how nodules are capable of synthesising such high quantities of haem without negative feedback regulation.

4.3 Phenotypic characterisation of *fech1b-1*

4.3.1 Growth phenotype of *fech1b-1*

The *fech1b-1* mutant (as described in Chapter 3) has a *Tnt1* insertion in the second exon, abolishing full-length transcript. Comparison of plant growth of the *fech1b-1* mutant with wild type on Terragreen/sand did not show any obvious differences in the architecture of aerial tissue or roots (Figure 4.2a), or fresh weight of nodulated plants, which was not significantly different from the wild type (Figure 4.2b).

a)

b)
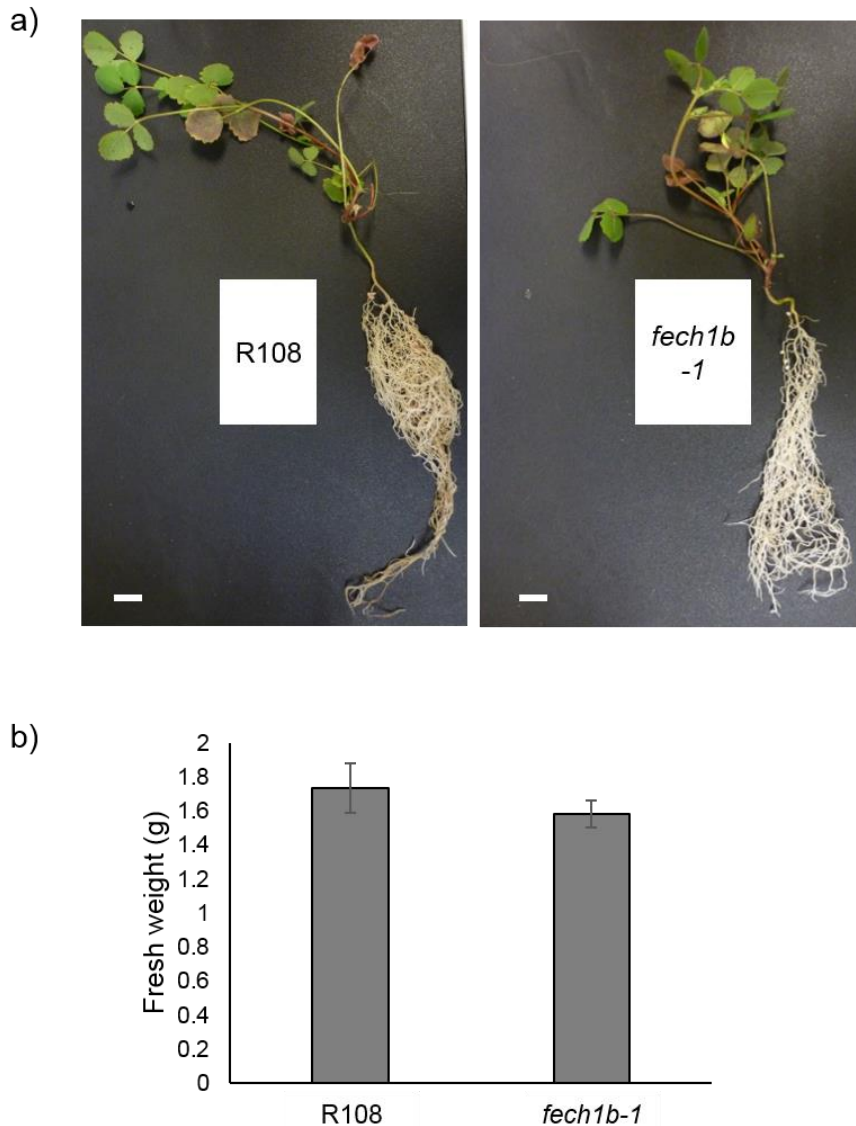
Figure 4.2 Phenotypes of *fech1b-1* mutant plants compared to R108 wild-type control
a) General plant growth at 35 days post inoculation (dpi). Plants were grown on Terragreen/sharp sand and irrigated with water. Representative images chosen from plants phenotyped. Scale bars represent 1 cm. b) Fresh weights of R108 (wild-type) and *fech1b-1* plants. Error bars represent standard error of the mean. Significance was inferred using Student's T-test assuming unequal variances. n=6 plants.

4.3.2 Nodule phenotype of *fech1b-1*

To determine whether lack of FECH1B affects nodule development, the number of pink and white nodules was counted on plants at 35 days post inoculation (dpi). There was no significant difference in the total number of nodules, however there were significantly fewer pink nodules compared to the wild type ($P \leq 0.035$) (Figure 4.3), with white nodules making up the difference. This decrease in the number of pink nodules is consistent with a decrease in levels of functional leghaemoglobin, due to less haem cofactor being produced.
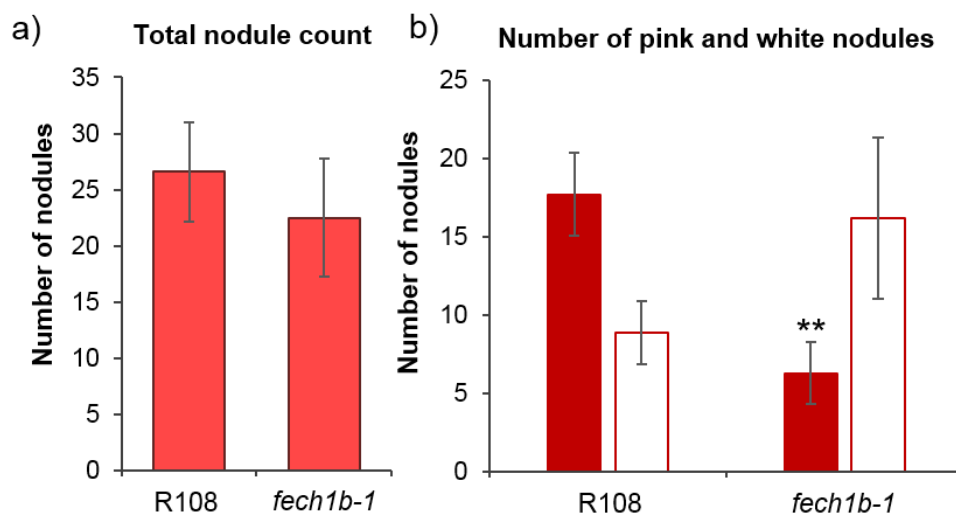
Figure 4.3 Average number of nodules per plant from R108 and *fech1b-1*
Plants grown on Terragreen/sand at 35 dpi. a) Average total number of nodules, b) average number of pink nodules (red bars) and white nodules (white bars). Significance was inferred using Student's T-test assuming unequal variances ** P ≤ 0.035. Error bars represent standard error of the mean, data from 10 plants, representative data shown from 3 biological replicates. (with Hannah Justice)

4.3.2.1 Nodule morphology of *fech1b-1* mutants

The higher proportion of white nodules on *fech1b-1* plants compared to the R108 wild type suggests that infection of nodule cells and/or bacteroids' development is impaired. To establish the extent of nodule colonisation in *fech1b-1*, three staining methods were used to detect (i) bacterial β-galactosidase expression; (ii) accumulation of plant starch; (iii) DNA.

S. meliloti 1021 expressing β-galactosidase were used to infect R108 and *fech1b-1* roots. 14 dpi fixed nodules were then stained using X-gal (5-bromo-4chloro-3-indolyl-β-D-galactopyranoside, a chromogenic substrate of β-galactosidase), before sectioning. Cleavage of X-gal by β-galactosidase produces 5,5'-dibromo-4,4'-dichloro-indigo, an intense blue precipitate. In R108 (wild type) nodules bacteria can be seen throughout the mature nodule, with some early infection visible in the late meristem (Figure 4.4a). In pink nodules from the *fech1b-1* mutant X-gal staining can also be seen throughout the nodule, from the infection zone down to the nitrogen fixation zone, albeit with some incomplete colonisation in the core of the nitrogen fixation zone. White *fech1b-1* nodules show X-gal staining from the infection zone to the basal zone of the nodule, similar in morphology to immature wild-type nodules, suggesting that the white nodule phenotype might just be due to retarded nodule development in *fech1b-1* plants.

Large amounts of sugars are transported to nodules to provide the carbon source for the symbionts. In infected cells this is thought to be converted into dicarboxylic acids such as malate, for import by the bacteroids (Udvardi and Poole, 2013). In uninfected nodule cells this sugar is instead stored as starch. To investigate whether starch is deposited in *fech1b-1* nodules, sections were stained with Lugol's solution, allowing identification of uninfected cells. Lugol's staining is more intense in both the infection and late nitrogen-fixing zones of *fech1b-1* nodules than in R108 (Figure 4.4 b). This suggests either poorer colonisation, or a lower respiratory activity of the infecting bacteria, therefore requiring less sugar to function. As nitrogen fixation is the most energetically demanding process in the nodule (requiring 16 molecules of ATP to reduce one molecule of dinitrogen to ammonia), this could imply lower nitrogen fixation efficiency.
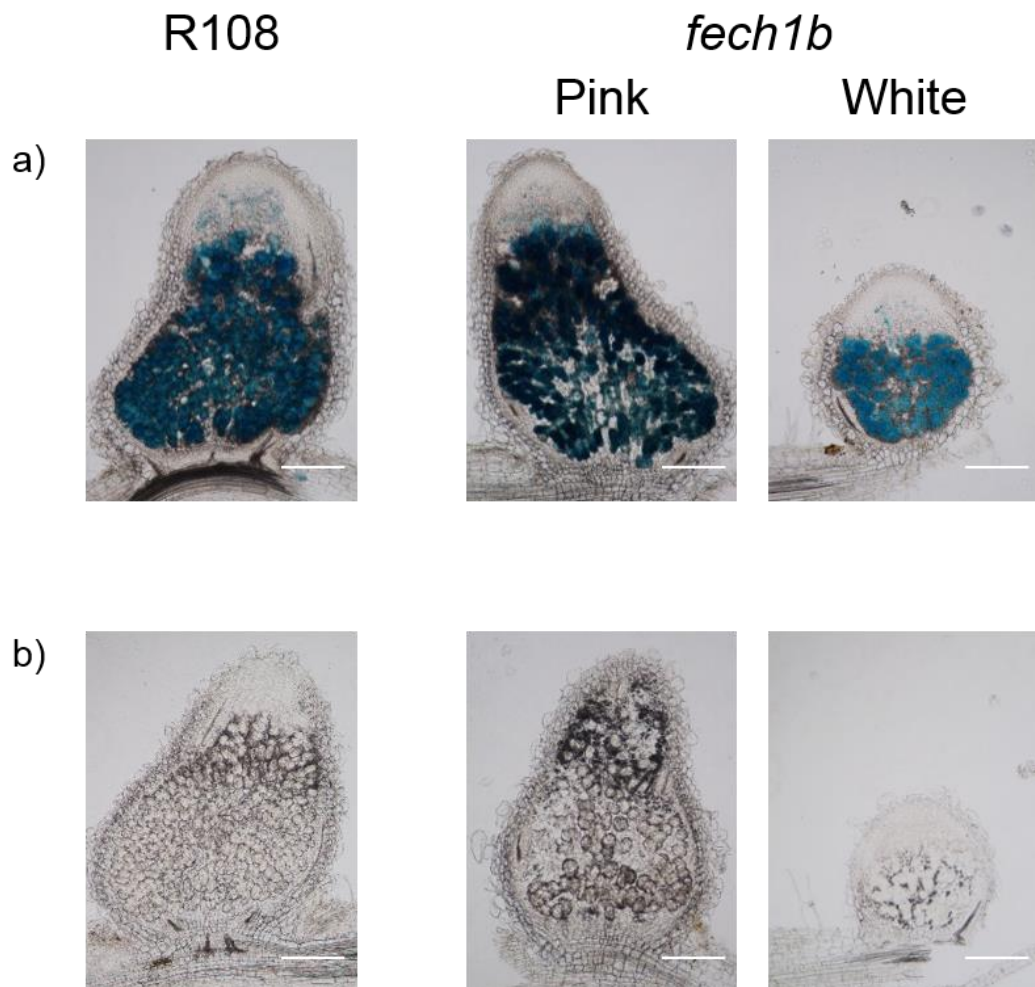
Figure 4.4 Morphology of nodules of R108 and *fech1b-1* plants grown on zeolite, irrigated with Gibson's trace media and infected with *Sm1021*
a) Nodules colonized by *lacZ*-expressing bacteria, fixed and stained with X-gal before embedding in agarose and sectioning to 75 μm thickness using a vibratome. b) Lugol's stain for presence of starch granules in nodule sections prepared as described above, omitting the X-gal staining.

Syto^{TM} 13 is a green fluorescent dye that binds to DNA, allowing higher resolution tracking of bacterial colonisation, and differentiation. Syto^{TM} 13 staining of R108 and *fech1b-1* nodule sections is shown in Figure 4.5. In R108, signal from the bacteria first appears in the infection zone, and is highest in the differentiation zone, where bacteria elongate and differentiate into bacteroids. Signal in the nitrogen fixation zone is weaker as the organisation of the bacteroids in the infected cells begins to deteriorate.

R108

fech1b

Pink   White

a)

b)

c)

d)
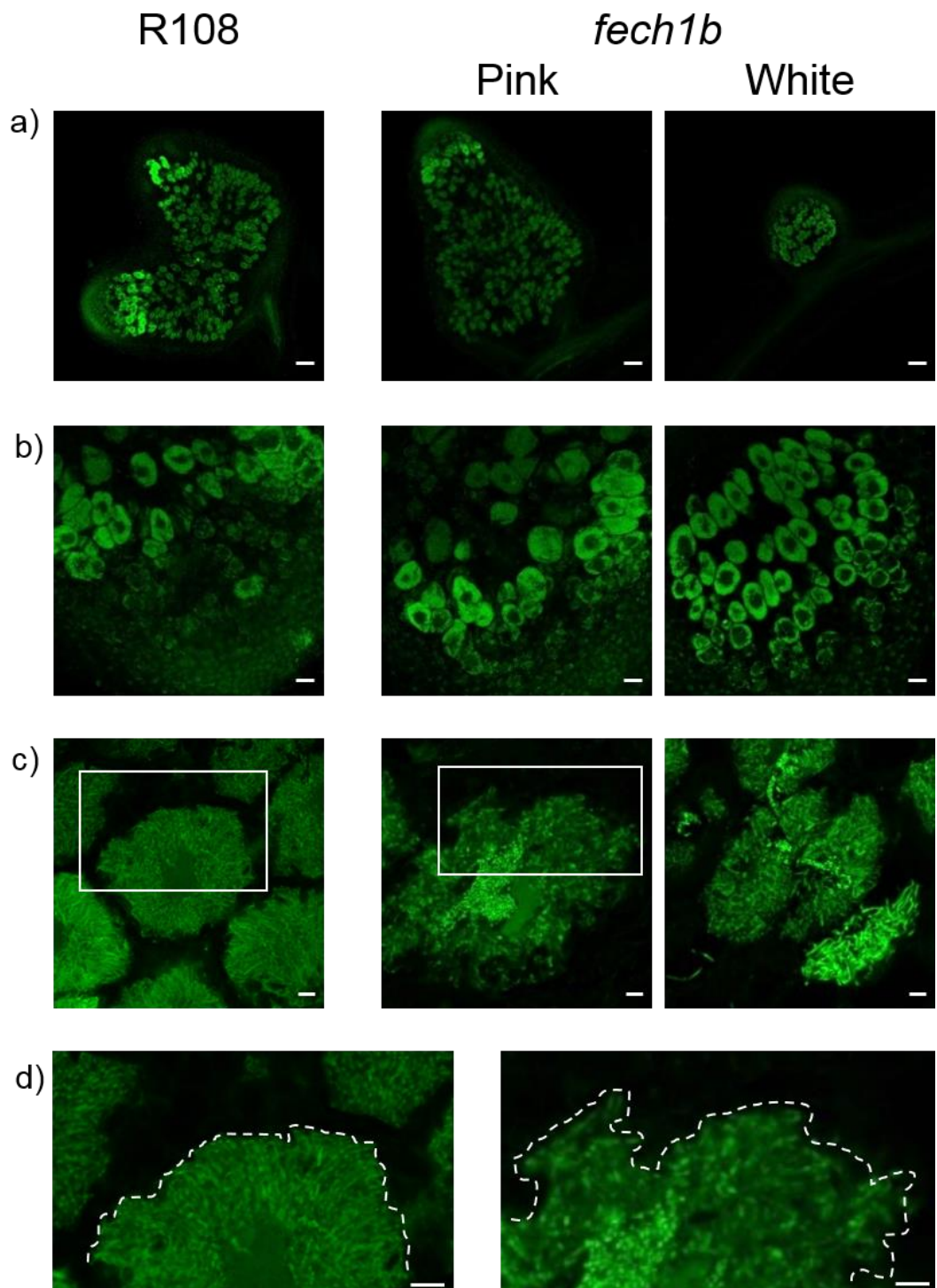
Figure 4.5 Syto™ 13 stained R108 and *fech1b-1* nodule sections
a) Overview of R108 and *fech1b-1* nodule sections using a 10 x objective, scale
bar represents 100 µm b) Higher magnification view of the meristem to the
interzone, 40 x objective, scale bar represents 25 µm c) High magnification
images of cells in the nitrogen fixation zones with insets (d) shown below.
White dashed lines indicate cell boundaries. Images shown in c) and d) were
captured using a 63 x objective, scale bar represents 5 µm. . Nodules were
fixed and embedded in agarose and 75 µm sections cut using a vibratome.
Images taken using Leica TCS SP8 confocal microscope, by Gergely Iski.

Pink *fech1b-1* nodules also show maximal staining in the differentiation zone with weaker fluorescence in the nitrogen fixation zone, similar to R108. White *fech1b-1* nodules show uniform brightness throughout. At 40 x magnification the size of the interzones of both pink *fech1b-1* and R108 nodules are comparable (Figure 4.5b).   It could be that the uniform brightness in white *fech1b-1* nodules is due to the majority of cells being at the differentiation stage (the brightest stage). This would again be consistent with the white nodules just showing a slower growth phenotype, as the interzone is the stage of nodule cell development before the nitrogen fixation zone.

Higher magnification of the nitrogen fixation zones are shown in Figure 4.5c. The most striking difference between the R108 wild-type and *fech1b-1* nodule sections is in this zone. Cell structure is severely disrupted in *fech1b-1*, and the cell boundaries are completely unclear, as highlighted in Figure 4.5d. This suggests that although mature nodules do seemingly develop in this mutant line, at a cellular level there are some abnormalities.

4.4 Tissue-specific expression of *FECH1B*

To examine the pattern of *FECH1B* expression in nodules, 1084 bp upstream of the start of *FECH1B* was synthesised for Golden Gate cloning. Although 2000 bp would have been preferable to be sure of encompassing the entire promoter region, 1084 bp was the distance to the next gene. For Golden Gate cloning, sequences often need "domesticating" to remove any internal restriction sites of Golden Gate enzymes. In this case one base of a *Bbs*I site was altered, destroying the site (Appendix 1). In coding sequences, it is comparatively easy to ensure mutations are silent but in promoters this is much more difficult unless there is a recognised promoter motif in the vicinity. *pFECH1B* was cloned adjacent to a β-glucuronidase (GUS) gene, with introns intact as this was shown to increase expression (Feike, D. personal communication). This was cloned along with the terminator from the octopine synthase gene (T-Ocs) into a backbone vector (EC50506), containing a kanamycin resistance cassette for bacterial selection, next to a constitutively expressed dsRed module as a transformation marker (Figure 4.6).
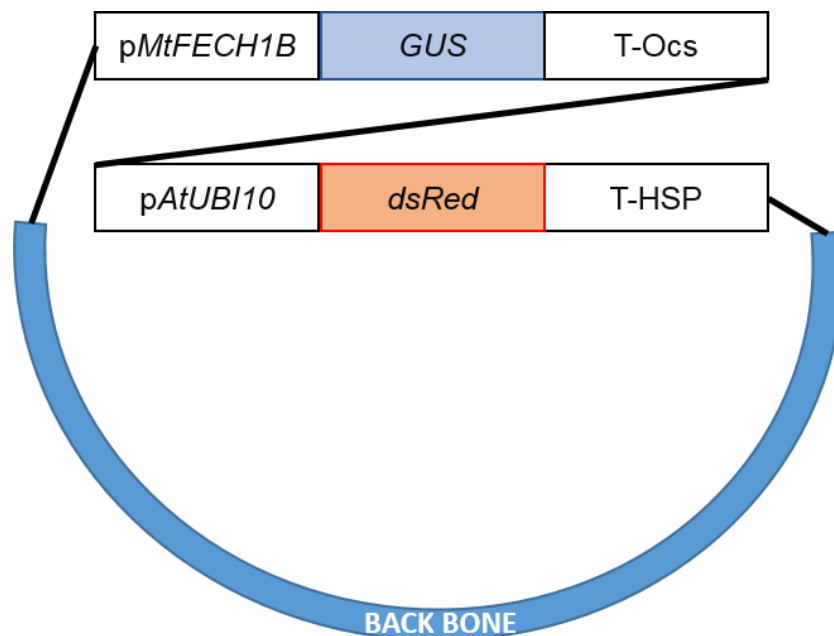
Figure 4.6 Schematic representation of plasmid for determination of *pMtFECH1B* expression pattern in nodules
Containing *GUS* gene under the control of *pMtFECH1B*.

4.4.1 Localisation of *pFECH1B-GUS* activity

Medicago "Jester" seedlings were transformed as described in Section 2.9.4. "Jester" is a commercial seed line, and were used for hairy root transformations for a couple of reasons. Firstly, the seeds do not require scarification, cutting out this step at the seed preparation stage. Secondly, site-wide problems with fungal contamination of developing seeds during bulking up have caused particularly bad problems when growing the plants on plates, so I decided to use a line for these experiments that had been bulked up elsewhere and so was less likely to be contaminated.

At 28 dpi plants were screened for the presence of the dsRed transformation marker. Nodules with good fluorescence were harvested and stained for GUS activity using X-Gluc (5-bromo-4-chloro-indolyl glucuronide), fixed and sectioned. GUS staining was localised to the nodule meristem, indicating activity of the *FECH1B* promoter in that zone (Figure 4.7, left panel). Interestingly this is different from the expression pattern reported in Roux et al. (2014), where *FECH1B* expression peaks in the interzone. The images used were representative of all nodules sectioned (~12, taken from multiple plants). Untransformed nodules were subjected to the same staining, fixing and sectioning treatment but no staining was seen in these (Figure 4.7, right panel).
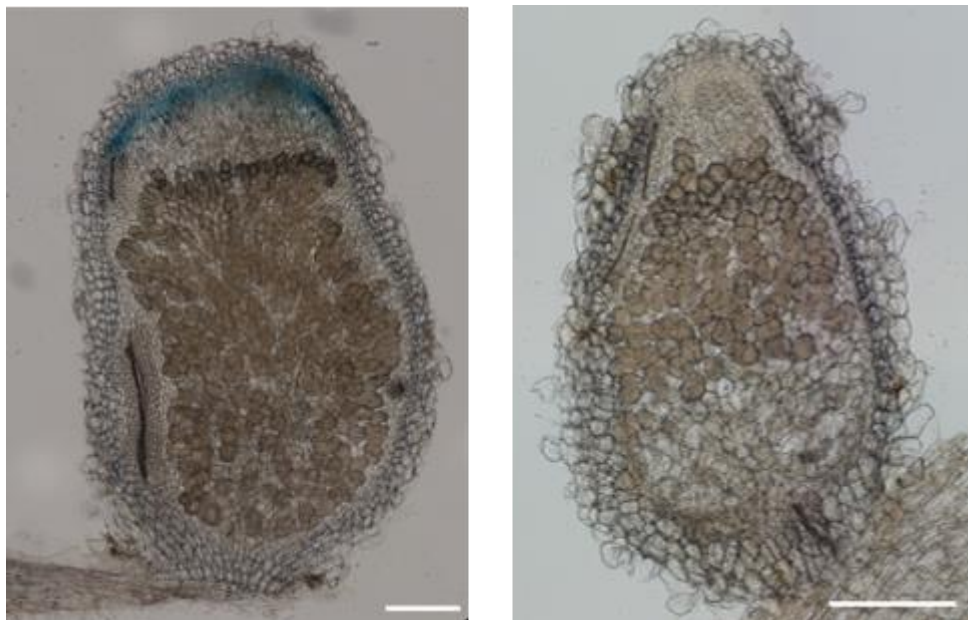
Figure 4.7 Localisation of *pFECH1B* expression by promoter-*GUS* fusion.
28 dpi nodules from plants grown on Terragreen/sand were harvested and immediately stained for GUS activity, before fixation and embedding in agarose. 75 µm sections were cut using a vibratome and GUS staining visualised using a Leica DM6000 microscope Left panel is nodule from a root transformed with *pFECH1B-GUS*, right panel is a nodule from an untransformed root. Scale bar represents 0.25 mm.

4.5 Determining the subcellular localisation of FECH1B protein using an eGFP fusion

To examine FECH1B protein localisation the coding sequence (1413 bp) was domesticated for Golden Gate cloning by the silent mutation of a *Bsa*I site and synthesised. The synthetic *FECH1B* gene was cloned downstream of the constitutive Lotus ubiquitin promoter (*pLjUBI*); and upstream of an in-frame *eGFP* gene, connected by a Gly-Gly linker. The octopine synthase terminator was used (as described in Section 4.4). This module was cloned into the EC50506 kanamycin-resistant backbone alongside a plastid marker-mCherry module, containing the small subunit of rubisco fused to mCherry, under the control of *pLjUBI* and the Acetyl CoA-synthetase 2 (Acs2) terminator. There was also a constitutively expressed, nuclear-localised cyPET module in case a further transformation marker was required (Figure 4.8a).

A control plasmid was constructed using the same components, with the exception that eGFP was constitutively expressed unbound to another protein, and the marker used was ScCOX4, a mitochondrial marker (Figure 4.8b)
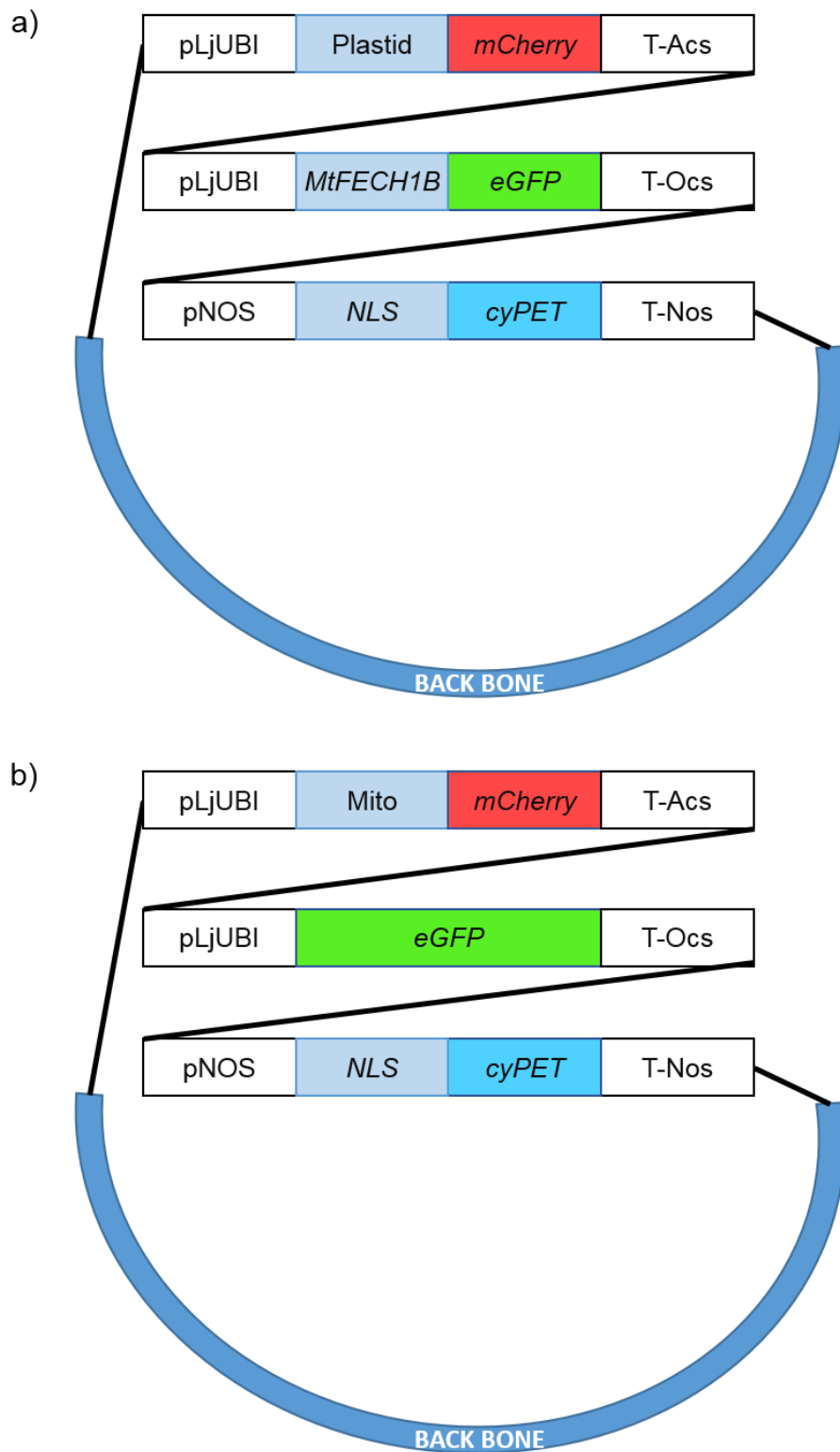
Figure 4.8 Schematic representation of plasmids for localisation of FECH1B a) plastid-localised mCherry and *FECH1B-eGFP*, b) mitochondrially-localised *mCherry* and the free *eGFP*.

4.5.1 FECH1B-eGFP is localised to the plastids

Medicago "Jester" seedlings were transformed as described in Section 2.9.4. Four weeks after transformation, plates were inoculated with *Sm1021* and grown for a further 3-4 weeks. Plates were screened for roots expressing eGFP and transformed roots were imaged using a confocal microscope. Roots were visualised instead of nodules as FECH1B is expressed in both organs and roots are easier to image because they are smaller and less dense.

| MtFECH1B-eGFP | Plastid-mcherry | Merge |

a)

b)

c)

d)

| Free eGFP | Mito-mcherry | Merge |

e)
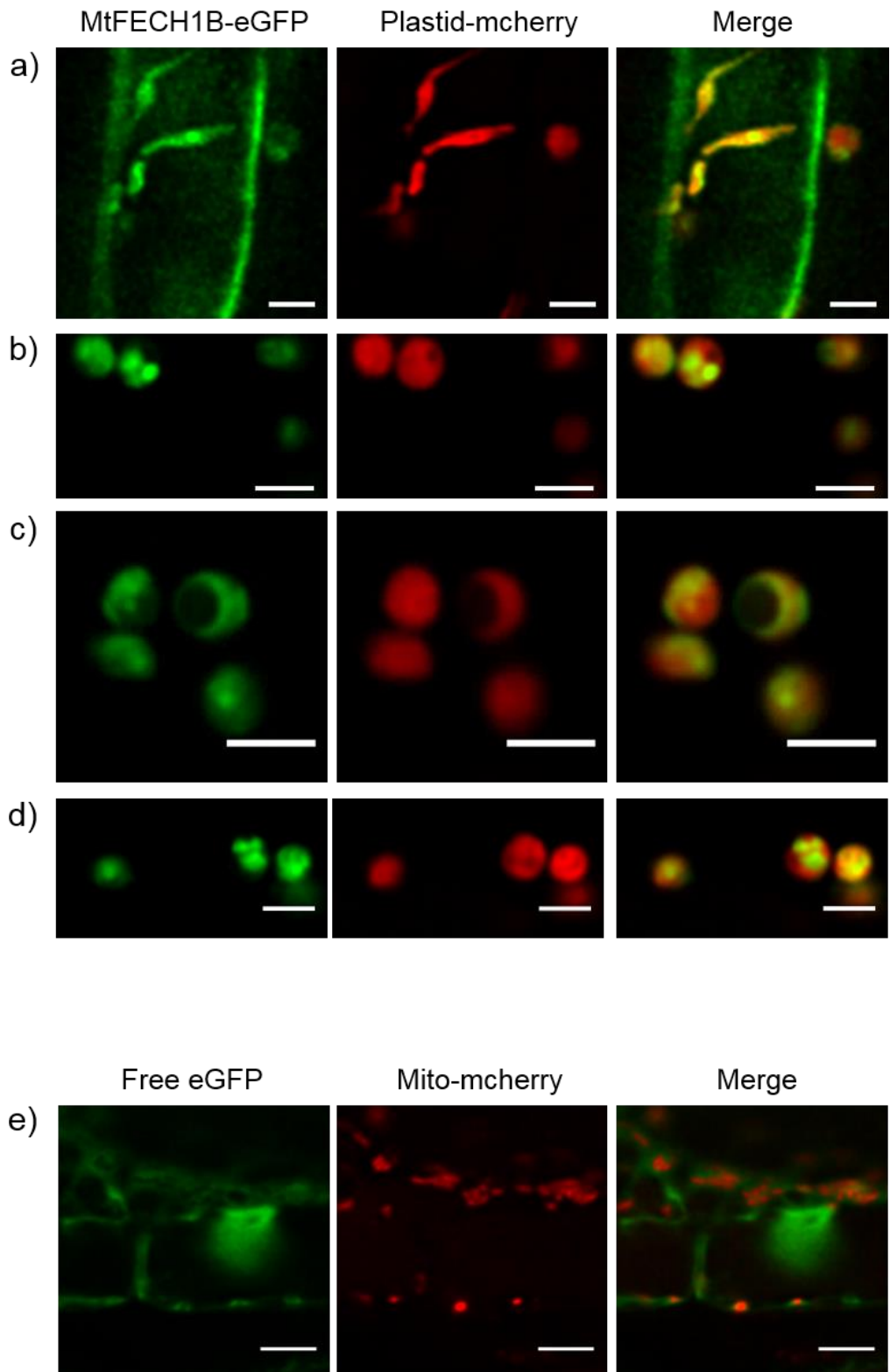
Figure 4.9 Localisation of FECH1B-eGFP to the plastids
Confocal microscope images of Medicago root expressing a-d) FECH1B-eGFP (left panel), plastid-localised mCherry (middle panel) and overlay of the two (right panel). e) Free-eGFP (left panel), mitochondrially localised mCherry (middle panel) and an overlay of the two. Images captured on Leica SP5 microscope with a 63x water objective, and deconvolved using Huygens software. Scale bar represents 3 µm in a)-d) and 5 µm in e).

FECH1B-eGFP localisation is shown in the left panel of Figure 4.9a-d, plastid-mcherry in the middle panel, and the overlay in the right panel. From this, FECH1B localises to the plastids. This can be seen from the signal overlap with the plastidial marker (Figure 4.9a-d, right panel). There does look to be some eGFP signal in the cytosol of Figure 4.9a, which could be due to the construct being expressed at very high levels, as it is driven by the Lotus ubiquitin promoter. Panel a) also appears to show elongated plastids, that resemble stromules; tubules that extend from plastids. Cross-sections of plastids show FECH1B-eGFP localised to distinct cores in central parts of the plastid, however due to the low signal to noise ratio it was impossible to resolve the localisation further (Figure 4.9b-d).

Free eGFP localises to the cytosol and nucleus (Figure 4.9e, left panel) and does not overlap with the mitochondrially targeted mCherry (Figure 4.9e, middle and right panels). It is possible to see the gaps in the cytosolic signal where the mitochondria are located. By comparing the pattern of FECH1B-eGFP with the mitochondrial marker, FECH1B was not localised to the mitochondria. This was also evident from the image stack, as the mitochondria moved rapidly in the live tissue, and there was no evidence of moving structures of an mitochondrially-appropriate size in the FECH1B-GFP stacks.

4.6 Discussion

The aim of this chapter was to investigate the role of *FECH1B* in the Medicago-Sinorhizobia symbiosis by characterising the *fech1b-1* mutant, investigating the expression pattern of *FECH1B* by promoter-GUS fusion and determining the localisation of FECH1B by confocal microscopy.

4.6.1 Nodule phenotype of *fech1b-1* mutants

Leghaemoglobin is essential for generating and maintaining the microaerobic environment unique to nodules and, by controlling the oxygen levels, controlling expression of rhizobial *Nif* genes (Ratet et al., 1989; Dixon and Kahn, 2004). It follows therefore, that the synthesis of the haem cofactor should be an essential process, without which nodules either will not form, or more likely will be severely depleted in leghaemoglobin, and non-functional. The presence of nodules suggests that deletion of *FECH1B* alone is not sufficient to stop nodule initiation or development. This is not entirely surprising as *FECH1A* is also expressed in nodules, seemingly functioning redundantly with *FECH1B*. This could be easily confirmed if a *FECH1A* mutant became available, by crossing the *FECH1A* and *FECH1B* mutants and examining this nodule phenotype. This was my intention but unfortunately the only mutant I could obtain in *FECH1A* was not usable as *FECH1A* transcript levels were the same as in the wild type. This could also be achieved using RNAi to suppress *FECH1A* in the *fech1b-1* mutant, which would have provided interesting data if time had allowed.

The *fech1b-1* mutation resulted in plants that produced fewer mature pink nodules than the wild type at 35 days post inoculation, a phenotype I hypothesise is due to delayed nodule development. This could be tested by carrying out a time course of nodule counts, for example at 21, 28, 35 and 42 dpi. I would expect to see the difference between wild-type and *fech1b* nodules decreasing at later timepoints, as the mutant nodules "caught up". Delayed nodule development could be due to lower levels of leghaemoglobin, as the remaining nodule ferrochelatase (FECH1A), that is normally expressed at much lower levels than FECH1B, is trying to compensate. The role of leghaemoglobin is in binding oxygen to create the microaerobic environment for nitrogen fixation. I hypothesise that it takes longer for the necessary haem to populate leghaemoglobin to be produced by FECH1A alone, so it is likely that this microaerobic environment will be generated more slowly. The *Nif* genes for nitrogenase cofactor biosynthesis are only expressed at low partial pressures of oxygen (David et al., 1988), so will be expressed at a correspondingly later timepoint. This could be tested by comparing the levels of leghaemoglobin protein in wild-type and *fech1b* nodules at

different timepoints and comparing this to *Nif* gene expression by qRT-PCR. Ott et al. showed that the nodules of leghaemoglobin RNAi lines of *L. japonicus* had higher levels of free oxygen than their wild-type counterparts, and consequently *Nif* genes were not expressed, rendering the nodules unable to fix nitrogen (Ott et al., 2009).

Pink *fech1b-1* and R108 nodule section structure was very similar when visualised using both X-gal and Lugol's staining. This suggests that if and when *fech1b-1* nodules reach maturity, they are developmentally unimpaired. However, the increased proportion of white nodules present on *fech1b-1* roots, combined with the observation that these white nodules were morphologically similar to immature R108 nodules again suggests a slight developmental delay in the formation of mature nodules in this line.

The disruption of cell structure in the nitrogen fixation zone of *fech1b-1* mutants as shown by Syto™ 13 staining could be due to the role of leghaemoglobin in the neutralisation of reactive nitrogen species in infected cells (Navascues et al., 2012). These nitrogen species are known to attack membranes in nodule senescence, leading to increase permeability and eventual breakdown, consistent with the phenotype observed in sections of *fech1b-1* nodules (Herrada et al., 1993).

4.6.2 Expression pattern of *pFECH1B-GUS*

FECH1B was predicted to be maximally expressed in the interzone, with high expression in the late infection zone and nitrogen fixation zone (Roux et al., 2014), however this study found that a promoter-GUS construct was exclusively expressed in the nodule meristem. There could be several explanations for this: As previously discussed (Section 3.7.1) the accurate sectioning of sufficient nodules for RNA extraction/sequencing is incredibly difficult, and it is likely that there was some variation in the cutting between zones, leading to some blurring between the zonal assignments of transcripts. Having said this, it is unlikely that such a drastic shift in the localisation of gene expression could come about from the slight variation between cutting points between nodules.

The next gene was localised only 1084 bases upstream of the start codon of *FECH1B*, and the size of this intergenic region was the limiting factor on the sequence length used for this study. It could be that the single base mutation (A to T) during the domestication process to remove the *Bbs*I site disrupted the promoter sufficiently to alter the expression pattern, however there were no known promoter elements present in this region and this would seem like a cruel twist of fate. This could be checked by cloning the promoter region using conventional restriction-ligation cloning, so removing the need for sequence domestication.

Gene expression is a complex process, controlled not only by the promoter of the gene but also by both up- and down-stream *cis*-acting elements, for example intronic DNA or the 5'/3' UTR, as well as more distant regulatory elements (Deyholos and Sieburth, 2000; Williams et al., 2016). One option would be to clone the genomic DNA encompassing the assumed promoter region and the gene including introns and fusing this product to *GUS*, possibly with the 3' UTR from *FECH1B* after the *GUS*.

4.6.3 Localisation of the FECH1B protein

In contrast to the human ferrochelatase, which is localised to the mitochondria (Medlock et al., 2015), in the *Viridiplantae* (organisms ranging from pea (*Pisum sativum*) and Arabidopsis to the alga *Chlamydomonas reinhardtii*), Type I ferrochelatases have been suggested to localise exclusively to the plastids (Lister et al., 2001; Cornah et al., 2002; van Lis et al., 2005). However some recent work showed that overexpression of Type I ferrochelatase in tobacco resulted in a small amount of ferrochelatase localising to the mitochondria, although it was not clear whether this was an artefact of overexpression (Hey et al., 2016). A study by Masuda et al. found that Type I ferrochelatases localise exclusively to plastids with stromules when exogenously expressed in onion cells (Masuda et al., 2003). The elongated structure of the plastids in the middle panel of Figure 4.9a and the other image stacks captured but not shown here, resemble stromules, which can allow exchange of molecules as large as GFP between neighbouring plastids (Gray et al., 2001). In the context of haem synthesis in nodules this could provide a route by which haem biosynthesis could be compartmentalised between plastids, resulting in suppression of the inherent feedback inhibition one would expect from such a highly upregulated system. The localisation of FECH1B to specific but unknown substructures in the plastids could also contribute to the spatial separation of haem from the earlier stages of the pathway. The plastid localisation of FECH1B in the context of nodules and their extremely upregulated haem biosynthesis pathway perhaps points towards a new form of specialised plastid, tentatively dubbed the haemoplast. More work is necessary for this assignment to be confirmed however it seems likely that the metabolic shifts necessary to effect this gear change in haem synthesis would require a specialised compartment for that purpose.

4.6.4 Conclusions

The upregulation of *FECH1B* in nodules suggests an important role for *FECH1B* in nodule development. Probably due to genetic redundancy with *FECH1A* it was not possible to see profound morphological differences between *fech1b-1* and R108 wild-type nodules. Localisation of FECH1B to the plastids was consistent with the localisation of Type I ferrochelatases in other plant species.

5 Phenotypic characterisation of Vacuolar iron Transporter-Like (VTL) mutants in *M.truncatula*

5.1 Introduction

The Vacuolar Iron Transporter (VIT) family is one of the major classes of iron transporters in plants. These, along with related proteins the VIT-like (VTL) family are described in detail in the following section.

5.1.1 The Vacuolar Iron Transporter family (VIT)

In Arabidopsis seeds, the vacuole is the main cellular Fe repository (Ravet and Pilon, 2013). Iron is loaded into the vacuole by AtVIT1, a protein closely related to the yeast Ccc1 (Ca^{2+}-sensitive cross-complementor 1) protein. AtVIT1 also transports Fe into the developing vasculature of embryos (Kim et al., 2006). Other VIT family members have been characterised as vacuolar Fe importers in species such as *Centaurea cyanus* (cornflower) and *Tulipa gesneriana* (tulip), where they are crucial for blue colour development in petals, and *Triticum aestivum* (bread wheat), where TaVIT2 has successfully been used for biofortification of wheat grains by expressing it in the endosperm (Yoshida and Negishi, 2013; Momonoi et al., 2012; Connorton et al., 2017b).

Recently, an orthologue of AtVIT1 from *Plasmodium falciparum*, the parasite causing malaria, has been characterised. Not only was PfVIT found to be essential for virulence, but functional studies of recombinant PfVIT suggested that VIT family proteins are Fe^{2+}/H^+ antiporters; the first evidence for their mode of action (Slavic et al., 2016; Labarbuta et al., 2017).

5.1.2 The VIT-like family (VTL)

The VTL (Vacuolar iron Transporter-Like) are a family of proteins thought to have a role in Fe homeostasis. These differ from VIT in the absence of a predicted cytosolic loop, approximately 40 amino acid residues in length, between the second and third transmembrane helices (Figure 5.1).

Five VTL-encoding genes have been identified in Arabidopsis, three of which (*VTL1, 2* and *5*) are downregulated in Fe deficient conditions (Gollhofer et al., 2011). The expression of *VTL1* is under the control of a WRKY transcription factor WRKY46, which is induced under low Fe conditions, binding to the W-boxes in the promoter of *VTL1*, suppressing transcription (Yan et al., 2016). In Medicago roots, *VTL1, 2* and *3* are all slightly downregulated in response to Fe deficiency, although the absolute expression levels are low, so this must be treated with caution (Rodríguez-Celma et al., 2013).

Promoter-GUS studies have shown *VTL1* expression to be concentrated around the vasculature of both the root and shoot under Fe replete conditions (Gollhofer et al., 2011). *AtVTL1* is the only gene that has been characterised in this way, however analysis of *vtl3* and *vtl5* mutant plants reveals Fe accumulation in both the shoot and root in *vtl3* plants and the root alone in *vtl5* mutants. Subcellular localisation experiments using transient expression of N-terminally tagged GFP-VTL1/VTL2 in onion epidermal cells with revealed that VTL1 is localised to the vacuolar membrane, while VTL2 is partially plasma membrane-localised (Gollhofer et al., 2014).

VTL1, and to a lesser extent *VTL2,* partially complemented the Fe-sensitive phenotype of Δ*ccc1* yeast. *VTL5* was not able to complement Δ*ccc1* (Gollhofer et al., 2014). Isolated vacuoles from this experiment showed increased Fe accumulation in cells expressing *VTL1, VTL2* and *VTL5*, posing the question of how complementation with *VTL5* could result in the accumulation of Fe to high levels in the vacuole while being unable to restore growth of the mutant on Fe replete media. *VTL1,2* and *5* also partially complement the *nramp3/nramp4* double mutant; AtNRAMP3 and NRAMP4 are redundantly functioning vacuolar Fe exporters, the deletion of which results in poor seedling root growth on synthetic media (Lanquar et al., 2005). AtVIT1 functions in an antagonistic manner to NRAMP3/4, transporting Fe into the vacuole in high Fe conditions to prevent toxic Fe build-up in the cytosol. *vit1* mutants show decreased seedling root growth on alkaline substrate (Kim et al., 2006). Complementation of *vit1* with *VTL1, 2* or *5* resulted in increased root growth above wild-type levels (Gollhofer et al., 2014). The ability of the VTLs to complement both of these antagonistic transporters implies lack of directionality, possibly due to the loss of the cytosolic loop region.

5.1.3 SEN1

SEN1 is a protein in the VTL family that has been shown to be required for the formation of symbiotically functional, pink nodules in *Lotus japonicus* (Hakoyama et al., 2012). Although it has been hypothesised to be a symbiosome membrane-localised Fe transporter this has not been proven. It is known that SEN1 is needed for bacteria to differentiate into bacteroids; Lotus *sen1* nodules were found to undergo early senescence, hypothesised to be caused by the activation of plant defence responses (Suganuma et al., 2003). In soybean the closest homologue is GmNOD21, which is known to be nodule-specific but has not been characterised. The Medicago homologue of *SEN1* (*Medtr4g094335*) remains uncharacterised.
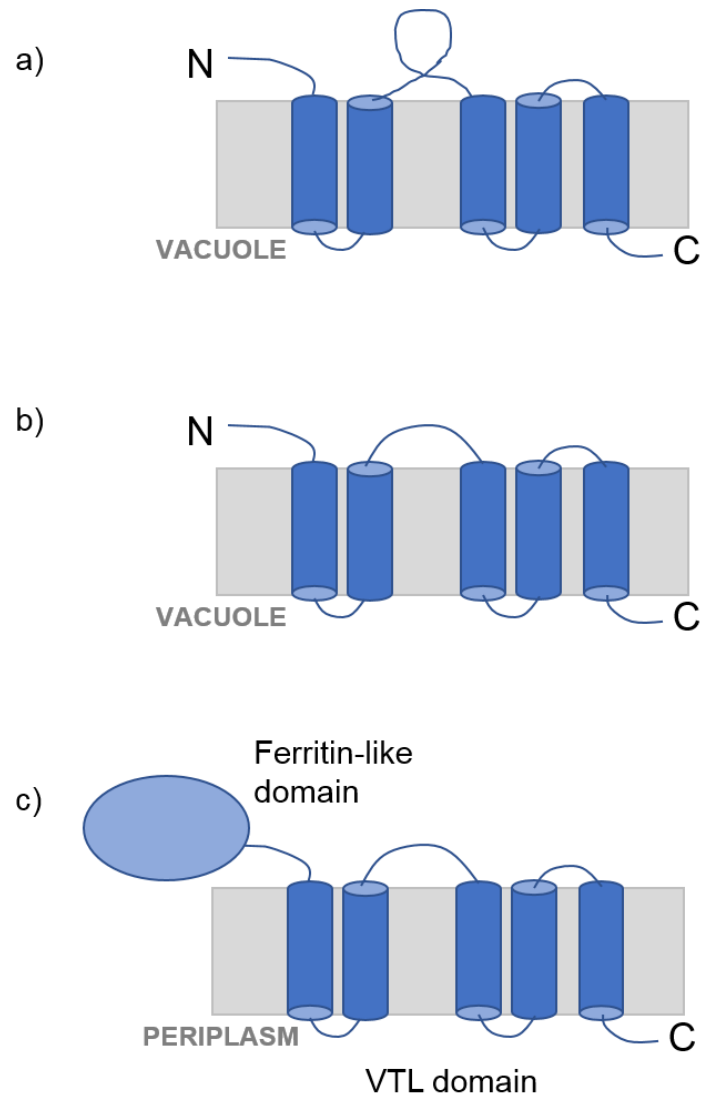
Figure 5.1 Predicted transmembrane domain structure of VIT proteins
a) VIT proteins, b) VTL proteins, and c) MbfA showing the truncated loop between transmembrane helices 2 and 3 in VTLs and the N-terminal ferritin-like domain in MbfA. Transmembrane predictions were the consensus of all five Arabidopsis proteins from Aramemnon version 8.1 (http://aramemnon.uni-koeln.de/index.ep)

5.1.4 MbfA

Membrane-bound ferritin A (MbfA) is a bacterial iron exporter localised to the inner membrane, that is highly conserved among the Rhizobiaceae, including *Sinorhizobium meliloti* (Sankari and O'Brian, 2014; Bhubhanil et al., 2014). It is proposed to be a homodimer consisting of two domains; an N-terminal ferritin-like domain located in the cytosol and a Ccc1-like transmembrane domain, that is actually more similar to the VTL family of proteins, as it is missing the characteristic cytosolic loop (Figure 5.1c, Sankari and O'Brian, 2014; Bhubhanil et al., 2014). In *Agrobacterium tumefaciens*, Δ*mbfa* mutants are more sensitive to high levels of iron and H_2O_2 stress than the wild type (Ruangkiattikul et al., 2012). *mbfA* expression is induced in response to iron; the *mbfA* promoter contains an Iron Control Element, or ICE box, which is the target of the Iron Responsive Regulator (Irr) (Singleton et al., 2010). When bound to the ICE box, Irr represses transcription, but under high iron conditions it binds haem, resulting in its degradation and subsequent relief of repression (Singleton et al., 2010). Thus, when iron levels are high, *mbfA* is expressed, resulting in iron export and subsequent relief of the high iron stress. Although it has been extensively characterised in a number of species of α-proteobacteria (Sankari & O'Brian, 2014; Bhubhanil et al., 2014; Ruankiattikul et al., 2012), the role of MbfA in symbiosis is unknown.

5.2 Aims and objectives

This chapter aims to study the roles of VTL proteins, previously identified by a bioinformatics approach, that are proposed to have an important role in Fe delivery to bacteroids for the synthesis of nitrogenase and other cofactors. The roles of host plant proteins MtVTL4 and MtSEN1 and bacterial protein SmMbfA will be studied in the context of the Medicago-Sinorhizobium symbiosis. Using the previously identified mutant lines *vtl4-1*, *vtl4-2* and 13U (Chapter 3), nodule phenotypes will be characterised. The 13U mutant will be complemented with constructs containing both *VTL4* and *SEN1* separately and in combination. As 13U is a fix- mutant (*i.e.* unable to fix nitrogen) the complementation will be measured by acetylene reduction assay ; a proxy for nitrogen fixation ability. Protein localisation will be characterised *in vivo* using mCherry fusions to pin down further the role these proteins play in functional nodule development. Iron levels perceived by bacteroids will be assessed in *vtl4* and 13U mutants by inoculation with *Sinorhizobium meliloti* (Sm1021) expressing an iron responsive construct consisting of the promoter of *mbfA* driving the *lux* operon. The effect of inoculation with an *mbfA* mutant strain of Sm1021 (Δ*mbfA*) on nodulation will also be characterised.

5.3 Phenotypic characterisation of *vtl4* and 13U

5.3.1 Growth phenotype of *vtl4* and 13U

As described in Section 3.5, *vtl4-1* has a *Tnt1* insertion 230 bp downstream of the start codon of the gene, in its only exon, and *vtl4-2* has a *Tnt1* insertion at 564 bp, near the end of the coding sequence, which is only 720 bp in total. In both cases, no *VTL4* transcript was detected by RT-PCR (Figure 3.10). The 13U mutation consists of a deletion of approximately 30 kb of *Medicago truncatula* A17 chromosome 4 (as described in Chapter 3), encompassing 6 *VTL* genes and a hairpin inducing-like protein, the function of which is uncharacterised.

The effects of these mutations on plant growth and nodule development were examined as described in Sections 2.5 and 2.11. Images of *vtl4-1* and *vtl4-2* plants show slight growth phenotypes associated with the mutations. In both *vtl4-1* and *vtl4-2* plants there looks to be decreased aerial tissue produced, and a slight increase in the amount of root tissue (Figure 5.2). When the fresh weight of *vtl4* mutants was compared with R108 wild-type plants, they were found to be significantly lower ($P \leq 0.01$ for *vtl4-1* and $P \leq 0.1$ for *vtl4-2*) (Figure 5.3). 13U plants were grown alongside Jemalong A17; the wild-type background for this mutation (Figure 5.4). In general, 13U plants are slightly smaller than Jemalong A17 (Figure 5.4a), the leaves of 13U plants are slightly paler than A17, and the plants altogether less luscious and bushy. This is most clearly seen when the plants are growing side by side, as shown in Figure 5.4b. It is also reflected in the fresh weights of 13U plants compared to A17, which is significantly lower ($P \leq 0.01$) (Figure 5.3). This is likely due to the lack of nitrogen fixation capability of *vtl4* and 13U nodules, as plants engaged in a healthy symbiosis have more nitrogen at their disposal and consequently grow better; one proxy for nitrogen fixation ability is dry weight of aerial tissue.
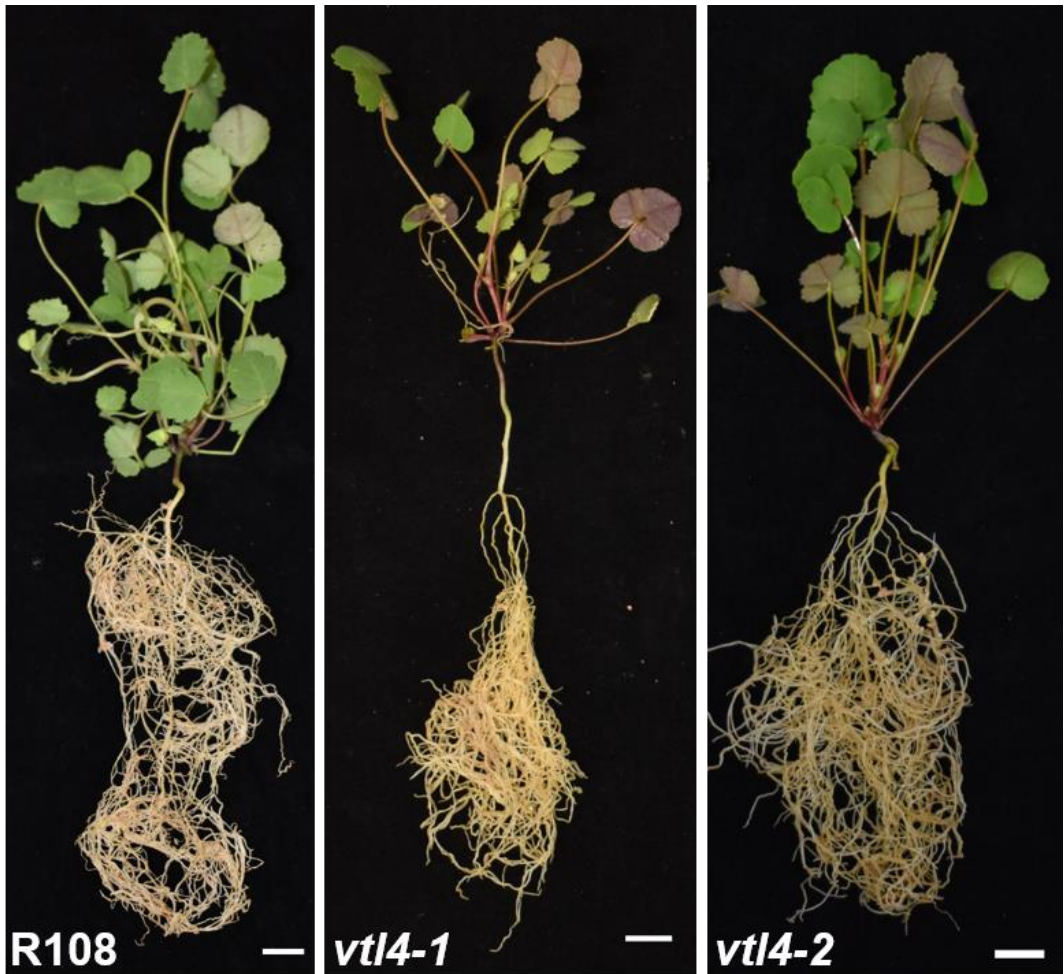
Figure 5.2 Phenotype of *vtl4* mutant plants compared to R108 wild-type
Plant growth at 35 dpi, representative images chosen from 6-8 plants, grown on Terragreen/sand and irrigated with water. Scale bar represents 1 cm.
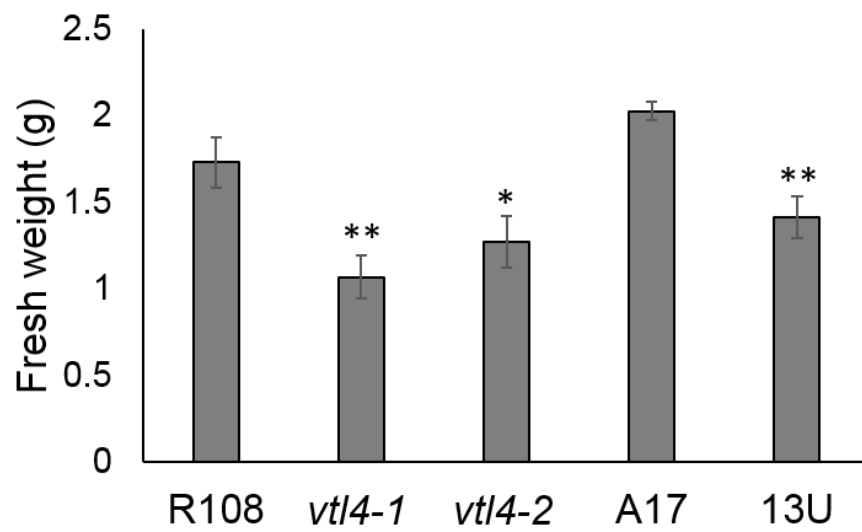
Figure 5.3 Fresh weights of whole *vtl4* and 13U mutant plants with their corresponding wild types. Error bars represent standard error of the mean. Significance was inferred using Student's T-test assuming unequal variances, * P ≤ 0.1, P ≤ 0.01, *** P ≤ 0.001, n = 6 plants.**

Figure 5.4 Phenotype of 13U mutant plants compared to Jemalong A17 wild-type
a) Plant growth at 28 dpi, representative images chosen from 10 plants. Scale bar represents 1 cm. b) Plant growth on Terragreen/sand at 35 dpi

5.3.2 Nodule phenotype of *vtl4* and 13U mutants

To determine the effects of the *vtl4* and 13U mutations on nodule development, the number of pink and white nodules was counted on plants at 35 dpi as described in Section 4.3.2.

Compared to the R108 wild type, *vtl4-1* plants showed no significant increase in total nodules, while *vtl4-2* shows a slight increase in the total number of nodules produced (P≤ 0.1) (Figure 5.5a). Significantly more white nodules were produced by both *vtl4-1* and *vtl4-2* (P ≤ 0.1 and P ≤ 0.01 respectively). There was no significant difference in the number of pink nodules formed by either *vtl4* line compared to the wild type.

13U mutants produce only white nodules (Figure 5.5b). The total number of nodules is significantly greater than the Jemalong A17 wild type (P ≤ 0.001), both the complete lack of pink nodules and the drastic increase in the number of white nodules are statistically significant (P ≤ 0.01 and P ≤ 0.001 respectively).
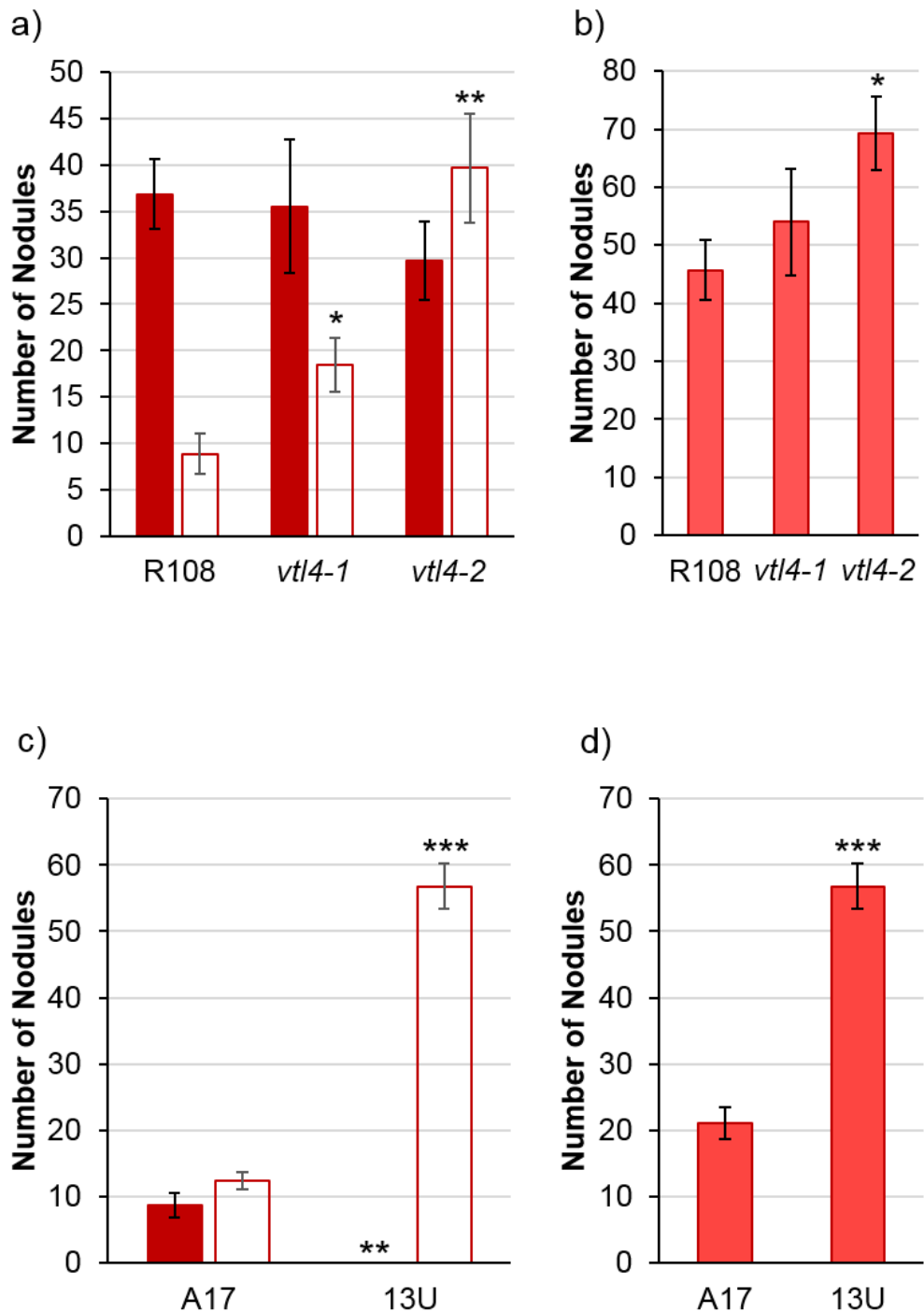
Figure 5.5 Average number of nodules per plant from R108, *vtl4*, A17 and 13U plants Grown on Terragreen/sand and nodules counted at 35 dpi .a) R108 and vtl4 mutant plants, b) A17 and 13U plants. Red bars represent pink nodules, white bars represent white nodules. Error bars represent standard error of the mean. Significance was inferred using Student's T-test assuming unequal variances, * P ≤ 0.1, P ≤ 0.01, *** P ≤ 0.001, n = 6 plants for a) and 10 plants for b). Significant changes in total nodules are denoted by asterisks above the white bar, significant changes in white nodules are denoted by asterisks inside the white bars at the top, and significant changes in the number of pink nodules are denoted by asterisks immediately above the red bars.**

5.3.3 Nodule morphology of *vtl4* and 13U mutants

To assess nodule colonisation and bacterial morphology, three methods were used; X-gal staining, Lugol's staining, and Syto™ 13 staining. For X-gal staining, plants were inoculated with *Sm1021* constitutively expressing *LacZ* as a reporter. Staining for β-galactosidase activity then reveals the extent of bacterial colonisation of nodule cells. Non-functional nodules accumulate more starch than wild-type nodules, as less carbon is required to feed the bacteroids (Aleman et al., 2010; Suganuma et al., 2003). Lugol's solution was used for staining starch granules to allow an estimation of the extent of colonisation. Syto™ 13 is a fluorescent stain that intercalates with nucleic acids, allowing high resolution interrogation of bacteroid development using confocal microscopy.

LacZ staining of longitudinal sections of R108 (wild-type) nodule showed intense staining across almost all cells of the mature nodule, indicating good colonisation of the nodule after infection (Figure 5.6a). Both pink and white nodules of *vtl4-1* showed slightly less staining with part of the basal section of the white nodule being comparatively unstained. The sections of pink and white nodule appeared to show similar levels of colonisation to each other using this method.

LacZ-mediated X-gal stained sections from the R108 wild type grown alongside *vtl4-2* mutants showed intense blue staining in the differentiation and early nitrogen fixation zones but slightly sparser staining in the more mature nitrogen fixation zone, suggesting incomplete colonisation of nodules when plants were grown under these slightly different conditions (R108 and *vtl4-2* plants were grown in the UK on Terragreen/sand, while R108, *vtl4-1*, A17 and 13U plants were grown in Hungary, on zeolite) (Figure 5.6b). *vtl4-2* sections show sparse staining from the infection zone to the nitrogen fixation zone. The basal part of the nodule is similar to the R108 wild type, however the intense stained band of the differentiation zone is no longer present.

In A17 nodule sections, staining with X-gal was strongest in the infection and differentiation zones, however the nitrogen fixation zone showed sparse colonisation (Figure 5.6c). This differs from the R108 wild type previously described (Figures 5.6a and b) as A17 and *Sm1021* are not ideal symbiotic partners. 13U nodules showed a completely different staining pattern, with some cells in the infection zone colonised. Furthermore, I observed blue staining in infection threads that penetrate as far as the basal zone of the nodules (in fix+ lines this would be described as the nitrogen fixation zone) but with no fully colonised cells beyond the infection zone. This suggests a "stalling" of the infection process, with the initial infection event taking place but the development of functional, nitrogen-fixing bacteroids not following on. From these data

it is not possible to speculate at which stage of bacteroid development infection abortion takes place.
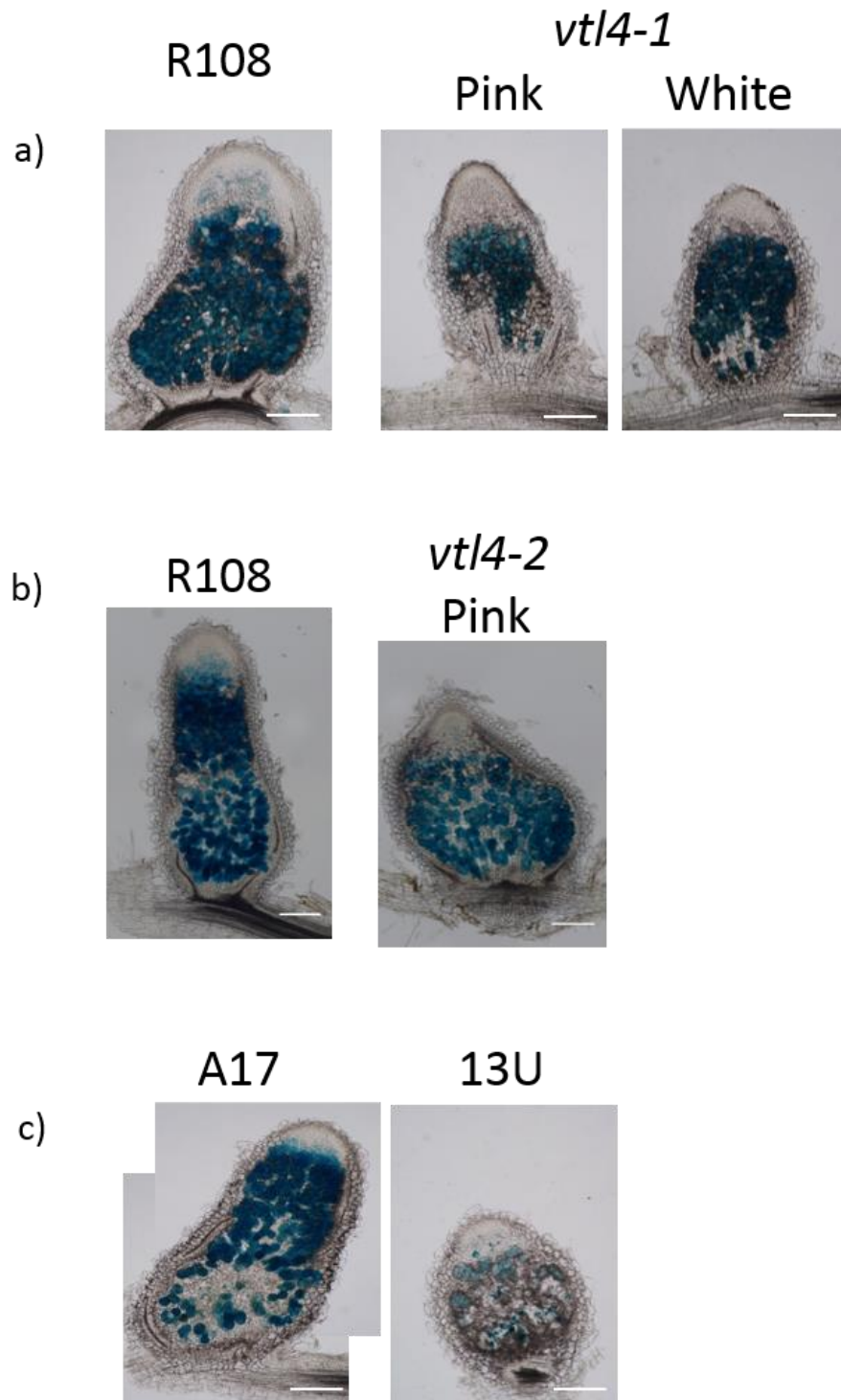
Figure 5.6 Morphology of 14 dpi nodules of R108 *vtl4*, A17 and 13U plants infected with Sm1021 *lacZ* and stained with X-gal.
a) R108 and *vtl4-1*, b) R108 and *vtl4-2*, c) A17 and 13U. Scale bar represents 0.25 mm. Nodules were initiated with *lacZ*-expressing bacteria, fixed and stained with X-gal before embedding in agarose and sectioned to 75 μm thickness using a vibratome.

Lugol's staining of R108 wild-type nodules (Figure 5.7 a and b, left panel) shows no staining in the meristem, a build-up of starch around the infection and differentiation zone of the nodule, and much weaker staining surrounding nodules in the nitrogen fixation zone. This suggests that colonisation increases with nodule maturation as the bacteria proliferate and become more metabolically active. By contrast in the *vtl4-1* mutant pink and white nodule sections, there is intense staining from the infection zone right down to the basal nitrogen fixation zone (Figure 5.7a, centre and right panel). This suggests that the basal (nitrogen fixation) zone of *vtl4-1* nodule sections is less well colonised than the wild type.

In *vtl4-2* sections the more intense Lugol's staining continues around the nodule periphery, right down into the nitrogen fixation zone (Figure 5.7b, right panel). This suggests that cells in the centre of the nodule are better colonised than those nearest the nodule epidermis and vasculature.

As for R108, Lugol's staining of A17 nodule sections is strongest in the infection and differentiation zones, with some cells in the nitrogen fixation zone also showing some staining (Figure 5.7c, left panel). 13U nodule sections show a significant amount of starch from the infection zone to the basal part of the nodule, and particularly around the vasculature and nodule periphery. This is consistent with the X-gal staining (Figure 5.6), suggesting poor colonisation of 13U nodules.
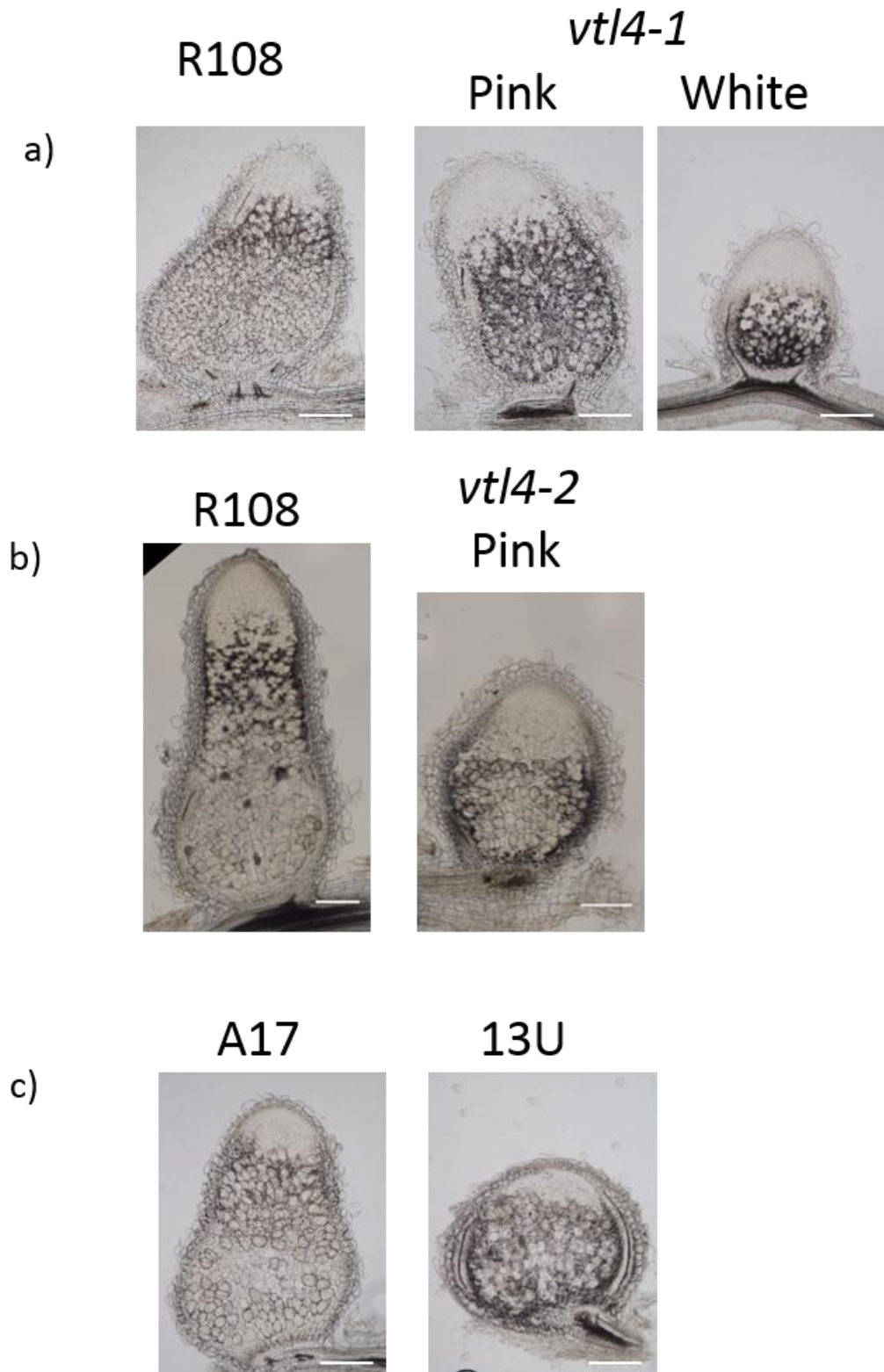
Figure 5.7 Lugol's stain for presence of starch granules in R108 *vtl4*, A17 and 13U nodules

14 dpi nodules of plants grown on zeolite were infected with *Sm1021*. a) R108 and *vtl4-1* b) R108 and *vtl4-2* c) A17 and 13U). Nodules were fixed embedded in agarose and sectioned to 75 μm thickness using a vibratome, before staining with Lugol's solution to detect starch. Scale bar represents 0.25 mm.

Syto^{TM} 13 is a fluorescent nucleic acid stain that can be used to interrogate bacteroid development in the nodule. In R108, signal in the meristem is localised to plant nuclei. In the infection zone the majority of the signal is due to the infecting bacteroids which are Tic Tac® shaped. Signal is the strongest in the differentiation zone, where bacteria are differentiating, elongating and radially orientating themselves around a central infected plant cell vacuole. The differentiation zone can be seen as the "bright band" across the nodule section in Figure 5.8a, before the nitrogen fixation zone, where the signal weakens and the organisation of the bacteroids within the infected cells becomes less ordered. This can be seen at 40 x magnification (Figure 5.8b) and 63 x magnification (Figure 5.8c).

Both pink and white nodules from *vtl4-1* exhibit the" bright band" of the differentiation zone, and the overview at 10 x magnification (Figure 5.8a) shows a significant proportion of cells in the nitrogen fixation zone are colonised, similar to the wild type. Bacteroids are elongated and arranged in an ordered fashion, similar to the wild type, but in the nitrogen fixation zone (or basal zone in white nodules) fluorescence is more pronounced in discrete dots within the bacteroids.

In nodules from R108 plants grown alongside *vtl4-2*, the "bright band" of the differentiation zone is clearly visible, with some cells in the nitrogen fixation zone having a stronger Syto^{TM} 13 signal and others (probably in different cell layers) a weaker signal (Figure 5.8d). The differentiation zone in *vtl4-2* nodule sections was less pronounced and bright than R108. One phenotype of *vtl4-2* nodules that is noticeable is their overall size, which is smaller than that of R108 wild-type nodules. This was consistently the case across all nodules sampled, and was noticeable at the point of harvesting. At 40 x magnification it becomes clearer that vacuoles in the infection and differentiation zones are larger in both mutants than the wild type. To establish whether this observed difference was statistically significant, I measured the vacuoles of all clearly identifiable differentiation zone cells in both *vtl4* mutants. On average, *vtl4-1* vacuoles are 3.2 μm and *vtl4-2* vacuoles are 2.7 μm larger than those in the R108 wild type (Figure 5.9a and b respectively).
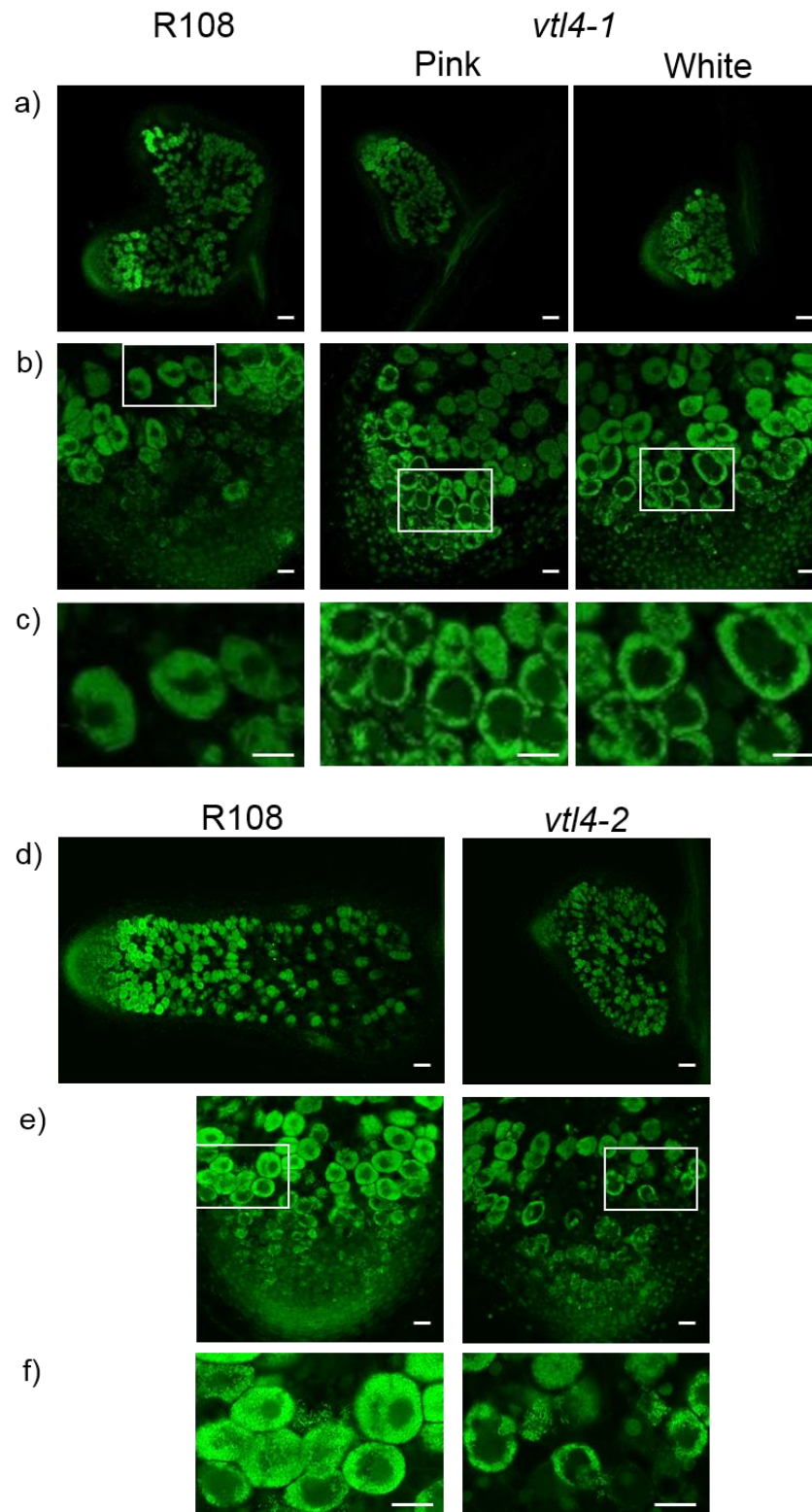
Figure 5.8 Syto™ 13 DNA stain of R108 and *vtl4* mutant nodule sections
a)-c) Show comparison between R108 and *vtl4-1* nodule sections and d)-f) R108 and *vtl4-2* nodule sections a) and c) Nodule overviews using a 10 x objective, scale bar represents 100 µm b) Higher magnification view of the meristem to the interzone, using a 40 x objective, scale bar represents25 µm, with insets c) and f), highlighting the difference in vacuole size between R108 and *vtl4* mutant differentiation zone cells, scale bar represents 25 µm. Nodules were fixed and embedded in agarose and sectioned to 75 µm thickness using a vibratome. Images a)-c) were captured using Leica TCS SP8 confocal microscope, by Gergely Iski, images d)-f) were captured using Leica SP5 confocal
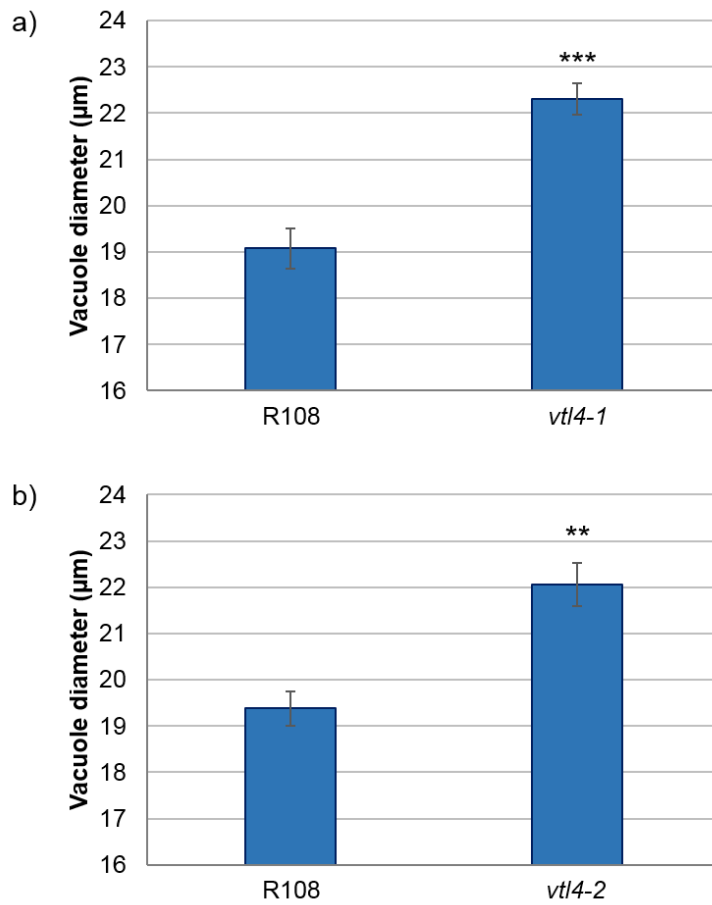
Figure 5.9 Vacuole diameters of differentiation zone cells of *vtl4* mutant nodules

a) *vtl4-1* and b) *vtl4-2* mutant nodules at 14 dpi, infected with *Sm1021*. Measurements were carried out using FIJI software on images represented in Figure 5.8. Error bars represent standard error of the mean. Significance was inferred using Student's T-test assuming unequal variances, ** P ≤ 1 x 10-4, *** P ≤ 1 x 10-7. n = 88 and 284 cells respectively for R108 and *vtl4-1* (a) and n = 122 and 89 cells for R108 and *vtl4-2*

As seen in the R108 wild type, A17 nodule sections are characterised by nuclear staining of plant cells in the meristem, a bright band of cells indicating the differentiation zone and good colonisation (the majority of cells fluorescing) in the nitrogen fixation zone, where the bacteroid structure is starting to become less ordered (Figure 5.10).

By contrast, 13U nodule sections are smaller and show significantly less fluorescence as a whole (Figure 5.10a). It is also clear that there is no bacterial colonisation of nodule cells in the basal zone of the nodule (that would normally be described as the nitrogen fixation zone), in agreement with the X-gal staining results. The only fluorescence that can be seen in this zone is either due to plant nuclei or infection threads repeatedly trying to infect cells even after infection has stalled after the differentiation zone. These infection threads can be seen spreading outwards from an infected cell in the differentiation zone in the Martian-like structure in the bottom 13U differentiation zone image (Figure 5.10c).

In summary, the histological data suggest that lack of *VTL4* caused slight changes to morphology of infected cells, such as increased vacuole size, however on the whole nodule development proceeds as in the wild type. Deletion of several genes missing in the 13U mutation is much more drastic, causes stalling of infection after the interzone and bacteria do not fully differentiate or elongate into the mature, nitrogen-fixing state.
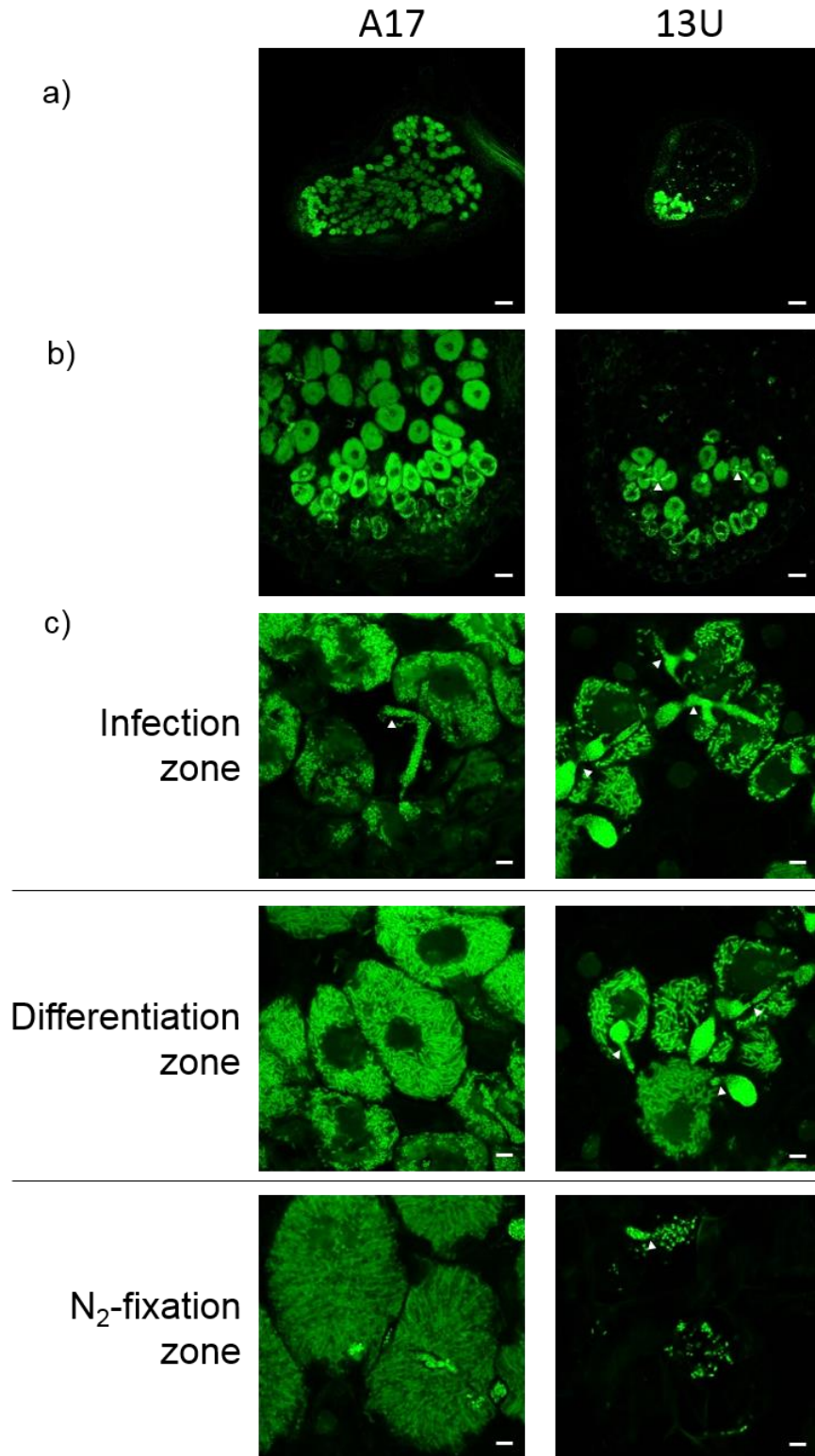
Figure 5.10 Syto^{TM} 13 DNA stain of A17 and 13U nodule sections.
a) Overview of A17 and 13U nodule sections using a 10 x objective, scale bar represents 100 μm b) Higher magnification view of the meristem to the interzone, 40 x objective, scale bar represents 25 μm c) High magnification images of cells in the infection, differentiation and nitrogen fixation/pseudo-nitrogen-fixation zones using a 63 x objective, scale bar represents 5 μm. Infection threads are indicated by arrows. Nodules were fixed and embedded in agarose and 75 μm sections cut using a vibratome. Images taken using Leica TCS SP8 confocal microscope, by Gergely Iski

5.3.4 Iron transport properties of VTL4

To investigate the Fe-transporting capabilities of VTL4, yeast complementation assays were carried out using mutants deficient in the iron transporters Fet3 or Ccc1. Fet3 is required for high-affinity transport of iron into yeast cells (Askwith et al., 1994). Cells lacking Fet3 can still take up Fe through the low-affinity Fet4 transporter, so bathophenanthroline disulphonic acid (BPDS) is added to the media to chelate free Fe and render the yeast Fe deficient, resulting in cell death.

The Δ*fet3* mutant was transformed with the empty vector (p416), *S. cerevisiae FET3*, and *MtVTL4* (Figure 5.12a). When BPDS is added to the growth media almost no growth is observed for Δ*fet3* + p416 and Δ*fet3* + *MtVTL4*, while for Δ*fet3* + *ScFET3* and BY4741 wild type, individual colonies can be seen in the 10^{-4} dilution (Figure 5.11a, left panel). Adding excess iron (200 μM) to the media resulted in the BY4741 wild type and Δ*fet3* mutants growing similarly well, as Fe will be taken up through the intact Fet4 transporter (Figure 5.11a, right panel). Thus, in this experimental set-up, *MtVTL4* does not complement Δ*fet3*.

Ccc1 is an iron transporter localised to the vacuolar membrane in yeast cells. As previously discussed (Chapter 1), Fe build-up in the cytosol is toxic to the cell; yeast cells sequester Fe to the vacuole, so lack of Ccc1 results in impaired growth on high Fe media. Wild-type BY4741 grows well regardless of Fe concentration in the range used for this experiment (5-10 mM, left to right, Figure 5.11b). By contrast there was no growth at any of the three Fe concentrations of Δ*ccc1* transformed with the empty vector (p416) or *MtVTL4* (Figure 5.11b,). By contrast reintroduction of the wild-type *ScCCC1* gene complemented the mutation, allowing growth on media with ≥ 5 mM Fe (Figure 5.11c, James Connorton). Again, using this experimental set-up, *MtVTL4* does not complement Δ*ccc1*.
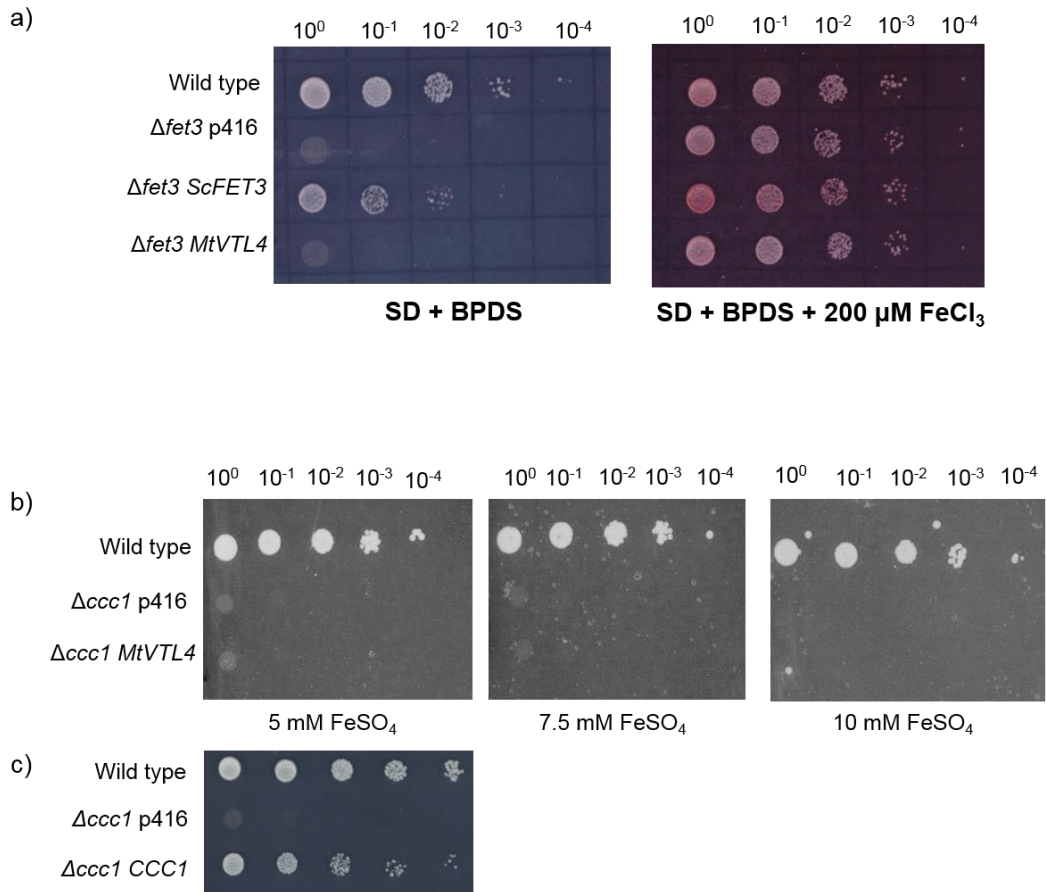
Figure 5.11 Yeast complementation assay with Mt*VTL4*
Complementation of *Δfet3* (a) and *Δccc1* (b and c) Drop assays were carried out on minimal media (SD) supplemented with histidine, leucine, methionine, BPDS and $FeCl_3$ (a) or $FeSO_4$ (b and c). c) shows the *Δccc1* mutant complemented with the wild type *S. cerevisiae CCC1* gene (data from James Connorton).

5.4 Functional complementation of the 13U mutant line

The white nodules produced by the 13U mutant line are completely lacking in the ability to fix nitrogen, as measured using the acetylene reduction assay method; a proxy for nitrogen fixation (Domonkos et al., 2013). Thus, complementation of 13U, even if partial, can be assayed using the ARA as any acetylene reduction measured is due to the presence of the complementation construct.

5.4.1 Construct design for complementation

To establish if the deletion of *SEN1* and *VTL4* is responsible for the 13U phenotype, constructs encoding these genes alone and in combination were designed (Figure 5.12). Usually when synthesising genes for use with Golden Gate cloning, internal restriction sites of the enzymes used in Golden Gate must be destroyed, in a process called domestication, however the *SEN1* and *VTL4* coding sequences were not domesticated as there were none of these internal restriction sites present in either coding sequence. Both genes were under the control of their native promoters. For *SEN1*, the 5'UTR (96 bases) and 2000 bases upstream was cloned to encompass the hypothesised promoter region. A larger region was chosen due to the differing localisation of the GUS staining under the control of the *MtFECH1B* promoter than predicted by the RNAseq data of Roux et al. (2014) (Section 4.4). The 5' UTR of *VTL4* (566 bases) includes a 321 base intron, which was cloned along with an additional 2000 bases upstream to encompass the promoter region. During the domestication process, two and one restriction sites had to be removed from the promoters of *SEN1* and *VTL4* respectively (Appendix 1).
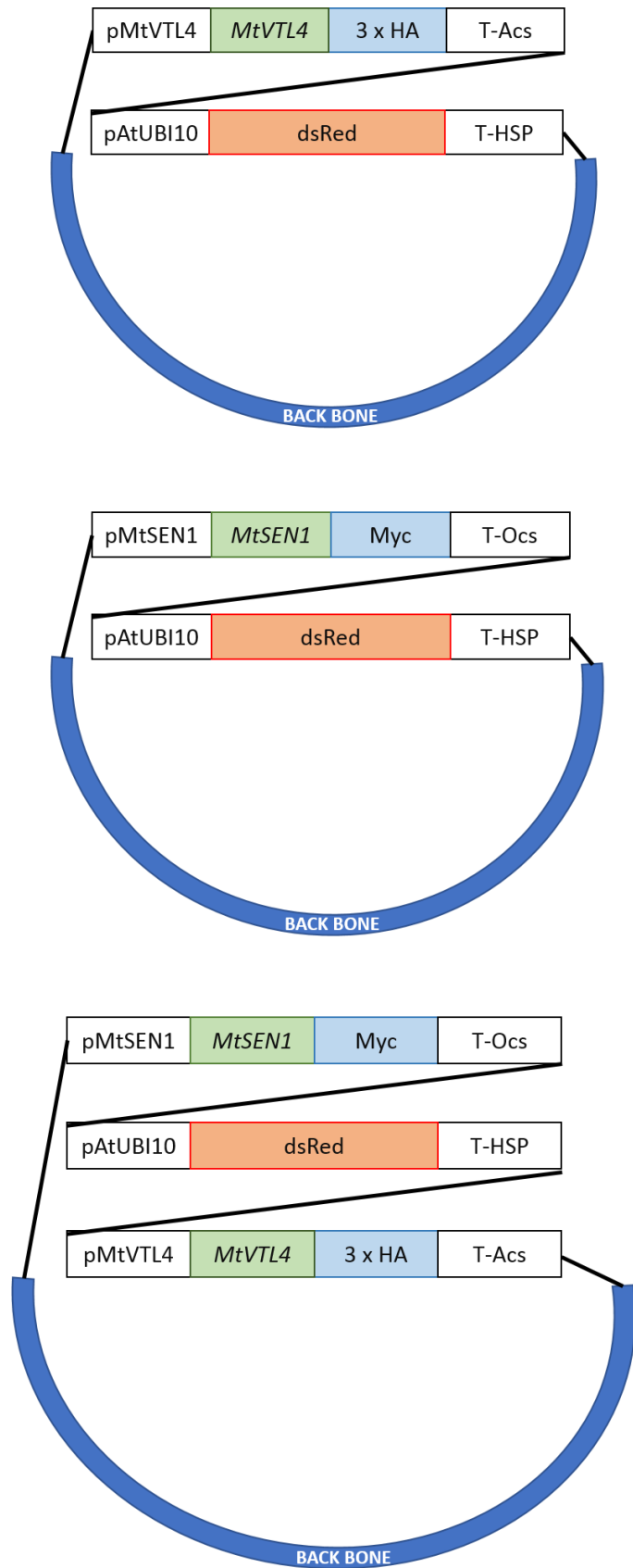
Figure 5.12 Schematic representation of plasmids for functional complementation of 13U mutant
Plasmids contain *VTL4*, *SEN1*, or both genes

5.4.2 Acetylene reduction assay to measure complementation

To establish whether transformation of Medicago roots with *VTL4*, *SEN1*, or both genes above can restore nitrogen fixation in 13U, plants were transformed as described in Section 2.9.4. After one week of growth on Mod FP medium, transformed seedlings were moved to Terragreen/sand and inoculated with *Sm1021*. Acetylene reduction rates were assessed at 28 and 35 dpi as a proxy for nitrogen fixation, which cannot easily be measured. A free GFP control plasmid was used to transform both Jemalong A17 and 13U roots for positive and negative controls respectively. Results are summarised in Figure 5.13.

Wild-type A17 plants at 28 dpi transformed with the control plasmid reduced acetylene to ethylene at an average rate of 1.038 µmoles per jar (5 plants) per hour, while 13U plants transformed with the same construct generated no ethylene in the same time (Figure 5.13a). 13U plants transformed with *VTL4* alone showed no increase in the amount of acetylene reduction relative to the negative control (Figure 5.13a). 13U plants transformed with *SEN1* showed a small amount of acetylene reduction (0.00872 µmoles per jar (5 plants) per hour, or 0.0025%), although the error bars are such that this cannot be treated as significant. 13U plants transformed with both *VTL4* and *SEN1* show a small but significant increase in ethylene production relative to the 13U control (0.01745 µmoles per jar per hour, or 0.005%), although this is still nowhere near the rate (~0.3%) of the wild-type control line.

At 35 dpi the only appreciable acetylene reduction capability was in the Jemalong A17 control, with an average rate of 4.65 µmoles ethylene per jar per hour (Figure 5.13b). This was also the case in subsequent experiments at 21 and 28 dpi, so the initial promising result could not be confirmed.
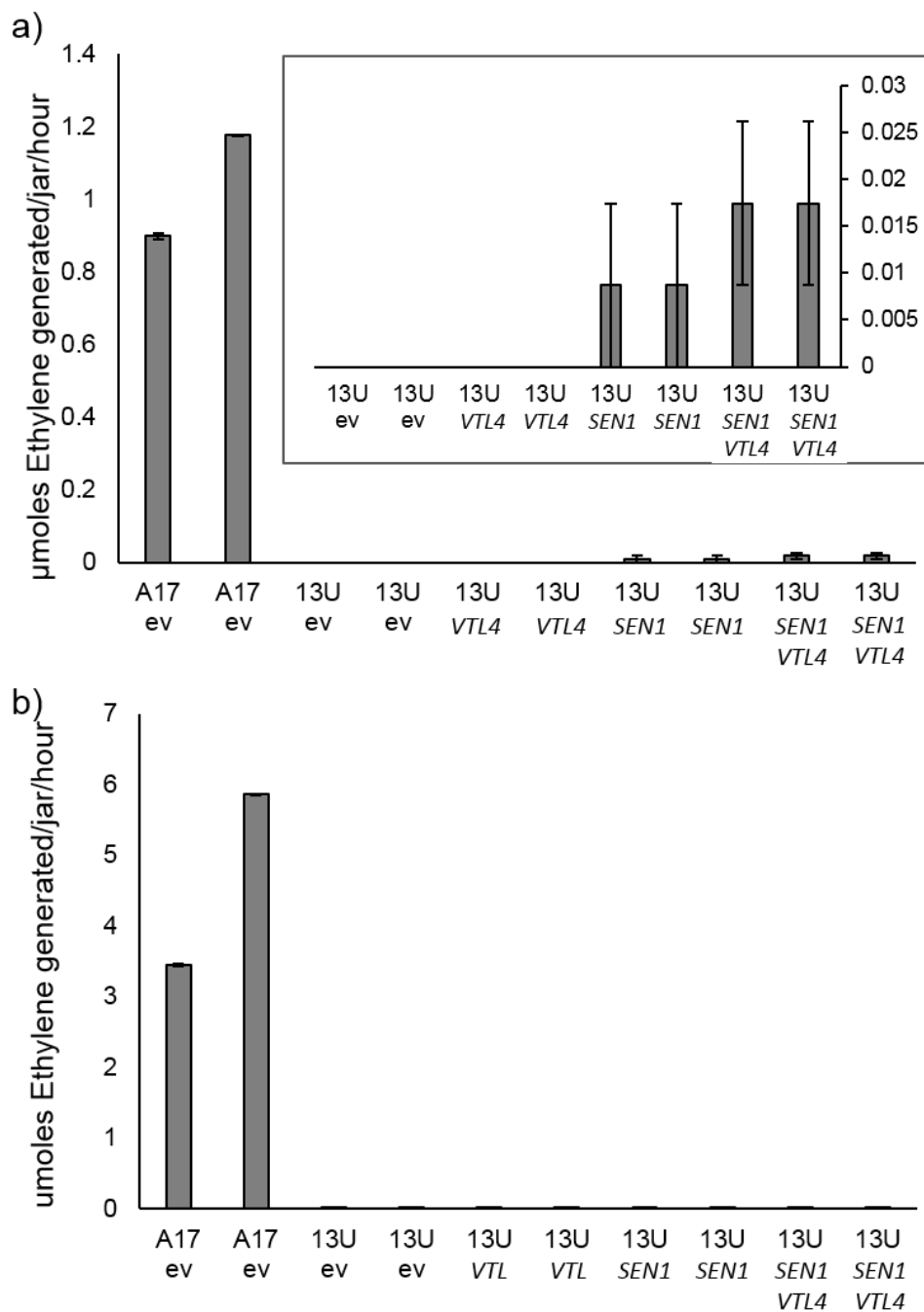
Figure 5.13 Functional complementation of 13U as measured using the acetylene reduction assay
Plants were grown on Terragreen/sand, irrigated with water and inoculated with Sm1021. They were assayed at a) 28 dpi or b) 35 dpi using acetylene reduction capability as a proxy for nitrogen fixation. Error bars represent standard error of the mean of three technical replicates.

5.5 Localisation of SEN1 and VTL4 proteins

5.5.1 Localisation predictions for SEN1 and VTL4

The amino acid sequence for VTL4 contains the LXXXLL motif, known to be important for tonoplast localisation in Arabidopsis (Komarova et al., 2012), (Figure 5.14). Symbiosome membranes are unique in their identity characteristics; during their development they acquire a variety of markers of endosomal, plasma and vacuolar membrane identity (Limpens et al., 2009), making it impossible to decipher which proteins will localise to the symbiosome membrane based on amino acid motifs alone. The LXXXVL motif in the SEN1 sequence is likely to function in a similar manner to the LXXXLL motif characterised in Arabidopsis due to the similarity between leucine and valine (both being apolar and a similar size).

```
VTL4      MKRLCFLHATLSSFSFNLGILRFLYFMASLGNHNNNEV
SEN1      MAVGTICDSESL---------------NKVITIPNPTK
cons      *     : .:                     .: .  *

VTL4      NDIEAKQNQTQDTEENNIDYSKRAQWLRAALLGANDGL
SEN1      EMEEKQIKIKENCKTSDIDYWQRAQWLRAAVLGANDGL
cons      :   * : : .:: : .:*** :*********:******

VTL4      VSITSLILGVGAVHEDIKTMLLAGFAGLIAGACSMGIG
SEN1      VSVASLMMGVGAVKTDSATMLVAGFAGLIAGACGMAIG
cons      **::**::*****: *   ***:***********.*.**

VTL4      EFVSVYTQFDIMVAQMKRENKINS------TFVEEEKQ
SEN1      EFVSVYTQYEVEIGQMMRDLGTSDRKEKELEIELEKRR
cons      ********::: :.** * :  ..        :  *:::

VTL4      LLPNPFQAAIASAIAFSFGATVPLLGAALVRDYKIRLF
SEN1      SLPNPLQAAAASAFSFSIGGLVPLLSGSFIRVYKIRII
cons       ****:*** ***::**:*. ****..::* ****::

VTL4      VVVGMASFALLVFGGVGAILGKTSVKMSCVRVVVGGWM
SEN1      AIMAIASLALVVFGGVGAMLGKTPKVKSSIRFLLGGWM
cons      .::.:**:**:*******:****.    *.:*.::****

VTL4      AMAITFGLTKFVGYS-S----L
SEN1      AMAITFGLTKLLAHCSGLDLEI
cons      **********::.:. .     :
```

Figure 5.14 Alignment of SEN1 and VTL4 sequences carried out using M-COFFEE, with established dileucine motif highlighted in blue (LXXXLL) and putative dileucine motif highlighted in red (LXXXVL). Identical residues are marked with the *symbol

5.5.2 Localisation of SEN1 and VTL4 using mCherry fusions

The flexible linker used by Timofeev et al. (2016) for tagging Arabidopsis VTLs was added to the C-terminal end of the gene *in silico* before being synthesised. They found that due to the small size of the VTLs (approximately 27 kDa, similar in size to mCherry) functionality was severely affected by tagging with a fluorescent protein without the addition of this flexible linker to separate the proteins.

5.5.3 Generation of symbiosome membrane-specific markers

There was no published marker for the symbiosome membrane, which is derived from the plasma membrane of the host cell; localisation to the symbiosome membrane is temporally controlled by the promoter and other unknown regulatory elements (Oldroyd, personal communication). In order to construct my own marker, I searched the available proteomic datasets that characterise symbiosome membrane proteins and identified the symbiosome membrane sulphate transporter LjSST1 as a good candidate (Wienkoop and Saalbach, 2003; Krusell et al., 2005; Clarke et al., 2015; Downie, personal communication). The coding sequence from *L. japonicus* was used as there are several Medicago sequences with high sequence similarity, which could make identification of the true orthologue difficult. In addition, the recommended promoter for constitutive expression in Medicago is the *L. japonicus* ubiquitin promoter, which is very effective in highly expressing genes in Medicago (Breakspear, personal communication) The coding sequence for *L. japonicus* SST1 was domesticated to remove four internal Golden Gate restriction enzyme sites and synthesised. To clone the *LjSST1* promoter, 2000 base pairs upstream of the start codon was domesticated, removing four Golden Gate sites, before the sequence was synthesised. The *SST1* coding sequence was fused to eGFP with a Gly-Gly linker with a Nos terminator (Figure 5.15).

AtPIP2A is a characterised plasma membrane marker (Cutler et al., 2000). The symbiosome membrane is derived from the plant plasma membrane, so AtPIP2A can be used as alternative to a symbiosome membrane-specific marker as the symbiosome membrane can be seen surrounding mature bacteroids, thus is distinguishable from the host cell plasma membrane.

The efficacy of both LjSST1 and AtPIP2A as symbiosome membrane markers will be described in parallel to the localisation of SEN1 and VTL4 in the following two sections.
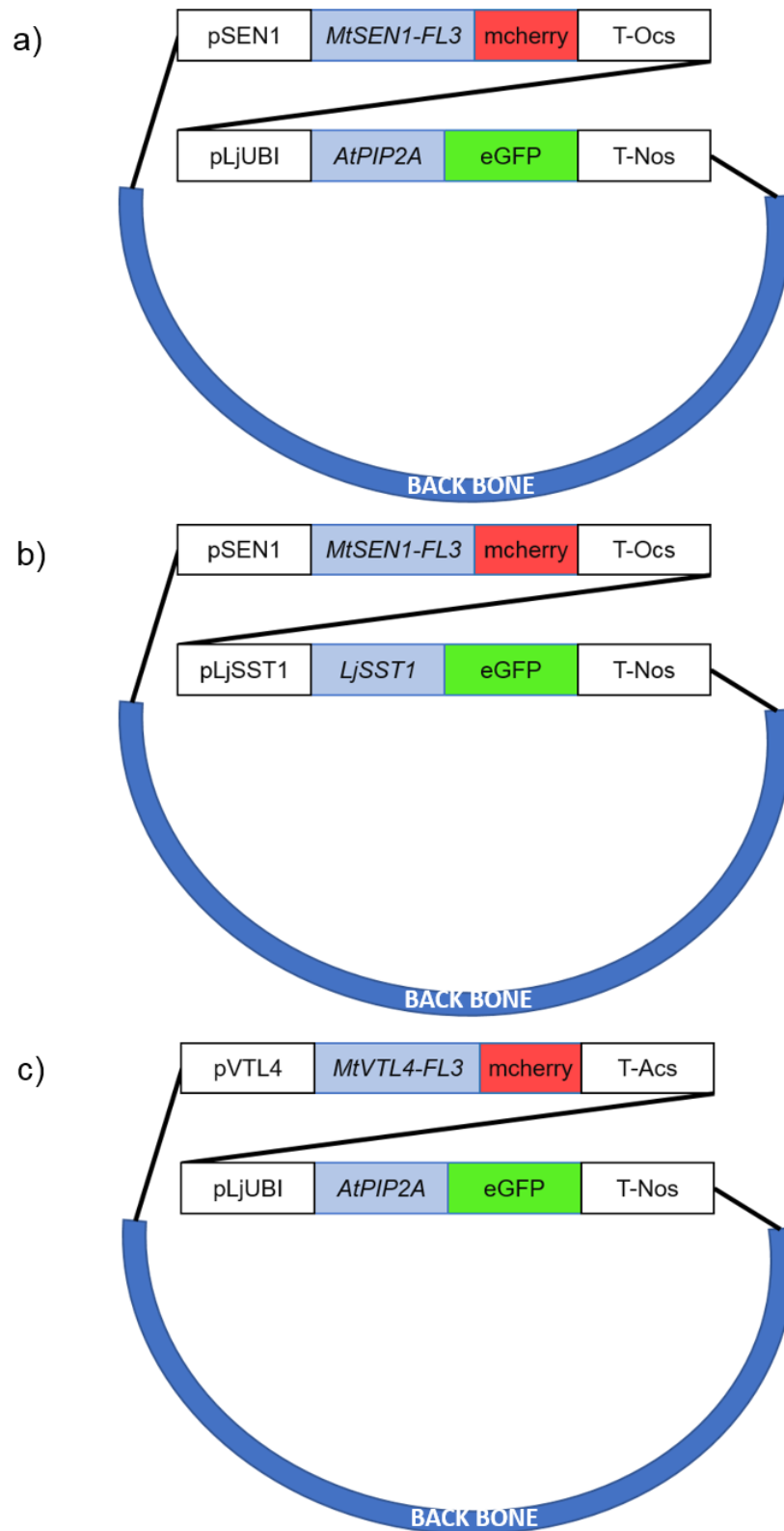
Figure 5.15 Schematic representation of plasmids for localisation of SEN1-mCherry and VTL4-mCherry
a) *SEN1-mCherry* and plasma membrane-localised *eGFP*, b) *SEN1-mCherry* and the symbiosome membrane marker *SST1* fused to *eGFP* and c) *VTL4-mCherry* and plasma membrane-localised eGFP.

5.5.4 SEN1 is localised to the symbiosome membrane in differentiation zone cells

For the localisation of SEN1, two two-gene constructs were made (Figure 5.15 a and b). One of these contained a plasma membrane marker (AtPIP2A) and the other the symbiosome membrane marker (LjSST1), as described in Section 5.5.3, cloned into the EC50506 kanamycin-resistant backbone. Laser scanning confocal microscopy was used to visualise hand-sectioned nodules from these two transformed lines.

Localisation of the SEN1-FL3-mCherry is predominantly to the plasma membrane and inside the infection thread with only very few loci of signal in the body of the cell (Figure 5.16a). The symbiosome membrane marker SST1-eGFP appears to localise predominantly to the plasma membrane of cells in the early infection zone, with some punctate signal also being detected inside the cell and in the cross section of the infection thread (Figure 5.16a). Examination of the DAPI staining suggests that this punctate localisation is either where bacteria or organelles are situated, suggesting that the bacteria are at an early stage of development into bacteroids, before elongation.

In the slightly more mature nodule cells of the late infection/early differentiation zone, MtSEN1-FL3-mCherry signal is localised to the plasma membrane (Figure 5.16b). Some small spots of fluorescence are also visible in the body of the cell. Again, most of the SST1-eGFP signal is localised to the plasma membrane, although some is localised to the centre of the cell, overlapping with the DAPI signal from the bacteria, suggesting that some SST1-eGFP localises to the symbiosome membrane, although this is not the majority of the signal detected (Figure 5.16b).

In late interzone cells SEN1-FL3-mCherry signal is localised to the plasma membrane of the infected plant cell, as well as surrounding individual bacteroids (Figure 5.16c). This pattern consistently overlays with that of the plasma membrane marker, AtPIP2A-eGFP, the signal for which can clearly be seen on the cell membrane of the infected cell and also outlining the elongating bacteroids inside the cell (Figure 5.16c). This indicates that SEN1-FL3-mCherry is localised to both the plasma and symbiosome membrane of infected cells in the nodule interzone.
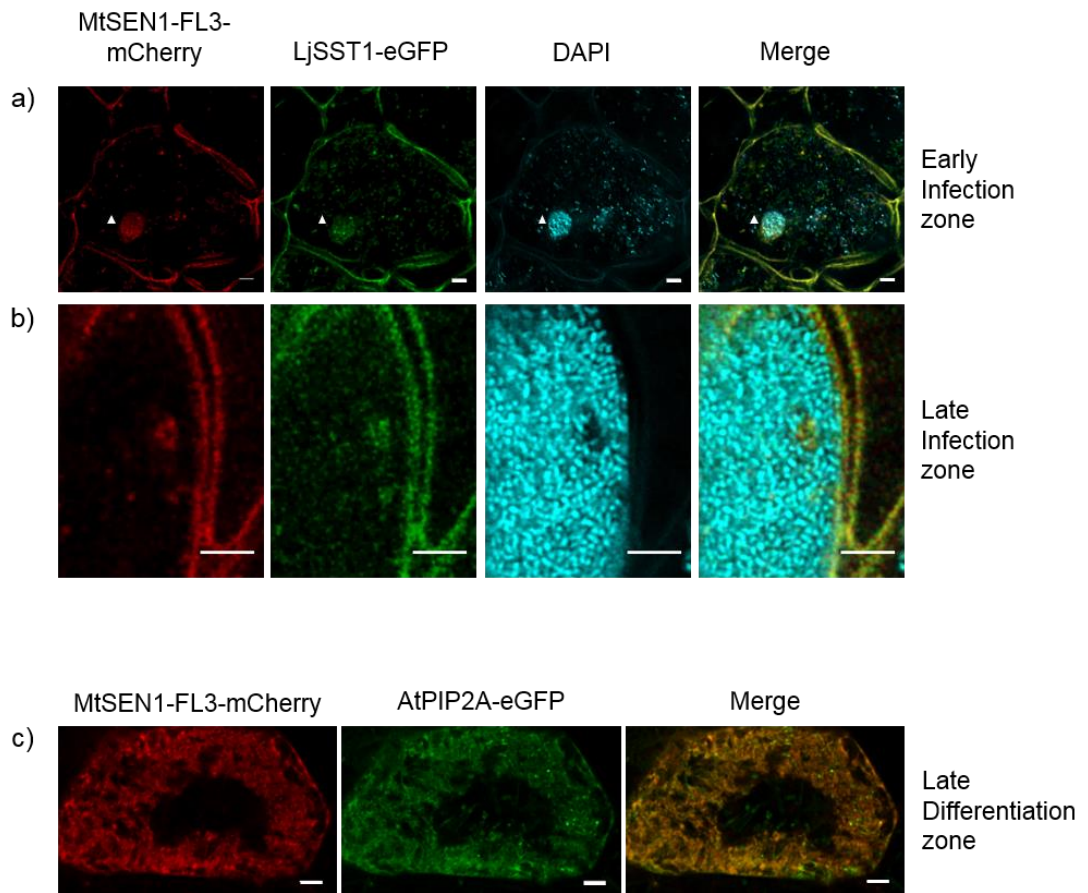
Figure 5.16 Localisation of MtSEN1-mCherry across different stages of nodule development
a) Early infection zone, infection thread indicated with an arrow; b) Late infection zone; c) late differentiation zone. a) and b) Left panels, MtSEN1-FL3-mCherry; second from left LjSST1-eGFP; second from right, DAPI nucleic acid stain; right panels, merge. c) Left panel, MtSEN1-FL3-mCherry; middle panel, AtPIP2A-eGFP; right panel, merge. Data collected using laser scanning confocal microscopy (LSCM) on a Leica SP5 microscope. Scale bar represents 5 μm in all images.

5.5.5 VTL4 is localised to the plasma and infection thread membranes of early infection zone cells

To study the localisation of VTL4, one construct was made, containing the plasma membrane marker AtPIP2A (Figure 5.15c).

VTL4-FL3-mCherry is localised to the plasma membrane and the membrane around the infection thread (Figure 5.17, top panel). There is some signal in the core of the infection thread (inset, right-hand panel) although the resolution is not sufficiently high for individual bacteria to be identified. AtPIP2A-eGFP localises to the plasma membrane of Medicago nodule cells (Figure 5.17, second from top panel), as well as the membrane around the infection thread (labelled IT in Figure 5.17) and around the individual bacteria within the infection thread as the symbiosome membrane. DAPI staining shows bacteria at a high density in the infection threads but not the rest of the cell, as this cell was in the infection zone and had not yet been colonised. The overlay of these three channels shows that there is co-localisation between the eGFP-tagged plasma membrane marker and VTL4-mCherry at the membrane of the uninfected cell and the infection thread membrane (Figure 5.17, bottom panel).
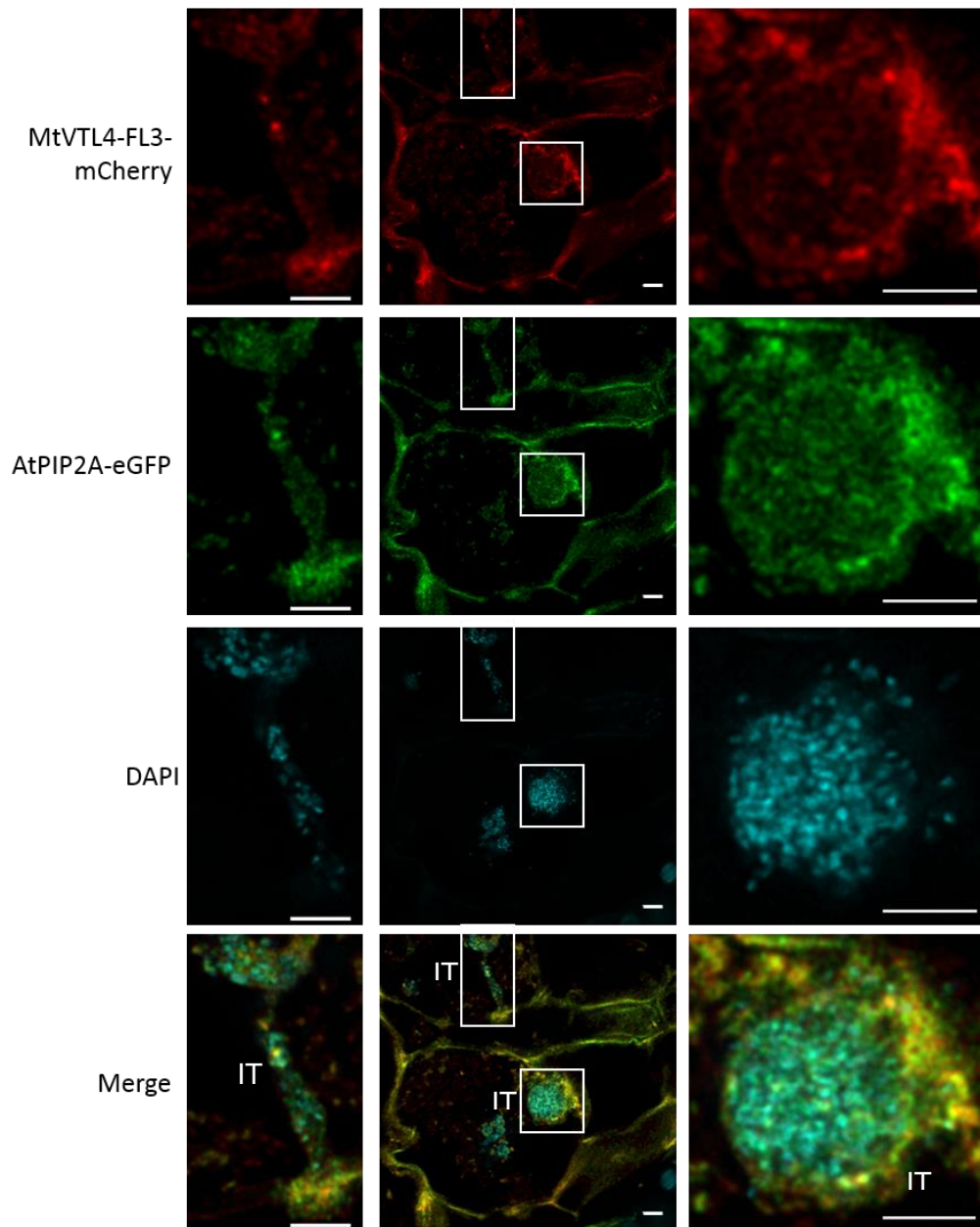
Figure 5.17 Localisation of MtVTL4-FL3-mCherry to the plasma and infection thread membranes of infection zone cells
VTL4-mCherry (top panels), AtPIP2A-eGFP (second from top panels) and DAPI nucleic acid marker (second from bottom panels) in an infection zone cell. The overlay of all three is shown in the bottom panels. The left and right panels are zoomed in views of an infection thread, labelled IT in two different planes. Data collected using laser scanning confocal microscopy (LSCM) on a Leica SP5 microscope. Scale bar represents 5 μm.

5.6 *Sm1021 mbfA* as an iron reporter

As previously discussed, *mbfA* (*SMc00359*) is transcribed in response to high iron in *S. meliloti* (Section 5.1.4). The mechanism of this Fe response is summarised in Figure 5.19a. Using the Roux et al. (2014) data set, we also found that *mbfA* is expressed in nodules (Figure 5.19a), specifically in the interzone (or differentiation zone). We hypothesise that *mbfA* is expressed in nodule cells when *S. meliloti* experience Fe influx in the interzone. Using this information we designed a *lux*-based reporter for bacterial Fe levels, by cloning the promoter of *Sm1021 mbfA* (P$_{mbfA}$) upstream of the *lux* operon genes and transforming *S. meliloti* with the reporter construct. This allowed us to interrogate subtle changes in the Fe levels perceived by bacteroids in *vtl4* and 13U mutant nodules.

5.6.1 Construction of P$_{mbfA}$*lux* reporter construct for nodule iron status (Rob Green)

To investigate the expression of *mbfA* in *Sinorhizobia* in both the free-living and symbiotic states, the promoter of *mbfA* was cloned into a vector containing the full *lux* operon (Figure 5.18). Molecular cloning was carried out by Rob Green. The positive control plasmid for this experiment was generated by mutating four of the key residues in the ICE box, thus abolishing the iron responsive element causing *mbfA* to be expressed at the maximal levels for its own promoter all the time (Figure 5.18b).

To examine the activity of the *mbfA* promoter in free-living *Sinorhizobia*, *Sm1021* transformed with either the ICE constitutive control or P$_{mbfA}$*lux* were pre-grown in TY media, then inoculated at a dilution of 1 in 100 into UMS either with or without the addition of 80 μM FeSO$_4$. These were grown overnight and both the OD600 and luminescence intensity measured at the end of this growth period. The luminescence intensity was then corrected using the OD600 of the bacteria. There was a slight increase in the luminescence intensity of the ICE mutant in high Fe, but a much higher increase in the luminescence intensity of the P$_{mbfA}$*lux* fusion-expressing bacteria (Figure 5.19b, Rob Green). When this is expressed proportionally, in the ICE box-*lux*-expressing strain luminescence increases by 456% in high Fe relative to low Fe conditions, while the strain expressing P$_{mbfA}$*lux* luminescence increases by 1233%.
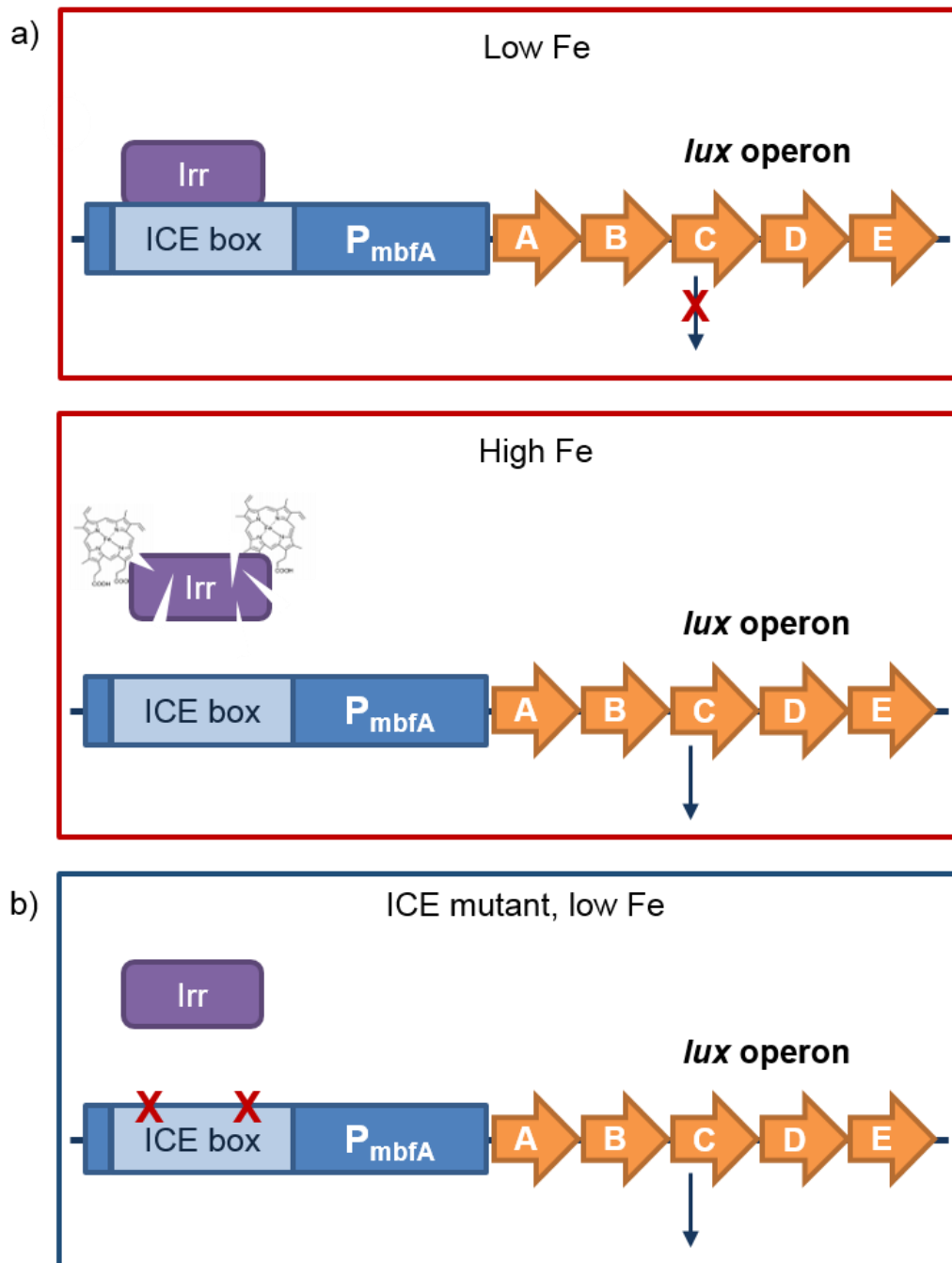
Figure 5.18 Construction and mode of action of the Fe-responsive P_{mbfA}:*lux* construct

Irr = Iron regulated repressor, ICE box = Iron control element. a) The mechanism of gene repression in Fe-limited conditions, and relief of suppression in Fe-replete conditions, b) Mode of action of ICE constitutive promoter.
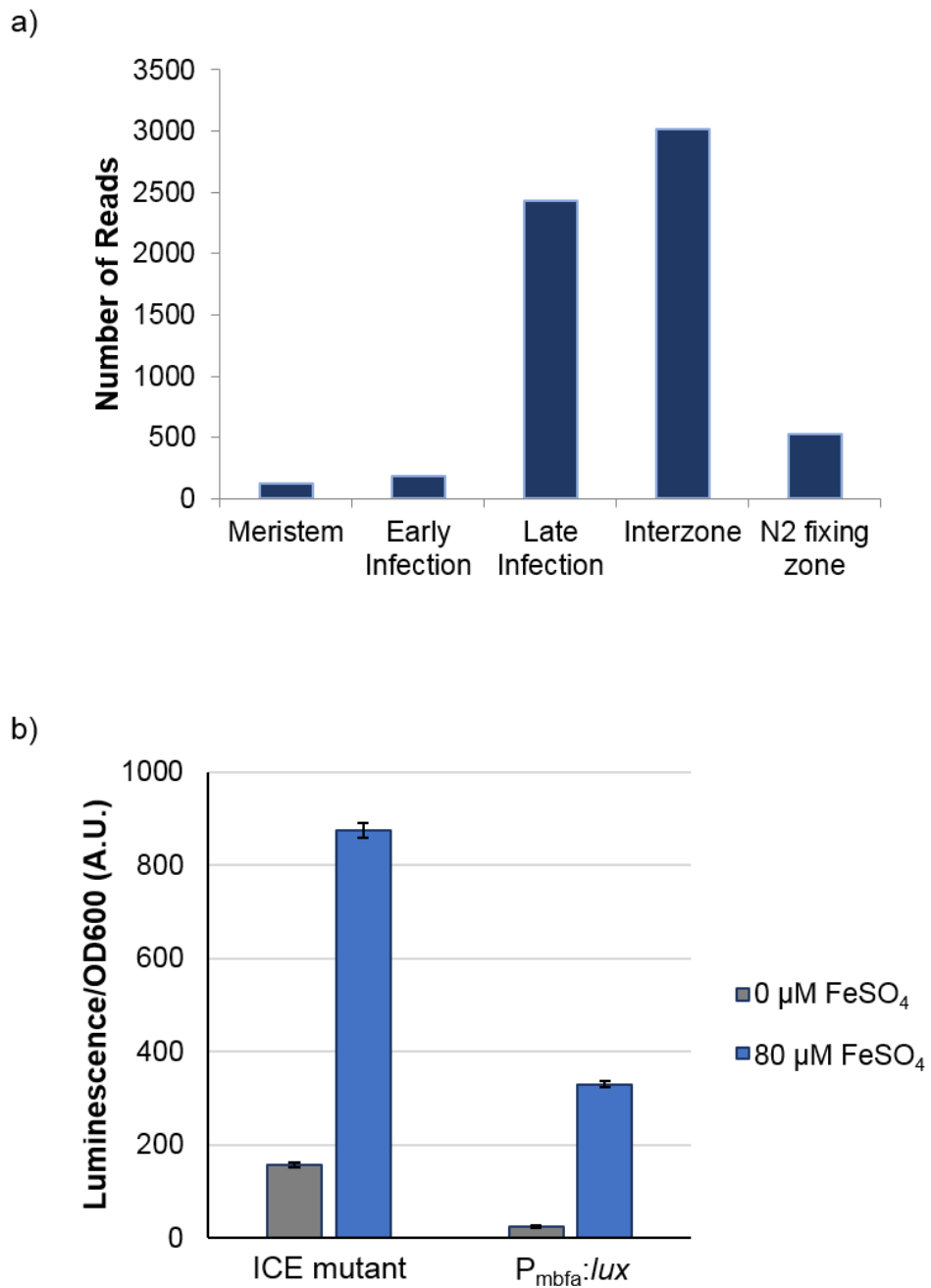
Figure 5.19 *mbfA* expression in nodules and free-living *S. meliloti* subjected to Fe stress
a) *mbfA* expression pattern over five zones of Medicago nodule development. Data taken from (Roux et al., 2014). b) Luminescence intensity of *Sm1021* carrying the lux operon under the control of the constitutive ICE promoter or P*mbfa*lux in zero Fe conditions (grey bars) or high (80 µm) Fe conditions (blue bars), corrected for OD600 as a proxy for bacterial cell density. Error bars represent the standard deviation of three technical replicates. A.U. = Arbitrary Units. Data from Rob Green.

5.6.2 Measuring bacteroid Fe perception using the P*mbfA*lux reporter

To investigate whether the *vtl4* and 13U mutations affect the amount of Fe perceived by endosymbiotic bacteria, these mutant lines (*vtl4-1*, *vtl4-2* and 13U) were grown on Terragreen/sand alongside their relevant wild types (R108 and A17), and inoculated with Sm1021 P*mbfA*lux and the ICE box control. At 21 and 28 dpi plants were dug up, washed and the roots photographed. The roots were then imaged using the NightOwl photon camera to detect luminescence.

There was no significant difference between the luminescence of the nodules from wild-type and mutant plants inoculated with the *Sm1021* ICE box constitutive control (Figure 5.20a).

All of the mutant lines tested showed decreased expression of P$_{mbfA}$*lux* compared to their relevant wild type (Figure 5.20b). Reduction in signal was most pronounced in the 13U mutant, with the average luminescence decreasing from 68462 photons s^{-1}mm^{-2} to 4528 ph s^{-1}mm^{-2}; a decrease of 93%. Although less striking than the difference between A17 and 13U, both *vtl4* mutants also showed a decrease in expression of P$_{mbfA}$*lux*. *vtl4-2* generated slightly less luminescence (19459 photons s^{-1}mm^{-2}) than *vtl4-1* (27005 photons s^{-1}mm^{-2}), compared to 54312 photons s^{-1}mm^{-2} detected from the R108 wild type, decreases of 64% and 50% respectively. This suggests that bacteria in both the 13U the *vtl4* mutant lines are perceiving less iron than those in the wild-type counterparts. Again, the *vtl4* mutants show a phenotype intermediate to the wild type and 13U mutant, further supporting that the phenotype of 13U is partially due to the lack of this gene. The similarity in constitutive expression levels regardless of the plant genotype suggests that any difference observed between plants inoculated with Sm1021 P$_{mbfA}$*lux* is due to a true difference in iron perceived by the bacteria rather than poor colonisation.
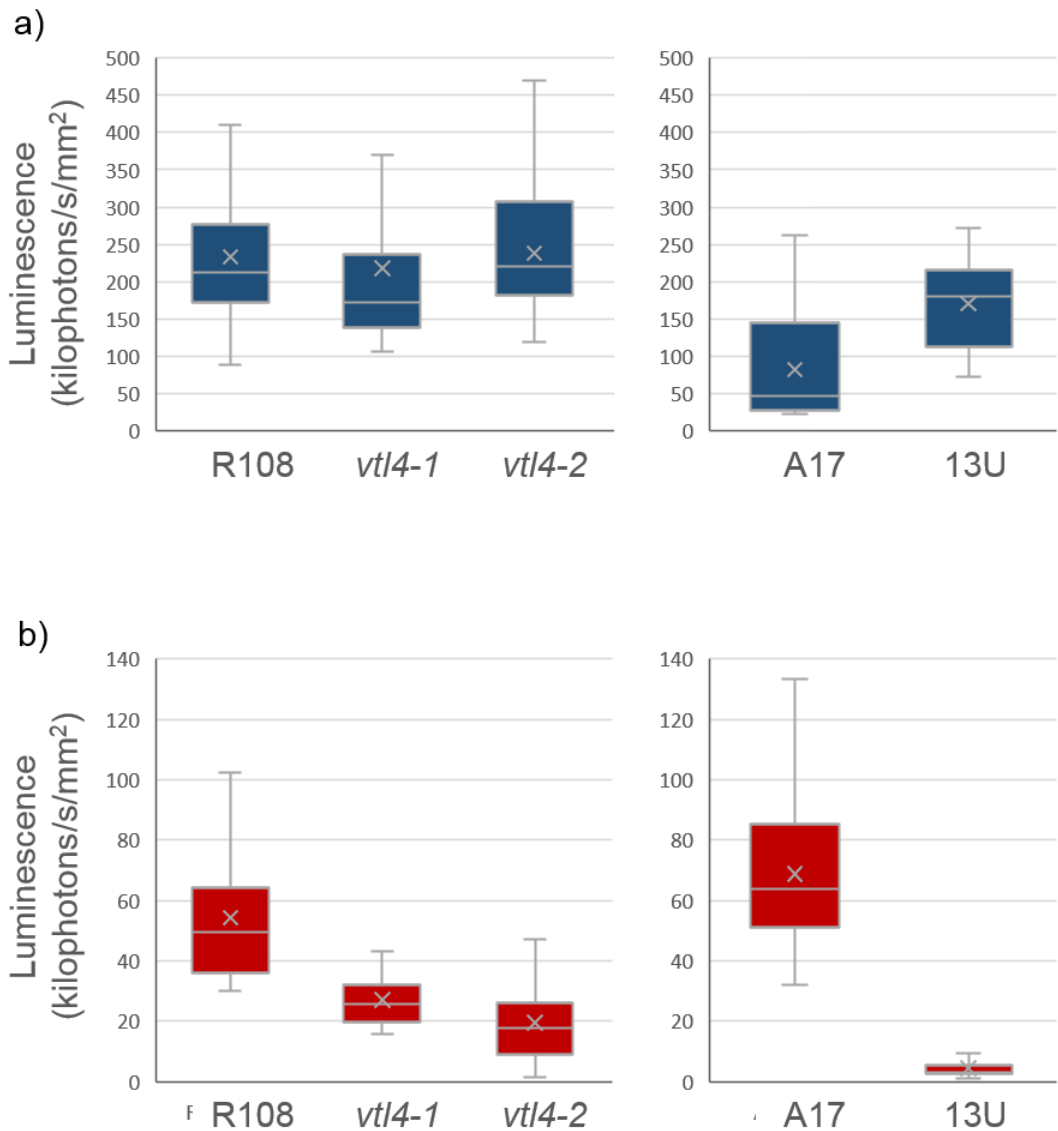
Figure 5.20 Expression of P_{mbfa}:lux in R108, *vtl4*, A17 and 13U nodules
Luminescence of nodules, corrected for nodule area. Plants were grown on Terragreen/sand and inoculated with a) *Sm1021* carrying the constitutive ICE control plasmid or b) *Sm1021* with the P_{mbfa}*lux* plasmid. Plants were grown to 21 days post inoculation and luminescence measured using the NightOwl camera (Berthold). Images were analysed using IndiGO software. The upper and lower limits of the box plots represent the upper and lower quartiles of the data, with the middle line representing the median. The mean is represented by the X symbol and error bars show the spread of the dataset.

5.7 The function of MbfA in the Medicago-Sinorhizobia symbiosis

Expression of *mbfA* is upregulated both in response to Fe stress and in the nodule (Figure 5.19). To determine if the product of this gene plays an important role in the nodulation process, plants were inoculated with wild-type *Sm1021* and a Δ*mbfa* strain (made by Jeremy Thornton). Fortunately, *mbfA* is not essential for nodulation to take place (Figure 5.21) as nodules are still produced when Medicago plants are inoculated with the mutant strain.

There is no significant difference in the number of pink nodules produced by either A17 or R108 plants inoculated with Δ*mbfa* compared to wild-type *Sm1021* (Figure 5.21). However there is a significant increase in the number of white nodules produced by both A17 and R108. This suggests that although MbfA is not essential for nodules to develop to maturity, there is a slight negative effect on nodule functionality, and demonstrates the importance of Fe homeostasis in the bacteroids during symbiosis.
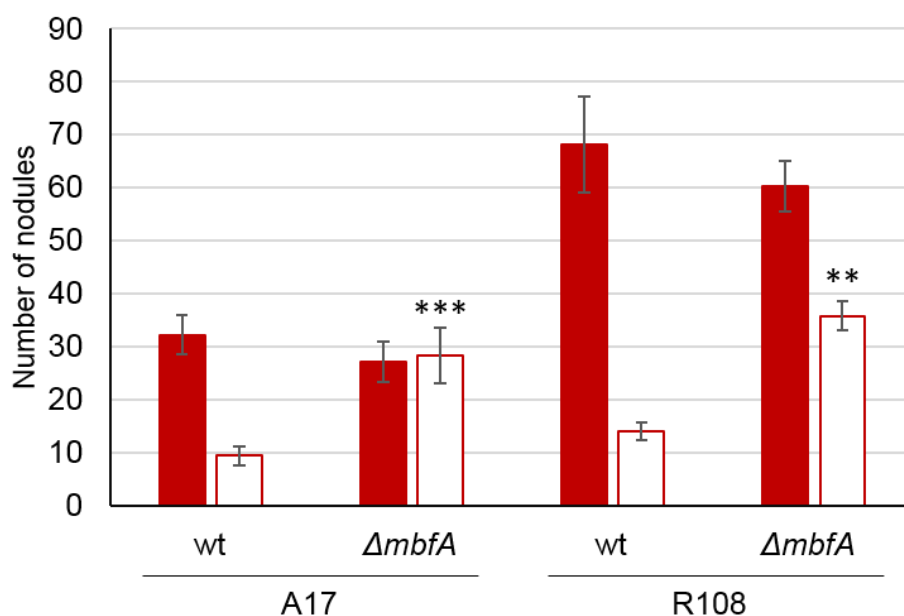
Figure 5.21 Average number of nodules per plant from A17 and R108 plants inoculated with wild-type or Δ*mbfa Sm1021*

Grown in Terragreen/sand, nodules counted at 28 dpi. Average number of pink nodules (red bars) and white nodules (white bars) shown. Number of nodules on plants inoculated with Δ*mbfa Sm1021* was compared to that of plants inoculated with wild-type *Sm1021*. Significance was inferred using Student's T-test assuming unequal variances, * $P \leq 0.1$,** $P \leq 0.01$, *** $P \leq 0.001$. Error bars represent standard error of the mean, data from 10 plants, representative of 2 biological replicates.

To determine whether the Δ*mbfa* mutation has an effect on bacterial differentiation within the developing nodules, A17 nodules infected with either *Sm1021* or Δ*mbfa* were fixed, embedded and sections prepared using a vibratome. These were then stained using Syto™ 13 to allow bacterial development to be monitored in the nodule.

Nodules infected with wild-type *Sm1021* develop into an elongated structure, with all the zones of development (from the meristem to the nitrogen fixation zone) clearly visible (Figure 5.22a and b, left panel). The *mbfA* mutant does not prevent nodules from developing to maturity (as previously noted); they still form the characteristic elongated shape, and are highly similar to the wild-type infected nodules (Figure 5.22a and b, right panel). Closer examination of the infection and differentiation zone of nodules infected with wild-type or mutant bacteria show no significant differences (Figure 5.22c). The nitrogen fixation zone of wild-type nodules is characterised by nodule cells being less intensely stained and "wispy" around the cell periphery as the structure starts to break down. In the nodules infected with Δ*mbfa* bacteria the nitrogen fixation cells seemed to retain a more defined outline. Additionally, they looked bigger, although due to the small sample size used (3-5 nodules) it was not appropriate to draw any firm conclusions from this observation.
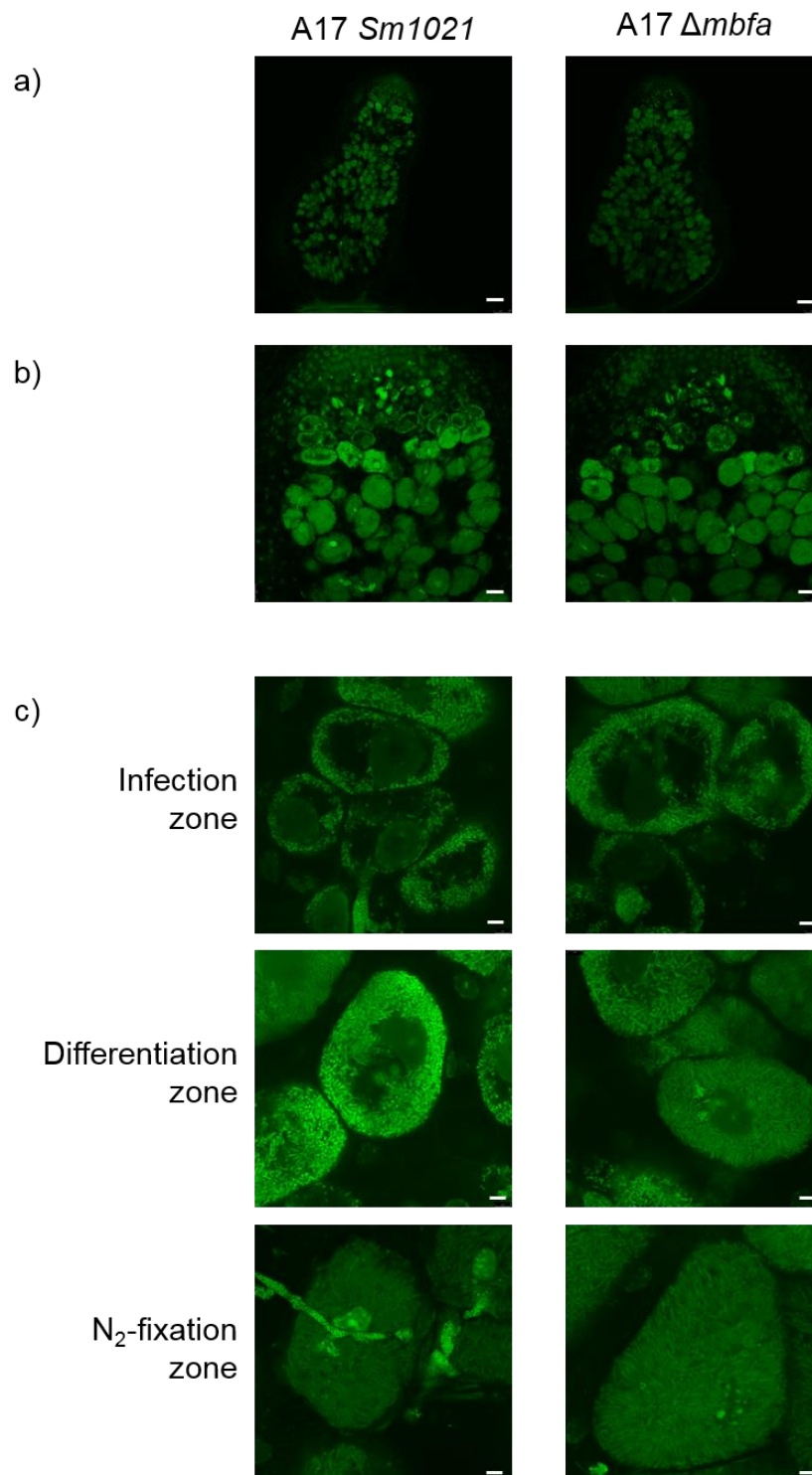
Figure 5.22 Syto^{TM} 13 DNA staining of nodule sections from A17 plants inoculated with wild-type or *Δmbfa Sm1021*

28 dpi A17 plants inoculated with wild-type *Sm1021* (left column) or *Δmbfa* mutants (right column). a) Overview of *Sm1021* and *ΔmbfA* nodule sections using a 10 x objective, scale bar represents 100 µm b) Higher magnification view of the meristem to the interzone, 40 x objective, scale bar represents 25 µm c) High magnification images of cells in the infection, differentiation and nitrogen fixation zones, 63 x objective, scale bar represents 5 µm. Images taken using Leica SP5 confocal microscope. Nodules were fixed and embedded in agarose and 75 µm sections cut using a vibratome.

5.8 Discussion

The aim of this chapter was to characterise the function of *VTL4* and *SEN1* in the context of the symbiotic Medicago nodule by characterising *vtl4* mutants and the 13U deletion mutant that encompasses both *VTL4* and *SEN1*. Characterisation included attempts to complement the 13U mutant, determination of the localisation of both VTL4 and SEN1 by confocal microscopy, and use of a novel bacterial Fe reporter to interrogate the Fe status of endosymbiotic bacteria in the nodules of mutant plants.

5.8.1 Nodule phenotypes were subtly different between the two *vtl4* lines

There were subtle differences between the total number of nodules produced by *vtl4-1* and *2* mutants when they were grown side by side. This is due to the slight but significant increase in white nodule production ($P \leq 0.1$). This difference could be due to the interference from another *Tnt1* insertion in the line, or it could be due to the differing positions of the insertions in the *VTL4* gene; *vtl4-2* is an insertion close to the beginning of the gene, while *vtl4-1* is nearer the end of the gene and it could be that a truncated protein with limited but significant functionality is produced, despite the lack of transcript observed. A greater understanding of the effect of mutations in the *VTL4* gene could be achieved by using electron microscopy to visualise the subtle nodule phenotypes at higher resolution.

5.8.2 *VTL4* failed to complement yeast *Δfet3* or *Δccc1*

One of the first discrepancies with the yeast complementation data was the fact that *ScFET3* does not fully complement the *Δfet3* mutant. This is likely because it is not under the control of its native promoter, but instead is under the control of the Met25 promoter, which is expressed at fairly low levels in the cell, accounting for the difference in growth of the yeast.

A more pressing problem with the yeast complementation experiments was the lack of complementation by *MtVTL4* of either of the yeast Fe transport mutants tested. There are a number of possible explanations for this outlined below.

Firstly it could be that the protein is not expressed in yeast. Often it is necessary to codon optimise genes for expression in heterologous systems. I did not do this after comparing the codons used in *MtVTL4* with those commonly used in yeast, as I found that there were no rare codons used in this gene, so this should not be necessary. To check the protein is expressed it would be necessary to tag it and detect the presence of the tag at the appropriate molecular weight by Western blot.

VTL4 is localised to the plasma membrane and infection thread membrane of cells in the nodule infection zone; the properties of these membranes in this context are not fully understood, nor the processes that drive proteins to localised to the infection thread membrane. There is potential for VTL4 to be mislocalised in yeast to a different membrane, preventing its effects from being detected in the context of these assays. In addition these were carried out when we predicted VTL4 to be localised to the vacuolar membrane, so these seemed like appropriate assays at the time.

Little is known about VTL orientation in the membrane and this could be another potential reason why no complementation was measured.

Finally, lack of complementation could be due to VTL4 not being able to transport Fe. Our current working hypothesis is that the loop region present in the VIT family proteins but missing in the VTL family is responsible for active Fe transport, and its removal render VTL only capable of acting as a "pressure valve" to prevent build-up of toxic Fe in the cytosol. This would mean it retains directionality but lacks the driving force necessary for active transport, explaining the lack of complementation. There is also the possibility that another plant protein is required for VTL4 to be an effective Fe transporter, although co-expression analysis did not suggest any likely candidates. Experiments to pin down the function of VTL4 and SEN1 are ongoing.

5.8.3 *vtl4* mutants have increased vacuoles in late infection/differentiation zone cells

The large vacuoles of differentiation zone cells in both *vtl4* mutants has proved somewhat puzzling; literature on increased vacuole size without a concomitant increase in cell size is sparse, making it difficult to hypothesise on the reason for this.

The vacuole is known to be a storage organelle for metals, to alleviate the toxic effects caused by their being in the cytosol (Sharma et al., 2016). In this chapter I have shown that VTL4 is localised to the plasma membrane of cells in the infection zone, in addition to the plasma membrane-derived infection thread membrane. From the expression pattern of ferritins, as well as X-ray fluorescence studies to localise metals within the nodule, it has been shown that there is high Fe in the infection zone, mainly sequestered in ferritin and the apoplast (Rodríguez-Haas et al., 2013). One plausible explanation for the larger vacuoles observed as standard in the infection zone of wild-type nodules would be if there were also iron sequestered there. We hypothesise that VTL4 functions as a "pressure valve" to aid Fe translocation from an area of high concentration (the cytosol) to an area of low concentration (the infection thread space). If this is the case, when NRAMP1 is active and cytosolic Fe is high, VTL4 will alleviate some of this high Fe stress

by removing Fe to the infection thread. This would result in less Fe being sequestered in the vacuole, decreasing the volume of fluid required to keep the Fe in solution in the vacuole, and so decreasing the vacuole size. In *vtl4* mutants by contrast, this equalisation could not occur in the infection zone (where *VTL4* would be maximally expressed), resulting in a large amount of Fe still being stored in the (larger) vacuoles in cells of the differentiation zone. This would be further augmented by the presence of VTL4 at the plasma membrane of infection zone cells; assuming that VTL4 is capable of unidirectional translocation, in order that it is orientated correctly to transport Fe into infection threads, it must effectively be pointing out of the cytosol, so the plasma membrane-localised VTL4 could also be removing excess Fe into the apoplast.

Another hypothesis is that the bacteroids are elongated more slowly than in the wild type so are shorter, with the plant vacuole increasing in size to make up for this bacterial length deficit. However, other than the large vacuole phenotype, bacteroid development appeared to be unaltered in *vtl4* mutant nodules.

5.8.4   *VTL4* and *SEN1* failed to complement the 13U mutant

Having seen the tantalising preliminary data, it initially seemed like there was some partial complementation of the 13U mutant both by the reintroduction of *SEN1* alone and in combination with *VTL4*. Unfortunately, this measurement could not be repeated, even when optimising the conditions of the experiment to incorporate more plants into a smaller jar, which boosted the percentage of acetylene reduced significantly in the case of the wild type. Hairy root transformation of Medicago is very inefficient, so it could simply be that the first time I tested this experiment I happened to have higher transformation efficiency than when it was repeated. This could easily be monitored by checking dsRed expression of transformed roots before carrying out the assay. In addition, the levels of *VTL4* and *SEN1* expression in could be assessed using qRT-PCR.

Another potential reason why addition of *SEN1* and *VTL4* did not complement the 13U phenotype could be that the phenotype is due to one of the other genes deleted in 13U. Although none of the other VTLs in deletion region are expressed in nodules, Medtr4g094338, a hypothetical protein of unknown function, is highly expressed in nodules. To establish whether lack of this gene is causing the 13U phenotype, further complementation experiments incorporating this gene would need to be carried out. It seems likely that lack of *SEN1* and *VTL4* is partially responsible for the phenotype, as the *sen1* mutant in *L. japonicus* is also a white nodule mutant. This Lotus *sen1* mutant was fully complemented by the full length wild type *SEN1* sequence (Hakoyama et al. 2012). An identical construct could be used to complement the 13U mutant; Lotus

promoters are normally functional in Medicago and this would eliminate the uncertainty due to the untested Medicago promoters.

Finally, although the HA and Myc tags on the C-terminus of VTL4 and SEN1 respectively are small, they could still result in the mislocalisation or inactivation of these proteins, rendering them incapable of complementing 13U.

5.8.5 Localisation experiments

To make a symbiosome membrane marker, the promoter and coding sequence of the Lotus sulphate transporter SST1 were synthesised and cloned into a binary vector for transformation into Medicago hairy roots. The SST1-eGFP fusion was localised to the plasma membrane of infection zone and differentiation zone cells, and there was no signal observed in more mature nodule cells. This is contrary to the expression pattern predicted for the Medicago orthologue in Roux et al. (2014), where it is shown to be maximally expressed in the nitrogen fixation zone of nodules. As there was no true symbiosome membrane in the zones where SST1-eGFP was expressed, it could not be used as a symbiosome membrane marker. In the differentiation zone cells expression SST-eGFP it is possible to see some evidence of localisation to early symbiosome membrane, but the signal is relatively weak, making it difficult to draw firm conclusions from these data.

Mislocalisation of this marker could have been due to the four mutations introduced into the promoter region during the domestication process to destroy the Golden Gate sites. These mutations were two G to C mutations and two C to G mutations, to remove two *Bbs*I sites, one *Bsa*I site and one *BsmB*I site. Alternatively, there could be *cis*-acting elements located in the introns or 3' UTR of SST1 which were omitted during the cloning process. Although I considered including them, there are 11 introns in SST1, one of which alone is 1.3 kilobases. There could also be subtly different promoter elements in Lotus than Medicago that could mean a slight shift in expression of SST1. As previously discussed, there is no symbiosome membrane localisation motif; it is purely controlled by timing of expression. Thus, any change in the time of expression caused by a slight discrepancy between Lotus and Medicago promoter elements could result in mislocalisation to the plasma membrane rather than the symbiosome membrane.

Although AtPIP2A is well characterised in Arabidopsis as a plasma membrane marker it has not been confirmed as such in Medicago, particularly in the unusual context of nodules. FM™ 4-64 is a fluorescent plasma membrane dye that could be used to check plasma membrane localisation of AtPIP2A in nodules, simultaneously confirming both

the efficacy of the marker and the localisation of VTL4 to the plasma membrane in the nodule infection zone.

5.8.6 Conclusions

The significant upregulation of both *SEN1* and *VTL4* in nodules, along with their subtly differing expression patterns suggests that they both play important and potentially interlinked roles in the developing nodule. *vtl4* mutants exhibit an increase in the number of white nodules produced, and 13U, a deletion mutant lacking both *VTL4* and *SEN1* only produces white nodules that are incapable of fixing nitrogen. VTL4 was found to be localised to the plasma membrane and infection thread membrane of cells in the nodule infection zone, while SEN1 was found to be localised to the symbiosome membrane in the differentiation zone, which corresponds to the expression patterns of *VTL4* and *SEN1* genes in Roux et al. (2014). This is summarised in Table 5.1.

Table 5.1 Summary of VTL4 and SEN1 localisation experiments

Construct	Nodule zone	Localisation
MtVTL4-FL3-mCherry	Infection	Plasma membrane Infection thread membrane
MtSEN1-FL3-mCherry	Early Infection	Plasma membrane
	Late Infection	Plasma membrane
	Late Differentiation	Symbiosome membrane

When tested with a novel bacterial Fe sensor, endosymbionts both in *vtl4* and 13U nodules were found to be exposed to much lower concentrations of Fe, suggesting that these genes are necessary for high levels of Fe to reach the developing bacteria. From these data, I hypothesise that the role of VTL is in Fe delivery to the developing infection thread, while SEN1 is involved in the transport of larger amounts of Fe across the symbiosome membrane as the bacteroids develop and cofactor synthesis comes into play. Despite this, attempts to complement the 13U mutant with *VTL4* and *SEN1* independently and in combination were unsuccessful, suggesting that another gene is at least partially responsible for the white nodule phenotype in this mutant. Finally, the effects of a mutation in the bacterial Fe exporter gene *mbfA* were assessed both in terms of nodulation of wild-type plants and bacteroid development within nodules. Despite the bacterial mutation resulting in a higher proportion of white nodules, bacteroid development was not affected.

Together these results have allowed us to generate a new model for Fe transport in specific nodule zones, supplementing the published literature to give a better idea of how Fe moves around the nodule to allow normal development.

6 General Discussion

6.1 Summary of key findings

Despite the clear importance of Fe in the legume-rhizobial symbiosis, relatively little is known both about Fe transport to the nodule and its route to the sites of cofactor biosynthesis within the nodule. In this study, I aimed to identify genes that play a key role in nodule iron homeostasis, using a reverse genetics approach in *Medicago truncatula*. I identified and isolated mutants in the major nodule ferrochelatase (*FECH1B*) and the putative Fe transporter *VTL4*. I then characterised mutants in these genes (*fech1b-1, vtl4-1* and *vtl4-2*), along with the 13U mutant, which has a 30 kilobase deletion encompassing *VTL4* and *SEN1,* another putative Fe transporter in the VTL family.

fech1b-1 mutant plants produced fewer pink nodules than the wild type, however the pink nodules that were produced were morphologically similar to the wild type in most respects. It is likely that FECH1A is able to compensate in part for the lack of FECH1B, meaning some nodules progress to maturity. FECH1B was found to be localised exclusively to the plastids, consistent with localisation data for Type I ferrochelatases in other plants (Cornah et al., 2002; Lister et al., 2001; van Lis et al., 2005).

The 13U mutation resulted in Medicago plants that only produced fix- white nodules as colonisation was aborted around the nodule differentiation zone (Domonkos et al., 2013). *vtl4* plants displayed an intermediate phenotype, producing more white nodules than wild-type plants, but still producing a similar number of pink nodules. In addition, *vtl4* nodules showed the novel phenotype of enlarged vacuoles in the differentiation zone. A novel bacterial Fe sensor construct, $P_{mbfA}lux$, allowed us to infer the amount of Fe perceived by *S. meliloti* in symbiotic nodules. When *vtl4* and 13U plants were infected with *S. meliloti* carrying this reporter plasmid, bacteria in 13U nodules had only 7% lux activity of wild-type nodules. The *vtl4* mutant nodules again showed an intermediate phenotype, with 36% (*vtl4*-2) and 50% (*vtl4*-1) lux activity of wild-type control.

SEN1-mCherry under its native promoter localised to the symbiosome membrane in the late infection zone and interzone. VTL4 localized to the plasma and infection thread membranes of the infection zone. Together these data suggests that SEN1 and VTL4 are important for Fe delivery to the bacteria, and given their high sequence homology to known Fe transporters, is strong evidence that they deliver Fe to the bacteria at different stages of the infection process.

Technical aspects of the data have been discussed in the relevant chapter discussions (Sections 3.7, 4.6 and 5.8). In the final discussion I will focus on the use of nodules as a

model for haem biosynthesis, and the possible function of the VTL family of iron transporters.

6.2   Nodules as a model for haem biosynthesis in plants

Although haem biosynthesis is conserved across land plants, the levels of haem and biosynthetic intermediates are extremely low and difficult to detect. By contrast, much has been learned from red blood cells in mammalian cells, where haem synthesis is strongly induced. Generally, it is necessary for haem biosynthesis to be tightly regulated; nodules, as the highest haem-producing organs in plants, provide a new and interesting experimental model for haem biosynthesis and its regulation.

The first point of regulation is GluTR, which catalyses GSA formation, controlling flux into the tetrapyrrole pathway. In photosynthetically active tissues, FLU is a negative regulator of the GluTR, encoded by *HEMA1* (Goslings et al., 2004). It came as a surprise that in non-photosynthetic nodules *HEMA1* (rather than *HEMA2* for example) was the most highly expressed *GluTR* gene, but also *FLU* (Table 3.3, Figure 4.1, Roux et al., 2014). *FLU* expression was approximately a tenth of the total expression of the *GluTR* genes, but interestingly its pattern of expression was very similar to that of *HEMA1, FECH1B* and a number of other enzymes in the pathway. It is not known how *FLU* expression is regulated, but this apparent co-expression could suggest that it is somehow linked to flux through, or upregulation of, the pathway.

HEMA1 is also subject to feedback inhibition by haem *b* (protohaem), (Huang and Wang, 1986; Goslings et al., 2004) to prevent the build-up of photosensitive tetrapyrrole intermediates in plant tissues when they are not needed. In order to circumvent this in nodules, it follows that there must either be spatial separation of the HEMA1 protein and ferrochelatase, or rapid efflux of the protohaem end product. FECH1B was localised to unknown substructures in the core of plastids; it would be interesting to compare this to that of HEMA1, the localisation of which is not defined. This could provide some information on a sub-organellar level as to how this separation might occur. Alternatively, there could be transfer of intermediates between plastids via the stromules, perhaps leading to plastids specialised in different stages of haem biosynthesis. There is evidence of complexation of enzymes in haem synthesis in other kingdoms, allowing intermediates to be funnelled between adjacent enzymes in the pathway (Lüer et al., 2005; Medlock et al., 2015), and it has also been tentatively suggested (based on the crystal structures of tobacco PPOX and human ferrochelatase) that these two enzymes could form a complex (Koch et al., 2004). This could potentially be extended to these protein complexes shuttling between plastids through the stromules, which are known to be capable of transporting proteins such as GFP between plastids (Köhler et al., 1997). The most likely of these options is the rapid export of haem for incorporation into

leghaemoglobin, and it should be possible to identify a transporter for this using a reverse genetics approach.

In most plant species, the majority of flux through the tetrapyrrole pathway is for chlorophyll biosynthesis. Along with chloroplast development, chlorophyll biosynthesis is largely light-regulated. In the dark, however, other regulatory mechanisms are thought to play more of a role, one of which is hormonal regulation. Cytokinins have been shown to upregulate GATA family transcription factors which activate transcription of a number of chlorophyll biosynthetic genes, including *HEMA1* (Hudson et al., 2011). The high levels of cytokinins associated with nodule formation could be activating *HEMA1* transcription in nodules, so increasing flux into the pathway. Chlorophyll synthesis can also be decoupled from light-dependence through DELLA interacting with the PIFs responsible for gene repression in the dark (Cheminant et al., 2011). Although this decoupling has not been extended to haem synthesis, the use of the usually light-responsive HEMA1 isoform in nodules points to this as a viable route of activation. DELLA proteins (that regulate gibberellic acid signalling) are already known to be important in nodulation as they interact with CCaMK in the early signalling stages of the pathway. *della* mutants showed impaired colonisation (Jin et al., 2016), which could also be due to haem synthesis not being as upregulated as in wild-type nodules, thus preventing sufficient leghaemoglobin levels for a successful symbiosis.

Although there are significant differences in the oxygen-binding properties of plant leghaemoglobins and mammalian haemoglobins (Kundu and Hargrove, 2003), the haem moiety is synthesised through similar pathways, with the main difference being how ALA is generated (Barupala et al., 2016). The majority of haem pathway regulation is at this point in both plants and mammals, with some interesting similarities. There are two ALA synthase (ALAS) isoforms in humans: ALAS1 is expressed in the liver, where haem is required for metabolic enzymes; and ALAS2 in the bone marrow, where haem is synthesised in red blood cells. Like HEMA1 in plants, ALAS1 is regulated in part by feedback inhibition by haem (Hamilton et al., 1991; Roberts et al., 2005). ALAS1 transcription is controlled by transcription factors in the GATA family, akin to the light-dependent regulation of HEMA1 in plants. In humans, these GATA transcription factors co-ordinate haem synthesis with red blood cell development and iron availability to ensure even supply and demand at all times (Wingert et al., 2005; Chung et al., 2017). The GATA transcription factors in humans also activate transcription of a protein kinase responsible for activating ferrochelatase (Chung et al., 2017). The role of phosphorylation of plant enzymes involved in tetrapyrrole biosynthesis is in its infancy, but this would provide another layer of regulation.

6.3 The function of the VTL family of proteins in plants

As previously described in Sections 3.4.3 and 5.1.2, the main difference between the VITs and the VTLs is the presence of a large, cytosolic loop between transmembrane helices 2 and 3 in the former (Figure 5.1). Arabidopsis *VIT1* is capable of functionally complementing the Fe-sensitive phenotype of yeast Δ*ccc1* (Kim et al., 2006), while *VTL4* is not (Figure 5.12).

One working hypothesis is that the loop region of VIT1 is necessary for active transport of Fe across the tonoplast membrane, while the truncated loop in the VTLs merely confers directionality to a passive channel. The VIT orthologue in the malaria parasite *Plasmodium falciparum* has recently been shown to function as an Fe^{2+}/H^+ antiporter, and like AtVIT1 is functional when expressed heterologously (Labarbuta et al., 2017). We hypothesise that the elongated loop region of VITs mediates this Fe^{2+}/H^+ antiport through an as-yet-uncharacterised mechanism. Without the loop, VTLs can only act as a unidirectional "pressure valve", equalising Fe concentrations across a membrane. This could be tested by expressing VTL proteins in artificial membrane systems and measuring their ability to transport radiolabelled Fe into resulting vesicles. When VTL4 is localised to the plasma membrane of early infection zone cells and thought to be pointing "out" of the cell it will be having little effect. By contrast when it is localised to the infection thread membrane (and thought to be pointing into the infection thread) there is likely to be more iron in the nodule cell than the infection thread, due to the high Fe demand of the bacteria. This would result in Fe passing through VTL4, down the concentration gradient. The observed similarity between infection thread and pollen tube growth is particularly interesting here as mislocalised Fe in the style results in arrested pollen tubes (Schuler et al., 2012). It can be hypothesised that insufficient Fe in the infection zone cells would have a similar effect on infection thread progression.

Plant vacuoles are known to be slightly acidic compartments, with a pH of around 5.5, which generates the proton-motive force necessary for transport across the tonoplast membrane (Mathieu et al., 1989). The symbiosome space in Medicago nodules has been shown to be even more acidic, with a pH between 4.5 and 5.0 (Pierre et al., 2013). This is interesting both in the context of the increased solubility of Fe at low pH, and also the proton-motive force available to drive nutrient uptake across the symbiosome membrane. If a VIT were the primary Fe transporter of the symbiosome membrane this uptake could proceed too quickly, driven by such a high proton gradient, resulting in oxidative stress to the bacteria. However, if the truncated loop region in VTL proteins reduces their affinity for protons this could allow tuning of Fe uptake mechanisms across the symbiosome (and other) membranes to regulate Fe influx at different pHs. Finally,

low pH-dependence could explain why the Arabidopsis VTLs are only capable of weakly complementing the yeast Δ*ccc1* mutant, as the yeast vacuole has a higher pH of around 6.2 (Preston et al., 1989; Gollhofer et al., 2014).

One possibility that I explored is that VTLs require a partner protein for active transport. In particular, MbfA proteins consist of a ferritin-like domain fused to a full VTL sequence. Removal or disruption of this ferritin-like domain of MbfA renders it incapable of Fe transport (Sankari and O'Brian, 2014; Simon Andrews, personal communication). Thus, it is possible that VTL proteins co-operate with ferritin or another protein to render them capable of active transport. *VTL* genes are only found in the embyrophytes (land plants) and Chlamydomonas; interestingly the genes have been lost in other chlorophytes such as *Ostreococcus spp.*. The need for a plant-specific partner protein would be another explanation for the lack of complementation of the yeast Δ*ccc1* mutant, in addition to low pH-dependence. Unfortunately, co-expression analysis did not yield any likely candidate partners, but this hypothesis should not be discounted completely in future studies.

6.4 Updated model of nodule iron transport and usage

Finally, the results discussed above, in combination with studies from other research groups, have led to the refinement of our initial model (Figure 1.5) to the current model of Fe uptake and transport in nodules (Figure 6.1).

Fe is initially taken up via the roots of infected plants, and transported to the nodule through the vasculature. In the early infection zone Fe is imported into the cell and subsequently the plastids by as-yet-unidentified transporters, and stored in ferritin (Lucas et al., 1998; Rodríguez-Haas et al., 2013). As the infection thread penetrates the cell, Fe is supplied to the bacteria within it through the VTL4 transporter (Section 5.5.5, Figure 5.18).

When the infection thread reaches its destination cell in the infection zone the bacteria begin to differentiate and elongate as they transmogrify into the nitrogen-fixing state. At this point, Fe is taken up across the plasma membrane of infected cells by NRAMP1 (Tejada-Jiménez et al., 2015).

In the late infection and interzone, a large amount of Fe is required for haem synthesis in the plastids to ensure leghaemoglobin is furnished with sufficient cofactor to lower the oxygen partial pressure in the cell. At this stage the bacteria also have a high demand for Fe, as low oxygen activates the suite of *Nif* genes responsible for nitrogenase cofactor biosynthesis, which accounts for the major part of the bacterial demand for Fe. In the late differentiation/interzone, bacteroids are individually encased in symbiosome membrane, the site of SEN1 localisation (Section 5.5.4, Fig 5.17). Fe can pass through SEN1 to the symbiosome space where it can be taken up by the bacteroids.

If Fe levels within the bacteroids reach toxic levels, (as appears to be the case in the late infection and interzone) transcription of *mbfA* is activated (Roux et al., 2014). MbfA can then pump Fe out of the bacteroids and back into the symbiosome space, so relieving some of the Fe stress.
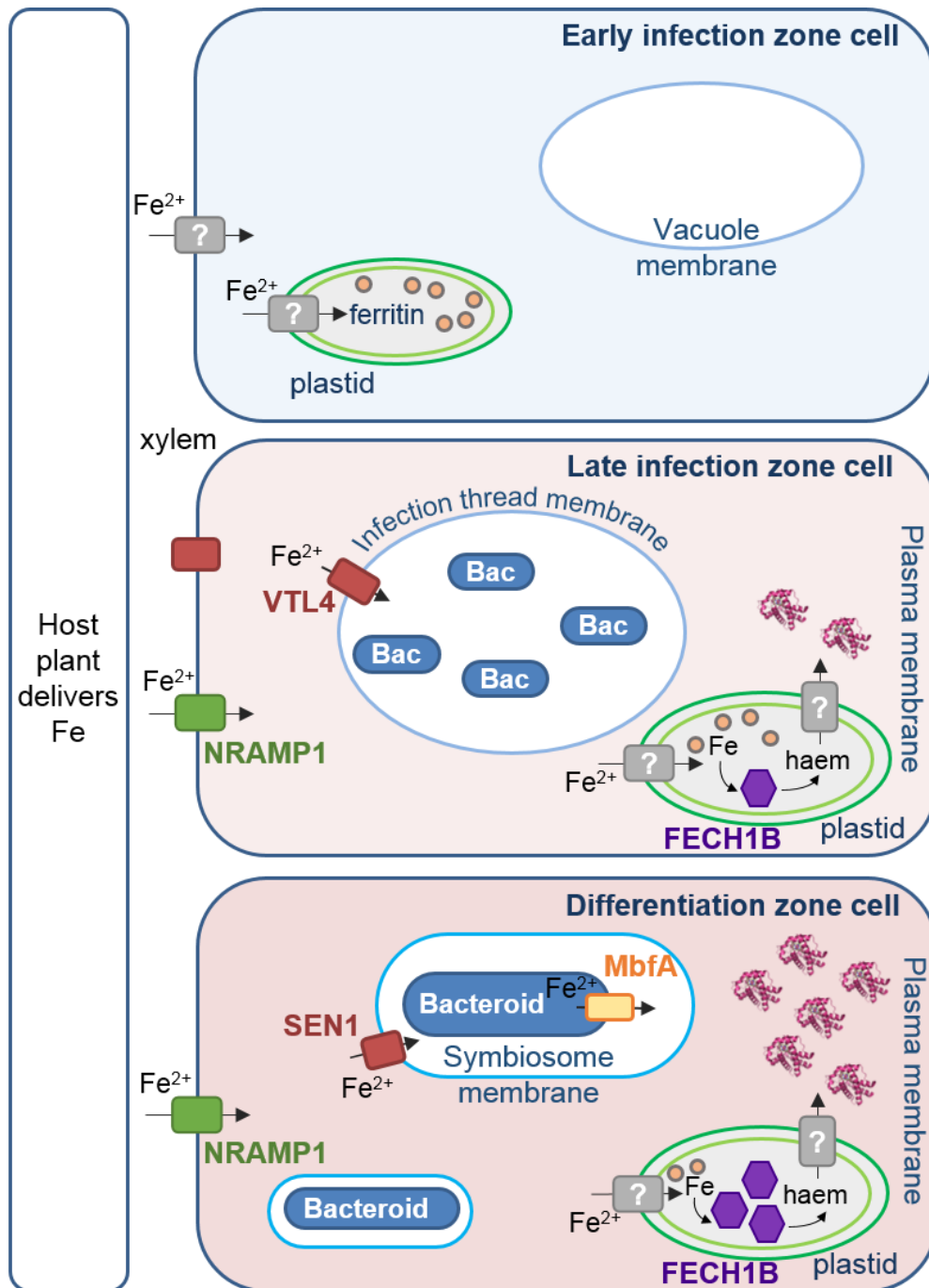
Figure 6.1 Model of Fe transport, storage and usage in the developing nodule Model based on literature and the findings of this study (Roux et al. 2014, Tejada-Jimenez et al. 2015, Rodriguez-Haas et al. 2013). Leghaemoglobin structures taken from PDB file 2GDM, using PyMol molecular modelling software.

6.5 Concluding remarks

The symbiotic nodule has proved to be a fascinating (if challenging) system for studying both iron transport and haem cofactor biosynthesis. In the last four years, there have been significant advances both in our lab and in the field as a whole, particularly with respect to the metal transporters involved in nodule development. The research carried out in this study has uncovered new lines of questioning, helping to pave the way for future researchers picking up the gauntlet of this exciting area of study.

7 References

Abadia, J., López-Millán, A.-F., Rombola, A., and Abadía, A. (2002). Organic acids and Fe deficiency : a review. Plant Soil **241**: 75–86.

Abreu, I., Saéz, Á., Castro-rodríguez, R., Escudero, V., Rodriguez-Haas, B., Senovilla, M., Larue, C., Grolimund, D., Tejada-Jiménez, M., Imperial, J., and González-Guerrero, M. (2017). *Medicago truncatula* Zinc-Iron Permease6 provides zinc to rhizobia-infected nodule cells. Plant, Cell Environ. DOI: 10.1111/pce.13035

Aleman, L., Ortega, J.L., Martinez-Grimes, M., Seger, M., Holguin, F.O., Uribe, D.J., Garcia-Ibilcieta, D., and Sengupta-Gopalan, C. (2010). Nodule-enhanced expression of a sucrose phosphate synthase gene member (*MsSPSA*) has a role in carbon and nitrogen metabolism in the nodules of alfalfa (*Medicago sativa* L.). Planta **231**: 233–244.

Andaluz, S., Rodríguez-Celma, J., Abadía, A., Abadía, J., and López-Millán, A.F. (2009). Time course induction of several key enzymes in Medicago truncatula roots in response to Fe deficiency. Plant Physiol. Biochem. **47**: 1082–1088.

Askwith, C., Eide, D., Van Ho, A., Bernard, P.S., Li, L., Davis-Kaplan, S., Sipe, D.M., and Kaplan, J. (1994). The *FET3* gene of S. cerevisiae encodes a multicopper oxidase required for ferrous iron uptake. Cell **76**: 403–410.

Avenhaus, U., Cabeza, R.A., Liese, R., Lingner, A., Dittert, K., Salinas-Riester, G., Pommerenke, C., and Schulze, J. (2016). Short-Term Molecular Acclimation Processes of Legume Nodules to Increased External Oxygen Concentration. Front. Plant. Sci. **6**: 1–13.

Balk, J. and Schaedler, T. A. (2014). Iron cofactor assembly in plants. Annu. Rev. Plant Biol. **65**: 125–53.

Barberon, M. (2017). The endodermis as a checkpoint for nutrients. New Phytol. **213**: 1604–1610.

Barney, B.M., Igarashi, R.Y., Santos, P.C. Dos, Dean, D.R., and Seefeldt, L.C. (2004). Substrate Interaction at an Iron-Sulfur Face of the FeMo-cofactor during Nitrogenase Catalysis J. Biol. Chem. **279**: 53621–53624.

Barupala, D.P., Dzul, S.P., Riggs-gelasco, P.J., and Stemmler, T.L. (2016). Synthesis, delivery and regulation of eukaryotic heme and Fe–S cluster cofactors. Arch. Biochem. Biophys. **592**: 60–75.

Bashir, K., Ishimaru, Y., Shimo, H., Nagasaka, S., Fujimoto, M., Takanashi, H., Tsutsumi, N., An, G., Nakanishi, H., and Nishizawa, N.K. (2011). The rice mitochondrial iron transporter is essential for plant growth. Nat. Commun. **2**: 1–7.

Benedito, V.A., Torres-Jerez, I., Murray, J.D., Andriankaja, A., Allen, S., Kakar, K., Wandrey, M., Verdier, J., Zuber, H., Ott, T., Moreau, S., Niebel, A., Frickey, T., Weiller, G., He, J., Dai, X., Zhao, P.X., Tang, Y., and Udvardi, M.K. (2008). A gene expression atlas of the model legume *Medicago truncatula*. Plant J. **55**: 504–513.

Bergersen, F.J. and Appleby, C.A. (1981). Leghaemoglobin within bacteroid-enclosing membrane envelopes from soybean root nodules. Planta **152**: 534–543.

Bernal, M., Casero, D., Singh, V., Wilson, G.T., Grande, A., Yang, H., Dodani, S.C., Pellegrini, M., Huijser, P., Connolly, E.L., Merchant, S.S., and Krämer, U. (2012) Transcriptome Sequencing Identifies *SPL7*-Regulated Copper Acquisition Genes *FRO4/FRO5* and the Copper Dependence of Iron Homeostasis in *Arabidopsis*. Plant Cell **24**: 738–761.

Bernard, D.G., Cheng, Y., Zhao, Y., and Balk, J. (2009). An Allelic Mutant Series of *ATM3* Reveals Its Key Role in the Biogenesis of Cytosolic Iron-Sulfur Proteins in Arabidopsis. Plant Physiol. **151**: 590–602.

Bernard, D.G., Netz, D.J.A., Lagny, T.J., Pierik, A.J., and Balk, J. (2013). Requirements of the cytosolic iron–sulfur cluster assembly pathway in Arabidopsis. Philos Trans R Soc L. B Biol Sci **368**: 1–10.

Bhubhanil, S., Chamsing, J., Sittipo, P., Chaoprasid, P., Sukchawalit, R., and Mongkolsuk, S. (2014). Roles of *Agrobacterium tumefaciens* membrane-bound ferritin (MbfA) in iron transport and resistance to iron under acidic conditions. Microbiology **160**: 863–871.

Bonneau, J., Baumann, U., Beasley, J., Li, Y., and Johnson, A.A.T. (2016). Identification and molecular characterization of the nicotianamine synthase gene family in bread wheat. Plant Biotechnol. J. **14**: 2228–2239.

Breakspear, A., Liu, C., Roy, S., Stacey, N., Rogers, C., Trick, M., Morieri, G., Mysore, K.S., Wen, J., Oldroyd, G.E.D., Downie, J.A., and Murray, J.D. (2014). The root hair "infectome" of *Medicago truncatula* uncovers changes in cell cycle genes and reveals a requirement for Auxin signaling in rhizobial infection. Plant Cell **26**: 4680–701.

Brear, E.M., Day, D.A., and Smith, P.M.C. (2013). Iron: an essential micronutrient for the legume-rhizobium symbiosis. Front. Plant Sci. **4**: 1–15.

Bughio, N., Yamaguchi, H., Nishizawa, N.K., Nakanishi, H., and Mori, S. (2002). Cloning an iron-regulated metal transporter from rice. **53**: 1677–1682.

Cannon, S.B., Sterck, L., Rombauts, S., Sato, S., Cheung, F., Gouzy, J., Wang, X., Mudge, J., Vasdewani, J., Schiex, T., Spannagl, M., Monaghan, E., Nicholson, C., Humphray, S.J., Schoof, H., Mayer, K.F.X., Rogers, J., Quétier, F., Oldroyd, G.E.D., Debellé, F., Cook, D.R., Retzel, E.F., Roe, B.A., Town, C.D., Tabata, S., Van de Peer, Y., and Young, N.D. (2006). Legume genome evolution viewed through the *Medicago truncatula* and *Lotus japonicus* genomes. Proc. Natl. Acad. Sci. **103**: 14959–14964.

Castaings, L., Caquot, A., Loubet, S., and Curie, C. (2016). The high-affinity metal Transporters NRAMP1 and IRT1 Team up to Take up Iron under Sufficient Metal Provision. Sci. Rep. **6**: 1–11.

Cerri, M.R., Wang, Q., Stolz, P., Folgmann, J., Frances, L., Katzer, K., Li, X., Heckmann, A.B., Wang, T.L., Downie, J.A., Klingl, A., de Carvalho-Niebel, F., Xie, F., and Parniske, M. (2017). The *ERN1* transcription factor gene is a target of the CCaMK/CYCLOPS complex and controls rhizobial infection in *Lotus japonicus*. New Phytol. **215**: 323–337.

Chao, T-C., Buhrmester, J., Hansmeier, N., Pühler, A., and Weidner, S. (2005). Role of the Regulatory Gene *rirA* in the Transcriptional Response of *Sinorhizobium meliloti* to Iron Limitation. Appl. Environ. Microbiol. **71**: 5969–5982.

Cheminant, S., Wild, M., Bouvier, F., Pelletier, S., Renou, J-P., Erhardt, M., Hayes, S., Terry, M.J., Genschik, P., and Achard, P. (2011). DELLAs Regulate Chlorophyll and Carotenoid Biosynthesis to Prevent Photooxidative Damage during Seedling Deetiolation in *Arabidopsis*. **23**: 1849–1860.

Cheng, N-H., Liu, J-Z., Brock, A., Nelson, R.S., and Hirschi, K.D. (2006). AtGRXcp, an *Arabidopsis* chloroplastic glutaredoxin, is critical for protection against protein oxidative damage. J. Biol. Chem. **281**: 26280–26288.

Chung, J., Wittig, J.G., Ghamari, A., Maeda, M., Dailey, T.A., Bergonia, H., Kafina, M.D., Coughlin, E.E., Minogue, C.E., Hebert, A.S., Li, L., Kaplan, J., Lodish, H.F., Bauer, D.E., Orkin, S.H., Cantor, A.B., Maeda, T., Phillips, J.D., Coon, J.J., Pagliarini, D.J., Dailey, H.A., Paw, B.H. (2017). Erythropoietin signaling regulates heme biosynthesis. eLife DOI: 10.7554/eLife.24767

Clarke, V.C., Loughlin, P.C., Gavrin, A., Chen, C., Brear, E.M., Day, D.A., and Smith, P.M.C. (2015). Proteomic Analysis of the Soybean Symbiosome Identifies New Symbiotic Proteins. Mol. Cell. Proteomics **14**: 1301–1322.

Connorton, J.M., Balk, J., and Rodríguez-Celma, J. (2017a). Iron homeostasis in plants – a brief overview. Metallomics **9**: 813–823.

Connorton, J.M., Jones, E.R., Rodríguez-Ramiro, I., Fairweather-Tait, S., Uauy, C., and Balk, J. (2017b). Vacuolar Iron Transporter TaVIT2 transports Fe and Mn and is effective for biofortification. Plant Physiol. **174**: 2434–2444

Cornah, J.E., Roper, J.M., Singh, D.P., and Smith, A.G. (2002). Measurement of ferrochelatase activity using a novel assay suggests that plastids are the major site of haem biosynthesis in both photosynthetic and non-photosynthetic cells of pea (*Pisum sativum* L.). Biochem. J. **362**: 423–432.

Couturier, J., Touraine, B., Briat, J-F., Gaymard, F., and Rouhier, N. (2013). The iron-sulfur cluster assembly machineries in plants: current knowledge and open questions. Front. Plant Sci. **4**: 259.

Curie, C., Panaviene, Z., Loulergue, C., Dellaporta, S.L., Briat, J-F., and Walker, E.L. (2001). Maize *yellow stripe1* encodes a membrane protein directly involved in Fe (III) uptake. **409**: 346–349.

Cutler, S.R., Ehrhardt, D.W., Griffitts, J.S., and Somerville, C.R. (2000). Random GFP :: cDNA fusions enable visualization of subcellular structures in cells of *Arabidopsis* at a high frequency. Proc. Natl. Acad. Sci. U.S.A. **97**: 3718–3723.

Danyal, K., Dean, D.R., Hoffman, B.M., and Seefeldt, L.C. (2011). Electron transfer within nitrogenase: Evidence for a deficit-spending mechanism. Biochemistry **50**: 9255–9263.

David, M., Daveran, M.L., Batut, J., Dedieu, A., Domergue, O., Ghai, J., Hertig, C., Boistard, P., and Kahn, D. (1988). Cascade Regulation of *nif* Gene-Expression in Rhizobium-Meliloti. Cell **54**: 671–683.

Deyholos, M.K. and Sieburth, L.E. (2000). Separable Whorl-Specific Expression and Negative Regulation by Enhancer Elements within the *AGAMOUS* Second Intron. Plant Cell **12**: 1799–1810.

DiDonato Jr, R.J., Roberts, L.A., Sanderson, T., Eisley, R.B., and Walker, E.L. (2004). Arabidopsis Yellow Stripe-Like2 (*YSL2*): a metal-regulated gene encoding a plasma membrane transporter of nicotianamine – metal complexes. Plant J. **2**: 403–414.

Di Tommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J-M., Taly, J-F., and Notredame, C. (2011). T-Coffee : a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. Nuc. Acid Res. **39**: 13–17.

Divol, F., Couch, D., Conéjéro, G., Roschzttardtz, H., Mari, S., and Curie, C. (2013). The *Arabidopsis* YELLOW STRIPE LIKE4 and 6 Transporters Control Iron Release from the Chloroplast. Plant Cell **25**: 1040–1055.

Dixon, R. and Kahn, D. (2004). Genetic regulation of biological nitrogen fixation. Nat. Rev. Microbiol. **2**: 621–631.

Domonkos, A., Horvath, B., Marsh, J.F., Halasz, G., Ayaydin, F., Oldroyd, G.E.D., and Kalo, P. (2013). The identification of novel loci required for appropriate nodule development in *Medicago truncatula*. BMC Plant Biol. **13**: 157–168

Durrett, T.P., Gassmann, W., and Rogers, E.E. (2007). The FRD3-Mediated Efflux of Citrate into the Root Vasculature Is Necessary for Efficient Iron Translocation. Plant Physiol. **144**: 197–205.

Duval, S., Danyal, K., Shaw, S., Lytle, A.K., Dean, D.R., Hoffman, B.M., Antony, E., and Seefeldt, L.C. (2013). Electron transfer precedes ATP hydrolysis during nitrogenase catalysis. Proc. Natl. Acad. Sci. U.S.A **110**: 16414–9.

Duy, D., Stübe, R., Wanner, G., and Philippar, K. (2011). The chloroplast permease PIC1 regulates plant growth and development by directing homeostasis and transport of iron. Plant Physiol. **155**: 1709–22.

Eide, D., Broderius, M., Fett, J., and Guerinot, M.L. (1996). A novel iron-regulated metal transporter from plants identified by functional expression in yeast. Proc. Natl. Acad. Sci. U.S.A. **93**: 5624–5628.

Einsle, O., Tezcan, F.A., Andrade, S.L.A., Schmid, B., Yoshida, M., Howard, J.B., and Rees, D.C. (2002). Nitrogenase MoFe-Protein at 1.16 Å Resolution: A Central Ligand in the FeMo-Cofactor. Science **297**: 1696–1701.

Engler, C., Gruetzner, R., Kandzia, R., and Marillonnet, S. (2009). Golden gate shuffling: A one-pot DNA shuffling method based on Type IIs restriction enzymes. PLoS One **4**. DOI: 10.1371/journal.pone.0005553

Engler, C., Kandzia, R., and Marillonnet, S. (2008). A One Pot , One Step , Precision Cloning Method with High Throughput Capability. **3**. DOI: 10.1371/journal.pone.0003647

Fay, A.W., Blank, M.A., Rebelein, J.G., Lee, C.C., Ribbe, M.W., Hedman, B., Hodgson, K.O., and Hu, Y. (2016). Assembly scaffold NifEN: A structural and

functional homolog of the nitrogenase catalytic component. Proc. Natl. Acad. Sci. U.S.A. **113**: 9504–9508.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39**: 783–791.

Figurski, D.H. and Helinski, D.R. (1979). Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in *trans*. Proc. Natl. Acad. Sci. U.S.A. **76**: 1648–52.

Frazzon, A.P.G., Ramirez, M. V., Warek, U., Balk, J., Frazzon, J., Dean, D.R., and Winkel, B.S.J. (2007). Functional analysis of *Arabidopsis* genes involved in mitochondrial iron-sulfur cluster assembly. Plant Mol. Biol. **64**: 225–240.

Gollhofer, J., Schläwicke, C., Jungnick, N., Schmidt, W., and Buckhout, T.J. (2011). Members of a small family of nodulin-like genes are regulated under iron deficiency in roots of *Arabidopsis thaliana*. Plant Physiol. Biochem. **49**: 557–564.

Gollhofer, J., Timofeev, R., Lan, P., Schmidt, W., and Buckhout, T.J. (2014). Vacuolar-iron-transporter1-like proteins mediate iron homeostasis in arabidopsis. PLoS One **9** DOI: 10.1371/journal.pone.0110468.

Goslings, D., Meskauskiene, R., Kim, C., Lee, K.P., Nater, M., and Apel, K. (2004). Concurrent interactions of heme and FLU with Glu tRNA reductase (HEMA1), the target of metabolic feedback inhibition of tetrapyrrole biosynthesis, in dark- and light-grown Arabidopsis plants. Plant J. **40**: 957–967.

Gray, J.C., Sullivan, J.A., Hibberd, J.M., and Hansen, M.R. (2001). Stromules : Mobile Protrusions and Interconnections Between Plastids. Plant boil. **3**: 223–233.

Guerinot, M.L. and Yi, Y. (1994). Iron: Nutritious, Noxious, and Not Readily Available. Plant Physiol. **104**: 815–820.

Hakoyama, T., Niimi, K., Watanabe, H., Tabata, R., Matsubara, J., Sato, S., Nakamura, Y., Tabata, S., Jichun, L., Matsumoto, T., Tatsumi, K., Nomura, M., Tajima, S., Ishizaka, M., Yano, K., Imaizumi-Anraku, H., Kawaguchi, M., Kouchi, H., and Suganuma, N. (2009). Host plant genome overcomes the lack of a bacterial gene for symbiotic nitrogen fixation. Nature **462**: 514–517.

Hakoyama, T., Niimi, K., Yamamoto, T., Isobe, S., Sato, S., Nakamura, Y., Tabata, S., Kumagai, H., Umehara, Y., Brossuleit, K., Petersen, T.R., Sandal, N., Stougaard, J., Udvardi, M.K., Tamaoki, M., Kawaguchi, M.,Kouchi, H., and Suganuma, N. (2012). The integral membrane protein SEN1 is required for symbiotic nitrogen fixation in lotus japonicus nodules. Plant Cell Physiol. **53**: 225–236.

Hamilton, J.W., Bement, W.J., Sinclair, P.R., Sinclair, J.F., Alcedo, J.A., and Wetterhahn, K.E. (1991). Heme Regulates Hepatic 5-Aminolevulinate Synthase mRNA Expression by Decreasing mRNA Half-Life and Not by Altering Its Rate of Transcription. Arch. Biochem. Biophys. **289**: 387–392

Heazlewood, J.L., Tonti-Filippini, J.S., Gout, A.M., Day, D.A., Whelan, J., and Millar, A.H. (2004). Experimental Analysis of the Arabidopsis Mitochondrial Proteome Highlights Signaling and Regulatory Components, Provides Assessment of Targeting Prediction Programs, and Indicates Plant-Specific Mitochondrial Proteins. Plant Cell **16**: 241–256.

Herrada, G., Puppo, A., Moreau, S., Day, D.A., and Rigaud, J. (1993). How is leghemoglobin involved in peribacteroid membrane degradation during nodule senescence? FEBS Lett. **326**: 33–38.

Hey, D., Ortega-Rodes, P., Fan, T., Schnurrer, F., Brings, L., Hedtke, B., and Grimm, B. (2016). Transgenic Tobacco Lines Expressing Sense or Antisense *FERROCHELATASE 1* RNA Show Modified Ferrochelatase Activity in Roots and Provide Experimental Evidence for Dual Localization of Ferrochelatase 1. Plant Cell Physiol. **57**: 2576–2585

Höfgen, R., Axelsen, K.B., Kannangara, C.G., Schüttke, I., Pohlenz, H-D., Willmitzer, L., Grimm, B., and von Wettstein, D. (1994). A visible marker for antisense mRNA expression in plants: inhibition of chlorophyll synthesis with a glutamate-1-semialdehyde aminotransferase antisense gene. Proc Natl Acad Sci U.S.A. **91**: 1726–1730.

Huang, D-D. and Wang, W-Y. (1986). Chlorophyll biosynthesis in *Chlamydomonas* starts with the formation of glutamyl-tRNA. J. Biol. Chem. **261**: 13451–13455.

Hudson, D., Guevara, D., Yaish, M.W., Hannam, C., Long, N., Joseph, D., Bi, Y-M., and Rothstein, S.J. (2011). *GNC* and *CGA1* Modulate Chlorophyll Biosynthesis and Glutamate Synthase (*GLU1 / Fd-GOGAT*) Expression in *Arabidopsis*. **6**. DOI: 10.1371/journal.pone.0026765

Imperial, J., Hoover, T.R., Madden, M.S., Ludden, P.W., and Shah, V.K. (1989). Substrate Reduction Properties of Dinitrogenase Activated in Vitro Are Dependent upon the Presence of Homocitrate or Its Analogues during Iron–Molybdenum Cofactor Synthesis. Biochemistry **28**: 7796–7799.

Inoue, H., Higuchi, K., Takahashi, M., Nakanishi, H., Mori, S., and Nishizawa, N.K. (2003). Three rice nicotianamine synthase genes, *OsNAS1*, *OsNAS2*, and *OsNAS3* are expressed in cells involved in long-distance transport of iron and differentially regulated by iron. Plant J. **36**: 366–381.

Inoue, H., Kobayashi, T., Nozoye, T., Takahashi, M., Kakei, Y., Suzuki, K., Nakazono, M., Nakanishi, H., Mori, S., and Nishizawa, N.K. (2009). Rice OsYSL15 is an iron-regulated Iron (III)-deoxymugineic acid transporter expressed in the roots and is essential for iron uptake in early growth of the seedlings. J. Biol. Chem. **284**: 3470–3479.

Jacobs, J.M. and Jacobs, N.J. (1987). Oxidation of protoporphyrinogen to protoporphyrin, a step in chlorophyll and haem biosynthesis. Purification and partial characterization of the enzyme from barley organelles. Biochem. J. **244**: 219–224.

Jain, A., Wilson, G.T., and Connolly, E.L. (2014). The diverse roles of FRO family metalloreductases in iron and copper homeostasis. Front. Plant Sci. **5**: DOI: 10.3389/fpls.2014.00100

Jeong, J., Cohu, C., Kerkeb, L., Pilon, M., Connolly, E.L., and Guerinot, M.L. (2008). Chloroplast Fe(III) chelate reductase activity is essential for seedling viability under iron limiting conditions. Proc. Natl. Acad. Sci. U.S.A. **105**: 10619–10624.

Jeong, J. and Connolly, E.L. (2009). Iron uptake mechanisms in plants: Functions of the FRO family of ferric reductases. Plant Sci. **176**: 709–714.

Jin, H.K., Füzéry, A.K., Tonelli, M., Ta, D.T., Westler, W.M., Vickery, L.E., and Markley, J.L. (2009). Structure and dynamics of the iron-sulfur cluster assembly scaffold protein IscU and its interaction with the cochaperone HscB. Biochemistry **48**: 6062–6071.

Jin, Y., Liu, H., Luo, D., Yu, N., Dong, W., Wang, C., Zhang, X., Dai, H., Yang, J., and Wang, E. (2016). DELLA proteins are common components of symbiotic rhizobial and mycorrhizal signalling pathways.Nat. Commun. **7** DOI: 10.1038/ncomms12433

Johnston, A.W.B., Todd, J.D., Curson, A.R., Lei, S., Nikolaidou-Katsaridou, N., Gelfand, M.S., and Rodionov, D.A. (2007). Living without Fur: The subtlety and complexity of iron-responsive gene regulation in the symbiotic bacterium *Rhizobium* and other α-proteobacteria. BioMetals **20**: 501–511.

Kaiser, B.N., Moreau, S., Castelli, J., Thomson, R., Lambert, A., Bogliolo, S., Puppo, A., and Day, D.A. (2003). The soybean NRAMP homologue, GmDMT1, is a symbiotic divalent metal transporter capable of ferrous iron transport. Plant J. **35**: 295–304.

Kelly, S., Radutoiu, S., and Stougaard, J. (2017). Legume LysM receptors mediate symbiotic and pathogenic signalling. Curr. Opin. Plant Biol. **39**: 152–158.

Kim, S.A., Punshon, T., Lanzirotti, A., Li, L., Alonso, J.M., Ecker, J.R., Kaplan, J., and Guerinot, M. L. (2006). Localization of iron in *Arabidopsis* seed requires the vacuolar membrane transporter VIT1. Science **314**: 1295–1298.

Klatte, M., Schuler, M., Wirtz, M., Fink-Straube, C., Hell, R., and Bauer, P. (2009). The analysis of Arabidopsis nicotianamine synthase mutants reveals functions for nicotianamine in seed iron loading and iron deficiency responses. Plant Physiol. **150**: 257–71.

Koch, M., Breithaupt, C., Kiefersauer, R., Freigang, J., Huber, R., and Messerschmidt, A. (2004). Crystal structure of protoporphyrinogen IX oxidase: a key enzyme in haem and chlorophyll biosynthesis. EMBO J. **23**: 1720–1728.

Köhler, R.H., Cao, J., Zipfel, W.R., Webb, W.W., and Hanson, M.R. (1997). Exchange of Protein Molecules Through Connections Between Higher Plant Plastids. **276**: 2039–2043.

Komarova, N.Y., Meier, S., Meier, A., Grotemeyer, M.S., and Rentsch, D. (2012). Determinants for Arabidopsis Peptide Transporter Targeting to the Tonoplast or Plasma Membrane. Traffic **13**: 1090–1105.

Korshunova, Y.O., Eide, D., Clark, W.G., Guerinot, M. Lou, and Pakrasi, H.B. (1999). The IRT1 protein from *Arabidopsis thaliana* is a metal transporter with a broad substrate range. Plant Mol. Biol. **40**: 37–44.

Krusell, L., Krause, K., Ott, T., Desbrosses, G., Krämer, U., Sato, S., Nakamura, Y., Tabata, S., James, E.K., Sandal, N., Stougaard, J., Kawaguchi, M., Miyamoto, A., Suganuma, N., and Udvardi, M.K. (2005). The Sulfate Transporter SST1 Is Crucial for Symbiotic Nitrogen Fixation in *Lotus japonicus* Root Nodules. Plant Cell **17**: 1625–1636.

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol. Biol. Evol. **33**: 1870–1874.

Kundu, S. and Hargrove, M.S. (2003). Distal heme pocket regulation of ligand binding and stability in soybean leghemoglobin. Proteins Struct. Funct. Genet. **50**: 239–248.

Labarbuta, P., Duckett, K., Botting, C.H., Chahrour, O., Malone, J., Dalton, J.P., and Law, C.J. (2017). Recombinant vacuolar iron transporter family homologue PfVIT from human malaria-causing Plasmodium falciparum is a Fe^{2+}/H^+ exchanger.Sci. Rep. **7** DOI: 10.1038/srep42850.

Labes, M., Rastogi, V., Watson, R., and Finan, T.M. (1993). Symbiotic Nitrogen Fixation by a *nifA* Deletion Mutant of *Rhizobium meliloti* : the Role of an Unusual *ntrC* Allele. J. Bacteriol. **175**: 2662–2673.

Lancaster, K.M., Roemelt, M., Ettenhuber, P., Hu, Y., Ribbe, M.W., Neese, F., Bergmann, U., and DeBeer, S. (2011). X-ray Emission Spectroscopy Evidences a Central Carbon in the Nitrogenase Iron-Molybdenum Cofactor. **334**: 974–978.

Lanquar, V., Lelièvre, F., Bolte, S., Hamès, C., Alcon, C., Neumann, D., Vansuyt, G., Curie, C., Schröder, A., Krämer, U., Barbier-Brygoo, H., and Thomine, S. (2005). Mobilization of vacuolar iron by AtNRAMP3 and AtNRAMP4 is essential for seed germination on low iron. EMBO J. **24**: 4041–4051.

Lee, J-W. and Helmann, J.D. (2007). Functional specialization within the fur family of metalloregulators. BioMetals **20**: 485–499.

Leustek, T., Smith, M., Murillo, M., Singh, D.P., Smith, A.G., Woodcock, S.C., Awan, S.J., and Warren, M.J. (1997). Siroheme Biosynthesis in Higher Plants. J. Biol. Chem. **272**: 2744–2752.

Li, G., Wang, B., Tian, Q., Wang, T., and Zhang, W-H. (2014). *Medicago truncatula* ecotypes A17 and R108 differed in their response to iron deficiency. J. Plant Physiol. **171**: 639–647.

Limpens, E., Ivanov, S., van Esse, W., Voets, G., Fedorova, E., and Bisseling, T. (2009). *Medicago* N_2-fixing symbiosomes acquire the endocytic identity marker Rab7 but delay the acquisition of Vacuolar Identity. Plant Cell **21**: 2811–2828.

Lister, R., Chew, O., Rudhe, C., Lee, M-N., and Whelan, J. (2001). *Arabidopsis thaliana* ferrochelatase-I and -II are not imported into *Arabidopsis* mitochondria. FEBS Lett. **506**: 291–295.

Liu, C-W. and Murray, J.D. (2016). The Role of Flavonoids in Nodulation Host-Range. Plants **5** DOI: 10.3390/plants5030033

López-Millán, A-F., Ellis, D.R., and Grusak, M.A. (2004). Identification and characterization of several new members of the ZIP family of metal ion transporters in *Medicago truncatula*. Plant Mol. Biol. **54**: 583–596.

Lucas, M.M., Van De Sype, G., Hérouart, D., Hernández, M.J., Puppo, A., and de Felipe, M.R. (1998). Immunolocalization of ferritin in determinate and indeterminate legume root nodules. Protoplasma **204**: 61–70.

Lüer, C., Schauer, S., Möbius, K., Schulze, J., Schubert, W-D., Heinz, D.W., Jahn, D., and Moser, J. (2005). Complex formation between glutamyl-tRNA reductase and glutamate-1-semialdehyde 2,1-aminomutase in *Escherichia coli* during the initial reactions of porphyrin biosynthesis. J. Biol. Chem. **280**: 18568–18572.

Masuda, T., Goto, F., Yoshihara, T., and Mikami, B. (2010). Crystal structure of plant ferritin reveals a novel metal binding site that functions as a transit site for metal transfer in ferritin. J. Biol. Chem. **285**: 4049–4059.

Masuda, T., Suzuki, T., Shimada, H., Ohta, H., and Takamiya, K. (2003). Subcellular localization of two types of ferrochelatase in cucumber. Planta **217**: 602–609.

Mathieu, Y., Guern, J., Kurkdjian, A., Manigault, P., Manigault, J., Zielinska, T., Gillet, B., Beloeil, J-C., and Lallemand, J-Y. (1989). Regulation of Vacuolar pH of Plant Cells. Plant Physiol. **89**: 19–26.

Medlock, A.E., Shiferaw, M.T., Marcero, J.R., and Vashisht, A.A., Wohlschlegel, J.A., Phillips, J.D., and Dailey, H.A. (2015). Identification of the Mitochondrial Heme Metabolism Complex. PLoS ONE **10** DOI: 10.1371/journal.pone.0135896

Meskauskiene, R., Nater, M., Goslings, D., Kessler, F., op den Camp, R., and Apel, K. (2001). FLU: A negative regulator of chlorophyll biosynthesis in *Arabidopsis thaliana*. October **98**: 12826–12831.

Miller, J.B., Pratap, A., Miyahara, A., Zhou, L., Bornemann, S., Morris, R.J., and Oldroyd, G.E.D. (2013). Calcium / Calmodulin-Dependent Protein Kinase Is Negatively and Positively Regulated by Calcium, Providing a Mechanism for Decoding Calcium Responses during Symbiosis Signaling. Plant Cell **25**: 5053–5066.

Mohanty, B.K. and Kushner, S.R. (2011). Bacterial/archaeal/organellar polyadenylation. Wiley Interdiscip. Rev. RNA **2**: 256–276.

Momonoi, K.M., Tsuji, T.T., Kazuma, K.K., and Yoshida, K. (2012). Specific Expression of the Vacuolar Iron Transporter , TgVit , Causes Iron Accumulation in Blue-Colored Inner Bottom Segments of Various Tulip Petals. Biosci. Biotechnol. Biochem. **76**: 319–325.

Moretti, S., Armougom, F., Wallace, I.M., Higgins, D.G., Jongeneel, C. V, and Notredame, C. (2007). The M-Coffee web server : a meta-method for computing multiple sequence alignments by combining alternative alignment methods. Nucleic Acids Res. **35**: 645–648.

Morrissey, J., Baxter, I.R., Lee, J., Li, L., Lahner, B., Grotz, N., Kaplan, J., Salt, D.E., and Guerinot, M.L. (2009). The ferroportin metal efflux proteins function in iron and cobalt homeostasis in *Arabidopsis.* Plant Cell **21**: 3326–3338.

Moser, J., Schubert, W-D., Beier, V., Bringemeier, I., Jahn, D., and Heinz, D.W. (2001). V-shaped structure of glutamyl-tRNA reductase, the first enzyme of tRNA-dependent tetrapyrrole biosynthesis. EMBO J. **20**: 6583–6590.

Mukherjee, I., Campbell, N.H., Ash, J.S., and Connolly, E.L. (2006). Expression profiling of the Arabidopsis ferric chelate reductase (*FRO*) gene family reveals differential regulation by iron and copper. Planta **223**: 1178–1190.

Murray, J.D. (2011). Invasion by Invitation : Rhizobial Infection in Legumes. Mol. Plant Microbe Interact. **24**: 631–639.

Navascués, J., Pérez-Rontomé, C., Gay, M., Marcos, M., Yang, F., Walker, F.A., Desbois, A., Abián, J., and Becana, M. (2012). Leghemoglobin green derivatives with nitrated hemes evidence production of highly reactive nitrogen species during aging of legume nodules. Proc. Natl. Acad. Sci. U.S.A. **109**: 2660–2665.

Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-Coffee : A Novel Method for Fast and Accurate Multiple Sequence Alignment. J. Mol. Biol. **302**: 205–217.

Nozoye, T., Nagasaka, S., Kobayashi, T., Takahashi, M., Sato, Y., Sato, Y., Uozumi, N., Nakanishi, H., and Nishizawa, N.K. (2011). Phytosiderophore Efflux Transporters Are Crucial for Iron Acquisition in Graminaceous Plants. J. Biol. Chem. **286**: 5446–5454.

Oldroyd, G.E.D., Murray, J.D., Poole, P.S., and Downie, J.A. (2011). The Rules of Engagement in the Legume-Rhizobial Symbiosis. Annu. Rev. Genet. **45**: 119–144.

Olsen, R.A., Clark, R.B., Bennett, J.H. (1981) The enhancement of soil fertility by plant roots. Am. Scientist. **69**. 378–384.

Orozco-Mosqueda, M.D.C.., Santoyo, G., Farías-Rodríguez, R., Macías-Rodríguez, L., and Valencia-Cantero, E. (2012). Identification and expression analysis of multiple FRO gene copies in *Medicago truncatula*. Genet. Mol. Res. **11**: 4402–4410.

Ott, T., Sullivan, J., James, E.K., Flemetakis, E., Günther, C., Gibon, Y., Ronson, C., and Udvardi, M. (2009). Absence of Symbiotic Leghemoglobins Alters Bacteroid and Plant Cell Differentiation During Development of *Lotus japonicus* Root Nodules. Mol. Plant Microbe Interact. **22**: 800–808.

Pedas, P., Ytting, C.K., Fuglsang, A.T., Jahn, T.P., Schjoerring, J.K., and Husted, S. (2008). Manganese Efficiency in Barley: Identification and Characterization of the Metal Ion Transporter HvIRT1. Plant Physiol. **148**: 455–466.

Pérez-Ruiz, J.M., Spínola, M.C., Kirchsteiger, K., Moreno, J., Sahrawy, M., Cejudo, F.J. (2006). Rice NTRC Is a High-Efficiency Redox System for Chloroplast Protection against Oxidative Damage. Plant Cell **18**: 2356–2368.

Pierre, O., Engler, G., Hopkins, J., Brau, F., Boncompagni, E., and Hérouart, D. (2013). Peribacteroid space acidification: A marker of mature bacteroid functioning in *Medicago truncatula* nodules. Plant, Cell Environ. **36**: 2059–2070.

Preston, R.A., Murphy, R.F., and Jones, E.W. (1989). Assay of vacuolar pH in yeast and identification of acidification-defective mutants. Proc. Natl. Acad. Sci. U.S.A. **86**: 7027–7031.

Qi, Z. and O'Brian, M.R. (2002). Interaction between the bacterial iron response regulator and ferrochelatase mediates genetic control of heme biosynthesis. Mol. Cell **9**: 155–162.

Ratet, P., Pawlowski, K., Schell, J., and de Bruijn, F.J. (1989). The *Azorhizobium caulinodans* nitrogen-fixation regulatory gene, *nifA* , is controlled by the cellular nitrogen and oxygen status. Mol. Microbiol. **3**: 825–838.

Raux-Deery, E., Leech, H.K., Nakrieko, K-A., McLean, K.J., Munro, A.W., Heathcote, P., Rigby, S.E.J., Smith, A.G., and Warren, M.J. (2005). Identification and characterization of the terminal enzyme of siroheme biosynthesis from *Arabidopsis thaliana*: A plastid-located sirohydrochlorin ferrochelatase containing a 2Fe-2S center. J. Biol. Chem. **280**: 4713–4721.

Ravet, K. and Pilon, M. (2013). Copper and iron homeostasis in plants: the challenges of oxidative stress. Antioxid. Redox Signal. **19**: 919–932.

Ravet, K., Touraine, B., Boucherez, J., Briat, J-F., Gaymard, F., and Cellier, F. (2009). Ferritins control interaction between iron homeostasis and oxidative stress in Arabidopsis. Plant J. **57**: 400–412.

Rellán-Álvarez, R., Giner-Martínez-Sierra, J., Orduna, J., Orera, I., Rodríguez-Castrillón, J.Á., García-Alonso, J.I., Abadía, J., and Álvarez-Fernández, A. (2010). Identification of a Tri-Iron(III), Tri-Citrate Complex in the Xylem Sap of Iron-Deficient Tomato Resupplied with Iron: New Insights into Plant Iron Long-Distance Transport. Plant Cell Physiol. **51**: 91–102.

Richter, A.S. and Grimm, B. (2013). Thiol-based redox control of enzymes involved in the tetrapyrrole biosynthesis pathway in plants. Front. Plant Sci. **4**: 1–11.

Richter, A.S., Peter, E., Rothbart, M., Schlicke, H., Toivola, J., Rintamäki, E., and Grimm, B. (2013). Posttranslational influence of NADPH-dependent thioredoxin reductase C on enzymes in tetrapyrrole synthesis. Plant Physiol. **162**: 63–73.

Roberts, A.G., Redding, S.J., and Llewellyn, D.H. (2005). An alternatively-spliced exon in the 5'-UTR of human ALAS1 mRNA inhibits translation and renders it resistant to haem-mediated decay. FEBS Lett. **579**: 1061–1066.

Robinson, N.J., Procter, C.M., Connolly, E.L., and Guerinot, M.L. (1999). A ferric-chelate reductase for iron uptake from soils. Nature **397**: 694–697.

Rodríguez-Celma, J., Lin, W-D., Fu, G-M., Abadía, J., López-Millán, A-F., and Schmidt, W. (2013). Mutually exclusive alterations in secondary metabolism are critical for the uptake of insoluble iron compounds by Arabidopsis and *Medicago truncatula*. Plant Physiol. **162**: 1473–1485.

Rodríguez-Celma, J., Vázquez-Reina, S., Orduna, J., Abadía, A., Abadía, J., Álvarez-Fernández, A., and López-Millán, A.F. (2011). Characterization of flavins in roots of Fe-deficient strategy I plants, with a focus on *Medicago truncatula*. Plant Cell Physiol. **52**: 2173–2189.

Rodríguez-Haas, B., Finney, L., Vogt, S., González-Melendi, P., Imperial, J., and González-Guerrero, M. (2013). Iron distribution through the developmental stages of *Medicago truncatula* nodules. Metallomics **5**: 1247–1253.

Roschzttardtz, H., Séguéla-Arnaud, M., Briat, J-F., Vert, G., and Curie, C. (2011). The FRD3 Citrate Effluxer Promotes Iron Nutrition between Symplastically Disconnected Tissues throughout *Arabidopsis* Development. Plant Cell **23**: 2725–2737.

Roux, B., Rodde, N., Jardinaud, M-F., Timmers, T., Sauviac, L., Cottret, L., Carrère, S.,Sallet, E., Courcelle, E., Moreau, S., Debellé, F., Capela, D., de Carvalho-Niebel, F., Gouzy, J., Bruand, C., and Gamas, P.. (2014). An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. Plant J. **77**: 817–837.

Ruangkiattikul, N., Bhubhanil, S., Chamsing, J., Niamyim, P., Sukchawalit, R., and Mongkolsuk, S. (2012). *Agrobacterium tumefaciens* membrane-bound ferritin plays a role in protection against hydrogen peroxide toxicity and is negatively regulated by the iron response regulator. FEMS Microbiol. Lett. **329**: 87–92.

Saitou, N. and Nei, M. (1987). The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees. Mol. Biol. Evol. **4**: 406–425.

Sankari, S. and O'Brian, M.R. (2014). A bacterial iron exporter for maintenance of iron homeostasis. J. Biol. Chem. **289**: 16498–16507.

Santana, M.A., Tan, F-C., and Smith, A.G. (2002). Molecular characterisation of coproporphyrinogen oxidase from *Glycine max* and *Arabidopsis thaliana*. Plant Physiol. Biochem. **40**: 289–298.

Santana, M.A., Pihakaski-Maunsbach, K., Sandal, N., Marcker, K.A., and Smith, A.G. (1998). Evidence that the Plant Host Synthesizes the Heme Moiety of Leghemoglobin in Root Nodules. Plant Physiol. **116**: 1259–1269.

Santi, S. and Schmidt, W. (2009). Dissecting iron deficiency-induced proton extrusion in Arabidopsis roots. New Phytol. **183**: 1072–1084.

Schaedler, T.A., Thornton, J.D., Kruse, I., Schwarzländer, M., Meyer, A.J., van Veen, H.W., and Balk, J. (2014). A conserved mitochondrial ATP-binding cassette transporter exports glutathione polysulfide for cytosolic metal cofactor assembly. J. Biol. Chem. **289**: 23264–23274.

Scharfenberg, M., Mittermayr, L., von Roepenack-Lahaye, E., Schlicke, H., Grimm, B., Leister, D., and Kleine, T. (2014). Functional characterization of the two ferrochelatases in *Arabidopsis thaliana*. Plant, Cell Environ. **38**: 280–298.

Schmid, N.B., Giehl, R.F.H., Döll, S., Mock, H-P., Strehmel, N., Scheel, D., Kong, X., Hider, R.C., and von Wirén, N. (2014). Feruloyl-CoA 6'-Hydroxylase1-Dependent Coumarins Mediate Iron Acquisition from Alkaline Substrates in Arabidopsis. Plant Physiol. **164**: 160–172.

Schuler, M., Rellán-Álvarez, R., Fink-Straube, C., Abadía, J., and Bauer, P. (2012). Nicotianamine Functions in the Phloem-Based Transport of Iron to Sink Organs , in Pollen Development and Pollen Tube Growth in *Arabidopsis*. Plant Cell **24**: 2380–2400.

Seefeldt, L.C., Hoffman, B.M., and Dean, D.R. (2009). Mechanism of Mo-Dependent Nitrogenase. Annu. Rev. Biochem. **78**: 701–722.

Serrato, A.J., Pérez-Ruiz, J.M., Spínola, M.C., and Cejudo, F.J. (2004). A novel NADPH thioredoxin reductase, localised in the chloroplast, which deficiency causes hypersensitivity to abiotic stress in *Arabidopsis thaliana*. J. Biol. Chem. **279**: 43821–43827.

Sharma, S.S., Dietz, K-J., and Mimura, T. (2016). Vacuolar compartmentalization as indispensable component of heavy metal detoxification in plants. Plant Cell Environ. **39**: 1112–1126.

Singh, S., Katzer, K., Lambert, J., Cerri, M., and Parniske, M. (2014). CYCLOPS , A DNA-Binding Transcriptional Activator , Orchestrates Symbiotic Root Nodule Development. Cell: **15** 139–152.

Singleton, C., White, G.F., Todd, J.D., Marritt, S.J., Cheesman, M.R., Johnston, A.W.B., and Le Brun, N.E. (2010). Heme-responsive DNA binding by the global iron regulator Irr from *Rhizobium leguminosarum*. J. Biol. Chem. **285**: 16023–16031.

Slavic, K., Krishna, S., Lahree, A., Bouyer, G., Hanson, K.K., Vera, I., Pittman, J.K., Staines, H.M., and Mota, M.M. (2016). A vacuolar iron-transporter homologue acts as a detoxifier in *Plasmodium*. Nat. Commun. **7**: 1–10.

Spatzal, T., Aksoyoglu, M., Zhang, L., Andrade, S.L.A., Schleicher, E., Weber, S., Rees, D.C., and Einsle, O. (2011). Evidence for Interstitial Carbon in Nitrogenase FeMo Cofactor. Science **334**: 940.

Stephens, B.W., Cook, D.R., and Grusak, M.A. (2011). Characterization of zinc transport by divalent metal transporters of the ZIP family from the model legume *Medicago truncatula*. BioMetals **24**: 51–58.

Stephenson, P.G., Fankhauser, C., and Terry, M.J. (2009). PIF3 is a repressor of chloroplast development. Proc. Natl. Acad. Sci. U.S.A. **106**: 7654–7659.

Strozycki, P.M., Szczurek, A., Lotocka, B., Figlerowicz, M., and Legocki, A.B. (2007). Ferritins and nodulation in *Lupinus luteus*: Iron management in indeterminate type nodules. J. Exp. Bot. **58**: 3145–3153.

Suganuma, N., Nakamura, Y., Yamamoto, M., Ohta, T., Koiwa, H., Akao, S., and Kawaguchi, M. (2003). The *Lotus japonicus Sen1* gene controls rhizobial differentiation into nitrogen-fixing bacteroids in nodules. Mol. Gen. Genomics **269**: 312–320.

Tadege, M., Wen, J., He, J., Tu, H., Kwak, Y., Eschstruth, A., Cayrel, A., Endre, G., Zhao, P.X., Chabaud, M., Ratet, P., and Mysore, K.S. (2008). Large-scale insertional mutagenesis using the *Tnt1* retrotransposon in the model legume *Medicago truncatula*. Plant J. **54**: 335–347.

Takanashi, K., Yokosho, K., Saeki, K., Sugiyama, A., Sato, S., Tabata, S., Ma, J.F., and Yazaki, K. (2013). LjMATE1: A citrate transporter responsible for iron supply to the nodule infection zone of *Lotus japonicus*. Plant Cell Physiol. **54**: 585–594.

Tanaka, R. and Tanaka, A. (2007). Tetrapyrrole Biosynthesis in Higher Plants. Annu. Rev. Plant Biol **58**: 321–346.

Tang, C., Robson, A.D., and Dilworth, M.J. (1990). The role of iron in nodulation and nitrogen fixation in *Lupinus angustifolius* L. New Phytol. **114**: 173–182.

Tejada-Jiménez, M., Castro-Rodríguez, R., Kryvoruchko, I., Lucas, M.M., Udvardi, M., Imperial, J., and González-Guerrero, M. (2015). *Medicago truncatula* Natural Resistance-Associated Macrophage Protein1 Is Required for Iron Uptake by Rhizobia-Infected Nodule Cells. Plant Physiol. **168**: 258–272.

Tejada-Jiménez, M., Gil-Díez, P., León-Mediavilla, J., Wen, J., Mysore, K.S., Imperial, J., and González-Guerrero, M. (2017). *Medicago truncatula* Molybdate Transporter type 1 (MtMOT1.3) is a plasma membrane molybdenum transporter required for nitrogenase activity in root nodules under molybdenum deficiency. New Phytol. DOI: 10.1111/nph.14739

Timofeev, R., Ghalamkari, Z., and Buckhout, T.J. (2016). Expression of the Vacuolar Iron Transporter-Like genes and their localization in Saccharomyces and Arabidopsis. ISINIP conference abstract.

Todd, J.D., Sawers, G., and Johnston, A.W.B. (2005). Proteomic analysis reveals the wide-ranging effects of the novel, iron-responsive regulator RirA in *Rhizobium leguminosarum* bv. *viciae*. Mol. Gen. Genomics **273**: 197–206.

Todd, J.D., Sawers, G., Rodionov, D.A., and Johnston, A.W.B. (2006). The *Rhizobium leguminosarum* regulator IrrA affects the transcription of a wide range of genes in response to Fe availability. Mol. Genet. Genomics **275**: 564–577.

Touraine, B., Boutin, J-P., Marion-Poll, A., Briat, J-F., Peltier, G., and Lobréaux, S. (2004). Nfu2: A scaffold protein required for [4Fe-4S] and ferredoxin iron-sulphur cluster assembly in *Arabidopsis* chloroplasts. Plant J. **40**: 101–111.

Trepp, G.B., Van de Mortel, M., Yoshioka, H., Miller, S.S., Samac, D.A., Gantt, J.S., and Vance, C.P. (1999). NADH-Glutamate Synthase in Alfalfa Root Nodules . Genetic Regulation and Cellular Expression. Plant Physiol. **119**: 817–828.

Tsang, E.W.T., Hu, Z., Chang, Q., McGregor, D.I., and Keller, W.A. (2003). Expression of a *Brassic napus* glutamate 1-semialdehyde aminotransferase in *Escherichia coli* and characterization of the recombinant protein. Protein Expr Purif. **29**: 193–201.

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

Ujwal, M.L., McCormac, A.C., Goulding, A., Kumar, A.M., Söll, D., and Terry, M.J. (2002). Divergent regulation of the *HEMA* gene family encoding glutamyl-tRNA reductase in *Arabidopsis thaliana*: expression of *HEMA2* is regulated by sugars, but is independent of light and plastid signalling. Plant Mol. Biol. **50**: 83–91.

van Brussel, A.A.N., Bakhuizen, R., van Spronsen, P.C., Spaink, H.P., Tak, T., Lugtenberg, B.J.J., and Kijne, J.W. (1992). Induction of Pre-Infection Thread Structures in the Leguminous Host Plant by Mitogenic Lipo-Oligosaccharides of *Rhizobium*. Science **257**: 70–72.

van Lis, R., Atteia, A., Nogaj, L.A., and Beale, S.I. (2005). Subcellular localization and light-regulated expression of protoporphyrinogen IX oxidase and ferrochelatase in *Chlamydomonas reinhardtii*. Plant Physiol. **139**: 1946–1958.

Vert, G., Briat, J-F., and Curie, C. (2001). Arabidopsis *IRT2* gene encodes a root-periphery iron transporter. Plant J. **26**: 181–189.

Vert, G., Grotz, N., Dédaldéchamp, F., Gaymard, F., Guerinot, M.L, Briat, J-F. and Curie, C. (2002). IRT1, an Arabidopsis Transporter Essential for Iron Uptake from the Soil and for Plant Growth. Plant Cell **14**: 1223-1233.

Vothknecht, U.C., Kannangara, C.G., and von Wettstein, D. (1997). Barley Glutamyl tRNA^{Glu} reductase: mutations affecting haem inhibition and enzyme activity. Phytochemistry **47**: 513–519.

Wallace, I.M., O'Sullivan, O., Higgins, D.G., and Notredame, C. (2006). M-Coffee : combining multiple sequence alignment methods with T-Coffee. **34**: 1692–1699.

Wang, J., Hou, Q., Li, P., Yang, L., Sun, X., Benedito, V.A., Wen, J., Chen, B., Mysore, K.S., and Zhao, J. (2017). Diverse functions of multidrug and toxin extrusion (MATE) transporters in citric acid efflux and metal homeostasis in *Medicago truncatula*. Plant J. **90**: 79–95.

Wienkoop, S. and Saalbach, G. (2003). Proteome analysis. Novel proteins identified at the peribacteroid membrane from *Lotus japonicus* root nodules. Plant Physiol. **131**: 1080–1090.

Wiig, J.A., Hu, Y., and Ribbe, M.W. (2015). Refining the pathway of carbide insertion into the nitrogenase M-cluster. Nat. Commun. **6**: 8034.

Williams, B.P., Burgess, S.J., Reyna-Llorens, I., Knerova, J., Aubry, S., Stanley, S., and Hibberd, J.M. (2016). An Untranslated *cis*-Element Regulates the Accumulation of Multiple C₄ Enzymes in *Gynandropsis gynandra*. **28**: 454–465.

Wingert, R.A., Galloway, J.L., Barut, B., Foott, H., Fraenkel, P., Axe, J.L., Weber, G.J., Dooley, K., Davidson, A.J., Schmidt, B., Paw, B.H., Shaw, G.C., Kingsley, P., Palis, J., Schubert, H., Chen, O., Kaplan, J., the Tübingen 2000 Screen Consortium, and Zon, L.I. (2005). Deficiency of *glutaredoxin 5* reveals Fe – S clusters are required for vertebrate haem synthesis. Nature **436**: 1035–1039.

Wollers, S., Layer, G., Garcia-Serres, R., Signor, L., Clemancey, M., Latour, J.M., Fontecave, M., and de Choudens, S.O. (2010). Iron-sulfur (Fe-S) cluster assembly: The SufBCD complex is a new type of Fe-S scaffold with a flavin redox cofactor. J. Biol. Chem. **285**: 23331–23341.

Wu, H., Li, L., Du, J., Yuan, Y., Cheng, X., and Ling, H-Q. (2005). Molecular and biochemical characterization of the Fe(III) chelate reductase gene family in *Arabidopsis thaliana*. Plant Cell Physiol. **46**: 1505–1514.

Xie, F., Murray, J.D., Kim, J., Heckmann, A.B., Edwards, A., Oldroyd, G.E.D., and Downie, J.A. (2012). Legume pectate lyase required for root infection by rhizobia. Proc. Nat. Acad. Sci. U.S.A. **109**: 633–638.

Yan, J.Y., Li, C.X., Sun, L., Ren, J.Y., Li, G.X., Ding, Z.J., and Zheng, S.J. (2016). A WRKY transcription factor regulates Fe translocation under Fe deficiency. Plant Physiol. **171**: 2017-2027.

Yoshida, K. and Negishi, T. (2013) The identification of a vacuolar iron transporter involved in the blue coloration of cornflower petals. Phytochemistry **94**: 60–67.

Zhang, X., Virtanen, A., and Kleiman, F.E. (2010). To polyadenylate or to deadenylate That is the question.Cell Cycle **9**: 4437–4449.

Zhuang, Y., Zhang, H., and Lin, S. (2013). Polyadenylation of 18S rRNA in algae. J. Phycol. **49**: 570–579.

Zuckerkandl, E. and Pauling, L. (1965). Evolutionary divergence and convergence in proteins. Evol. Genes Proteins: 97–166.

Appendix 1

Components of Level 2 and Level 1 Golden Gate plasmids used

Table A1.1 Components of Level 2 plasmids used for hairy root transformation of Medicago

Plasmid code	Full name of construct	Backbone	Position 1	Position 2	Position 3	Position 4
EC64032	pL2V-KAN-pMtFeCh1.2::GUS-dsRed-64032	EC50506	pL1-R1-pMtFeCh1.2::GUS-TOCS-64026	pL1M-R2-pAtUBI10::dsRed-tHSP-10938		EC41744
EC64082	pL2V-KAN-plastid_mcherry-MtFeCh1.2_eGFP-NLS_cyPET-64082	EC50506	pL1M-R1-pLjUBI::Plastid_mcherry-TAcs-64076	pL1M-R2-pLjUBI::MtFeCh1.2_eGFP-TOCS-64016	pL1M-R3-pNOS::NLS_cyPET-TNOS-64017	EC41766
EC64083	pL2V-KAN-Mit_mcherry-eGFP-NLS_cyPET-64083	EC50506	pL1M-R1-pljUBI::Mit_mcherry-TACS-64074	pL1-R2-pLjUBI::eGFP-TOCS-64035	pL1M-R3-pNOS::NLS_cyPET-TNOS-64017	EC41766
EC64109	pL2V-KAN-SEN1_Myc-dsRed-VTL_HA-64109	EC50506	pL1M-R1-pSen1::SEN1_FL3_Myc-TOcs-64097	pL1M-R2-pAtUBI10::dsRed-tHSP-10938	pL1M-R3-pVtl::VTL_FL3_HA-TAcs-64101	EC41766
EC64110	pL2V-KAN-SEN1_Myc-dsRed-64110	EC50506	pL1M-R1-pSen1::SEN1_FL3_Myc-TOcs-64097	pL1M-R2-pAtUBI10::dsRed-tHSP-10938		EC41744
EC64111	pL2V-KAN-VTL_HA-dsRed-64111	EC50506	pL1M-R1-pVtl::VTL_FL3_HA-TAcs-64099	pL1M-R2-pAtUBI10::dsRed-tHSP-10938		EC41744
EC64112	pL2V-KAN-SEN1_mcherry-SST1_eGFP-64112	EC50506	pL1M-R1-pSen1::SEN1_FL3_mcherry-TOcs-64103	pL1M-R2-pSST1-SST1_eGFP-Tnos-64107		EC41744
EC64119	pL2V-KAN-SEN1_mcherry-PM_eGFP-64119	EC50506	pL1M-R1-pSen1::SEN1_FL3_mcherry-TOcs-64103	pL1-R2-pLjUBI::PM-eGFP-Tnos-64118		EC41744
EC64120	pL2V-KAN-VTL_mcherry-PM_eGFP-64120	EC50506	pL1M-R1-pVtl::VTL_FL3_mcherry-TAcs-64105	pL1-R2-pLjUBI::PM-eGFP-TNos-64118		EC41744

Table A1.2 Level 1 components used to make Level 2 plasmids

Plasmid Code	ENSA Standard name	Backbone		Promoter		CDS1		CDS2		Terminator	
EC10938	pL1M-R2-pAtUBI10::dsRed-THSP-10938	pL1M-R2	EC47811	pAtUBI10	EC15062	dsRed				THSP	
EC64016	pL1M-R2-pLjUBI::MtFECH1B_eGFP-TOCS-64016	pL1-R2	EC47811	pLjUBI	EC15251	MtFECH1B	EC64011	eGFP	EC15095	TOCS	EC41432
EC64017	pL1M-R3-pNOS::NLS_cyPET-TNOS-64017	pL1-R3	EC47822	pNOS	EC15057	NLS	EC15101	cyPET	EC15096	TNOS	EC41421
EC64026	pL1-R1-pMtFECH1B::GUS-TOCS-64026	pL1-R1	EC47802	pMtFECH1B	EC64001	GUS	EC75111			TOCS	EC41432
EC64035	pL1-R2-pLjUBI::eGFP-TOCS-64035	pL1-R2	EC47811	pLjUBI	EC15251	eGFP	EC15112			TOCS	EC41432
EC64074	pL1M-R1-pLjUBI::Mit mcherry-TACS-64074	pL1-R1	EC47802	pLjUBI	EC15251	Mito marker	EC10037	mcherry	EC15097	TACS	EC15321
EC64076	pL1M-R1-pLjUBI::Plastid_mcherry-TACS-64076	pL1-R1	EC47802	pLjUBI	EC15251	Plastid marker	EC10036	mcherry	EC15097	TACS	EC15321
EC64097	pL1M-R1-pSEN1::SEN1_FL3_Myc-TOCS-64097	pL1-R1	EC47802	pSEN1	EC64095	SEN1_FL3	EC64089	3xMyc	EC15212	TOCS	EC41432
EC64099	pL1M-R1-pVTL::VTL_FL3_HA-TACS-64099	pL1-R1	EC47802	pVTL	EC64096	MtVTL_FL3	EC64090	3xHA	EC15185	TACS	EC15321

Plasmid code	ENSA Standard name	Backbone	Promoter	CDS1	CDS2	Terminator
EC64101	pL1M-R3-pVTL::VTL_FL3_HA-TACS-64101	pL1-R3 EC47822	pVTL EC64096	MtVTL_FL3 EC64090	3xHA EC15185	TACS EC15321
EC64103	pL1M-R1-pSEN1::SEN1_FL3_mcherry-TOCS-64103	pL1-R1 EC47802	pSEN1 EC64095	SEN1_FL3 EC64089	mcherry EC15110	TOCS EC41432
EC64105	pL1M-R1-pVTL::VTL_FL3_mcherry-TACS-64105	pL1-R1 EC47802	pVTL EC64096	MtVTL_FL3 EC64090	mcherry EC15110	TACS EC15321
EC64107	pL1M-R2-pSST1-SST1_eGFP-TNOS-64107	pL1-R2 EC47811	pSST1 EC64094	SST1 EC64093	eGFP EC15095	TNOS EC41421
EC64118	pL1-R2-pLjUBI::PM-eGFP-TNOS-64118	pL1-R2 EC47811	pLjUBI EC15251	PM marker EC10039	eGFP EC15113	TNOS EC41421

Domestication of sequences for Golden Gate Cloning

All Level 0 plasmids with a plasmid code beginning "EC64" were domesticated and synthesised in this study. All other initial codes (*e.g.* EC15, EC47) were supplied by the Engineering Nitrogen Symbiosis for Africa (ENSA) consortium (Eleni Soumpourou and Andy Breakspear).

EC64001

Native sequence

```
GGATGGATGAGATAAGAGAAGCAATGTTGCTTGCTCCGATGTGCAGTTTTTGTTTGTGGCGGTC
AACGTAACATGTTCCTTCAATGTCACTTCATGGCTTGTTTAAATTTGTAATCATTGTTTCCATG
ACTACAAAATTTATACATAATTAAATGCTAATCTGGATTGACATAGTCTTTTTAAAGCTGTTGG
CTGACGAGTGCAATGAGAAAATCACCATCAACTTTATGTGTTCCTTCCATGTACATTTATAAAT
TTCCTAGAATTTATTGAGTGTTGTGAAGAAGAATATGAATCTTAACTTAAATTATATAGGCTAT
TTTTTTTTACTAAAATCGGTCTATAAGTTCTTTTTCGATTTCACTCTGGTTTCTTTAATTCTTT
AGTATTTTTCGTTCTATTTTAGTGAGAAACATGTAATTAATTCATAATTGTTTTACAAGTTAGC
TCATCTCACGTCAGGATTGGATAGAAAACAAAGTTGTATAATATTTTTTTTAACGATGAATAAA
TAGTTTAGATGAGTATGAATCAGATCTTTGACATCACGGTTAGAATAGACGTGAAGACTCAATA
AACAATTACATAATTACTTACGAATAATTTACCAAAAGTTATTTACATCATTTCAATTGCAAGG
ACTAAAATAATTTAATAAGTAGATATTAAAAAAAGTGACATAATGAGAAGCAAGAATGGTACCA
AGTGTGCTGGGAAAGACAACGCCATGAAATATTGTGGGAGGTTAAAAAGTCTTAATTGTTTGTT
TGAGTAGAATAACTCAGCTTTTTTCAACAAAAAAAAAAAAATGAAATAAAAAAGGAAAGTGGGG
CTAGTTTCGTTTTCAATCCCAAAAATATATAGTGAGAAATTTTCTACGAGAATTGAGAAGTGGT
TAACCAACAATAATACAGTCCAGTGGGACAGAGATTCGTGATTTGGTATTGACTTTTATTTTGG
TTTAGCTCACCATCTTTGACGTACACGCAAACTATTTTCCTTTTAATAATACTTTCTTTCACCT
AATCTTCCTTTCTTCCTCACGCACGCCGACCAACTCTTTCCGTCGTCGTCCATCTCAGTT
```

Domesticated sequence

```
cactctgtggtctcaggagGGATGGATGAGATAAGAGAAGCAATGTTGCTTGCTCCGATGTGCA
GTTTTTGTTTGTGGCGGTCAACGTAACATGTTCCTTCAATGTCACTTCATGGCTTGTTTAAATT
TGTAATCATTGTTTCCATGACTACAAAATTTATACATAATTAAATGCTAATCTGGATTGACATA
GTCTTTTTAAAGCTGTTGGCTGACGAGTGCAATGAGAAAATCACCATCAACTTTATGTGTTCCT
TCCATGTACATTTATAAATTTCCTAGAATTTATTGAGTGTTGTGAAGAAGAATATGAATCTTAA
CTTAAATTATATAGGCTATTTTTTTTTACTAAAATCGGTCTATAAGTTCTTTTTCGATTTCACT
CTGGTTTCTTTAATTCTTTAGTATTTTTCGTTCTATTTTAGTGAGAAACATGTAATTAATTCAT
AATTGTTTTACAAGTTAGCTCATCTCACGTCAGGATTGGATAGAAAACAAAGTTGTATAATATT
TTTTTAACGATGAATAAATAGTTTAGATGAGTATGAATCAGATCTTTGACATCACGGTTAGAA
TAGACGTGATGACTCAATAAACAATTACATAATTACTTACGAATAATTTACCAAAAGTTATTTA
CATCATTTCAATTGCAAGGACTAAAATAATTTAATAAGTAGATATTAAAAAAAGTGACATAATG
AGAAGCAAGAATGGTACCAAGTGTGCTGGGAAAGACAACGCCATGAAATATTGTGGGAGGTTAA
AAAGTCTTAATTGTTTGTTTGAGTAGAATAACTCAGCTTTTTTCAACAAAAAAAAAAAAATGAA
ATAAAAAAGGAAAGTGGGGCTAGTTTCGTTTTCAATCCCAAAAATATATAGTGAGAAATTTTCT
ACGAGAATTGAGAAGTGGTTAACCAACAATAATACAGTCCAGTGGGACAGAGATTCGTGATTTG
GTATTGACTTTTATTTTGGTTTAGCTCACCATCTTTGACGTACACGCAAACTATTTTCCTTTTA
ATAATACTTTCTTTCACCTAATCTTCCTTTCTTCCTCACGCACGCCGACCAACTCTTTCCGTCG
TCGTCCATCTCAGTTaatgtgagaccacgaagtg
```

EC64011

Native sequence

```
AACGCACTTTCACATTCTTCTCTTCTTCCCAACCGATACCCTCAATCCCTTCATCGCCGCCGAT
TTTCACTGTCATGTTCTGATATTCAAAATCTTACCCGTGTTACTTGCTACTCGGATTGTAACAA
GTCTACATCTCAGGCATCCTTGTTTTTGTGTCCTGGCTCGACCAATACAAGAATTGGCAGCAGA
AACTTAGTTTCTCGCTCATTTTATTCTGCAGATGCAAGCACTTATAATGGGCTTACTGTACAAT
CTCCTACTCACGCTGCCCAAGAAAAGTTGGTGTGTTGCTTCTCAATCTAGGTGGACCCGAGAC
ACTGGACGATGTTCAACCTTTCCTTTTTAATCTTTTCGCAGATCCTGATATCATTCGGCTCCCA
AGATTGTTTCGGTTTCTCCAGCAACCATTAGCAAAACTGATATCTACGCTCCGGGCTCCTAAAT
CCAAGGAAGCATATGCTTCTATTGGAGGTGGCTCTCCTTTACGCAAAATTACAGATGACCAGGC
ACTTGCACTTAAAAGGGCTTTGGAAGCAAAGGGCCTCTCTTCAAATATCTATGTTGGGATGCGG
TACTGGTACCCATTCACTGAAGAAGCAATTCAACAAATTAAGAAGGATGGAATAACAAGGCTCG
TGGTGTTACCCCTTTATCCCCAGTTTTCTATATCCACTACTGGGTCAAGCATCAGTGTTCTGGA
GCAAACGTTCAGGGAAGATGCTTATTTGTCTAGGCTTCCTGTTTCCATTATAAACTCTTGGTAT
CAAAGGGAAGGTTATATCAAGTCAATGGCTGACTTAATTGAGAAAGAACTCGAGAGCTTTTCTG
AGCCAAAGGAAGCAATGATATTTTTTAGTGCCCATGGTGTTCCTGTCAGTTATGTTGAGAATGC
TGGGGATCCATACCGAGATCAAATGGAGGAGTGCATCTTCTTGATCATGCAGGAATTGAAGGCT
AGAGGAATTAGTAATGAGCACACTCTTGCTTATCAGAGTCGAGTGGGCCCTGTACAGTGGTTGA
AACCATATACAGATGAAGTTCTAGTTGAGCTTGGCCAGAAAGGTGTGAAGAGTCTTTTAGCTGT
TCCAGTGAGCTTTGTGAGTGAGCACATAGAGACTCTTGAAGAAATTGACATGGAGTACAGGGAA
TTGGCTCTTGAATCTGGCATCAAGAATTGGGCACGTGTCCCGGCTCTTGGTCTCACCCCTTCCT
TCATCATGGATCTTGCAGATGCGGTGATAGAAGCTCTACCATCGGCAGCAGCAATAAATGCCCC
TACCAGCACCTCTGAAGATATGGATAAAGACCCAGTTAAATACTTCGCCAAGATGTTTTTTGGT
TCAATTTTGGCTTTCCTCTTGTTTTTCTCGCCCAAAATGATTACTGCGTTCAGGAATCATGTCA
TT
```

Domesticated sequence

```
cactctgtggtctcaaatgAACGCACTTTCACATTCTTCTCTTCTTCCCAACCGATACCCTCAA
TCCCTTCATCGCCGCCGATTTTCACTGTCATGTTCTGATATTCAAAATCTTACCCGTGTTACTT
GCTACTCGGATTGTAACAAGTCTACATCTCAGGCATCCTTGTTTTTGTGTCCTGGCTCGACCAA
TACAAGAATTGGCAGCAGAAACTTAGTTTCTCGCTCATTTTATTCTGCAGATGCAAGCACTTAT
AATGGGCTTACTGTACAATCTCCTACTCACGCTGCCCAAGAAAAGTTGGTGTGTTGCTTCTCA
ATCTAGGTGGACCCGAGACACTGGACGATGTTCAACCTTTCCTTTTTAATCTTTTCGCAGATCC
TGATATCATTCGGCTCCCAAGATTGTTTCGGTTTCTCCAGCAACCATTAGCAAAACTGATATCT
ACGCTCCGGGCTCCTAAATCCAAGGAAGCATATGCTTCTATTGGAGGTGGCTCTCCTTTACGCA
AAATTACAGATGACCAGGCACTTGCACTTAAAAGGGCTTTGGAAGCAAAGGGCCTCTCTTCAAA
TATCTATGTTGGGATGCGGTACTGGTACCCATTCACTGAAGAAGCAATTCAACAAATTAAGAAG
GATGGAATAACAAGGCTCGTGGTGTTACCCCTTTATCCCCAGTTTTCTATATCCACTACTGGGT
CAAGCATCAGTGTTCTGGAGCAAACGTTCAGGGAAGATGCTTATTTGTCTAGGCTTCCTGTTTC
CATTATAAACTCTTGGTATCAAAGGGAAGGTTATATCAAGTCAATGGCTGACTTAATTGAGAAA
GAACTCGAGAGCTTTTCTGAGCCAAAGGAAGCAATGATATTTTTTAGTGCCCATGGTGTTCCTG
TCAGTTATGTTGAGAATGCTGGGGATCCATACCGAGATCAAATGGAGGAGTGCATCTTCTTGAT
CATGCAGGAATTGAAGGCTAGAGGAATTAGTAATGAGCACACTCTTGCTTATCAGAGTCGAGTG
GGCCCTGTACAGTGGTTGAAACCATATACAGATGAAGTTCTAGTTGAGCTTGGCCAGAAAGGTG
TGAAGAGTCTTTTAGCTGTTCCAGTGAGCTTTGTGAGTGAGCACATAGAGACTCTTGAAGAAAT
TGACATGGAGTACAGGGAATTGGCTCTTGAATCTGGCATCAAGAATTGGGCACGTGTCCCGGCT
CTTGGACTCACCCCTTCCTTCATCATGGATCTTGCAGATGCGGTGATAGAAGCTCTACCATCGG
CAGCAGCAATAAATGCCCCTACCAGCACCTCTGAAGATATGGATAAAGACCCAGTTAAATACTT
CGCCAAGATGTTTTTTGGTTCAATTTTGGCTTTCCTCTTGTTTTTCTCGCCCAAAATGATTACT
GCGTTCAGGAATCATGTCATTggtgtgagaccacgaagtg
```

EC64089

Native sequence

```
ATGGTTGGTACAATATGTGATTCTGAGTCACTAAACAAAGTAATCACAATTCCAAATCCTACAA
AGGAAATGGAAGAAAAGCAAATAAAAATCAAAGAAAATTGTAAAACTTCAGATATTGATTATTG
GCAAAGGGCTCAATGGCTTCGCGCAGCCGTATTAGGAGCTAATGATGGATTAGTTTCTGTTGCT
TCACTAATGATGGGTGTAGGAGCTGTTAAAACAGACAGTGCAACAATGTTAGTCGCGGGTTTTG
CGGGATTAATTGCAGGAGCATGTGGTATGGCAATAGGAGAATTTGTTTCTGTGTACACACAATA
TGAAGTTGAAATTGGTCAAATGATGAGAGATTTAGGAACAAGTGATAGAAAAGAGAAAGAATTG
GAGATTGAGTTGGAGAAAAGAAGATCATTGCCTAATCCATTGCAAGCTGCTGCAGCTTCTGCAT
TCTCATTTTCTATTGGTGGTTTGGTTCCTTTACTTAGTGGTTCATTTATAAGAGTTTATAAGAT
TAGGATTATTGCGATTATGGCAATAGCTAGTTTGGCTTTGGTTGTGTTTGGAGGTGTTGGTGCT
ATGCTTGGTAAAACACCAAAGGTAAAATCTTCTATTAGGTTCCTTCTTGGTGGATGGATGGCTA
TGGCTATCACTTTTGGGTTAACCAAATTACTTGCACATTGTAGTGGTTTGGATTTGGAAATT
```

Domesticated sequence (with FL3 linker)

```
cactctgtggtctcaaatgGCCGTTGGTACAATATGTGATTCTGAGTCACTAAACAAAGTAATC
ACAATTCCAAATCCTACAAAGGAAATGGAAGAAAAGCAAATAAAAATCAAAGAAAATTGTAAAA
CTTCAGATATTGATTATTGGCAAAGGGCTCAATGGCTTCGCGCAGCCGTATTAGGAGCTAATGA
TGGATTAGTTTCTGTTGCTTCACTAATGATGGGTGTAGGAGCTGTTAAAACAGACAGTGCAACA
ATGTTAGTCGCGGGTTTTGCGGGATTAATTGCAGGAGCATGTGGTATGGCAATAGGAGAATTTG
TTTCTGTGTACACACAATATGAAGTTGAAATTGGTCAAATGATGAGAGATTTAGGAACAAGTGA
TAGAAAAGAGAAAGAATTGGAGATTGAGTTGGAGAAAAGAAGATCATTGCCTAATCCATTGCAA
GCTGCTGCAGCTTCTGCATTCTCATTTTCTATTGGTGGTTTGGTTCCTTTACTTAGTGGTTCAT
TTATAAGAGTTTATAAGATTAGGATTATTGCGATTATGGCAATAGCTAGTTTGGCTTTGGTTGT
GTTTGGAGGTGTTGGTGCTATGCTTGGTAAAACACCAAAGGTAAAATCTTCTATTAGGTTCCTT
CTTGGTGGATGGATGGCTATGGCTATCACTTTTGGGTTAACCAAATTACTTGCACATTGTAGTG
GTTTGGATTTGGAAATTCTTGGTGGAGGTGGATCTGGTGGAGGTGGATCAGGTGGAGGTGGATC
TGCTGCAGCTggtgtgagaccacgaagtg
```

EC64090

Native sequence

ATGAAGAGACTGTGTTTTCTACATGCAACCCTTTCCTCATTTTCCTTTAATCTTGGTATCTTAA
GATTTTTGTACTTCATGGCTTCTCTTGGTAACCATAATAACAATGAAGTTAATGACATAGAAGC
AAAGCAAAACCAAACTCAAGACACTGAAGAAACAACATTGACTATTCCAAAAGAGCACAATGG
CTAAGAGCTGCTTTGTTAGGAGCCAATGATGGTTTGGTATCAATAACATCATTGATATTAGGTG
TTGGAGCTGTGCATGAGGACATCAAAACAATGCTTCTAGCTGGATTTGCAGGACTAATTGCAGG
AGCATGTAGTATGGGAATTGGAGAGTTTGTATCAGTGTACACTCAGTTTGACATAATGGTAGCT
CAAATGAAAAGAGAGAACAAAATCAACAGCACTTTTGTTGAAGAGGAAAAGCAGCTACTACCAA
ATCCTTTTCAAGCTGCTATTGCATCAGCAATTGCATTTTCATTTGGTGCTACTGTTCCATTGTT
AGGAGCAGCATTGGTAAGGGACTATAAGATAAGGTTGTTTGTTGTTGTTGGTATGGCTAGTTTT
GCATTGTTGGTGTTTGGTGGTGTAGGAGCAATACTTGGTAAAACATCAGTGAAAATGTCTTGTG
TTAGAGTGGTGGTTGGAGGTTGGATGGCTATGGCTATTACCTTTGGCTTAACCAAATTCGTTGG
CTATAGTTCA

Domesticated sequence (with FL3 linker)

cactctgtggtctcaaatgAAGAGACTGTGTTTTCTACATGCAACCCTTTCCTCATTTTCCTTT
AATCTTGGTATCTTAAGATTTTTGTACTTCATGGCTTCTCTTGGTAACCATAATAACAATGAAG
TTAATGACATAGAAGCAAAGCAAAACCAAACTCAAGACACTGAAGAAACAACATTGACTATTC
CAAAAGAGCACAATGGCTAAGAGCTGCTTTGTTAGGAGCCAATGATGGTTTGGTATCAATAACA
TCATTGATATTAGGTGTTGGAGCTGTGCATGAGGACATCAAAACAATGCTTCTAGCTGGATTTG
CAGGACTAATTGCAGGAGCATGTAGTATGGGAATTGGAGAGTTTGTATCAGTGTACACTCAGTT
TGACATAATGGTAGCTCAAATGAAAAGAGAGAACAAAATCAACAGCACTTTTGTTGAAGAGGAA
AAGCAGCTACTACCAAATCCTTTTCAAGCTGCTATTGCATCAGCAATTGCATTTTCATTTGGTG
CTACTGTTCCATTGTTAGGAGCAGCATTGGTAAGGGACTATAAGATAAGGTTGTTTGTTGTTGT
TGGTATGGCTAGTTTTGCATTGTTGGTGTTTGGTGGTGTAGGAGCAATACTTGGTAAAACATCA
GTGAAAATGTCTTGTGTTAGAGTGGTGGTTGGAGGTTGGATGGCTATGGCTATTACCTTTGGCT
TAACCAAATTCGTTGGCTATAGTTCA<u>CTTCTTGGTGGAGGTGGATCTGGTGGAGGTGGATCAGG
TGGAGGTGGATCTGCTGCAGCT</u>ggtgtgagaccacgaagtg

EC64093

Native sequence

```
ATGGGTACGATAGGCAACAATTCCCATGAGGGTGATCATCATGGGGTGAACTTCACAGCCCAGA
GAGGGTTTTACACAAAACTCAAATCCGGGTTGAAGGAGACGTTCTTCCCTGATGATCCGTTCAG
GCAGATCAAGGAGGAGGAGAACCGGTCGCGGAGGATTATTAAAGGGGTCCAGTACTACGTCCCC
ATCTTCGAGTGGCTCCCAAATTACACACTGCGCCTCTTCATTTCCGACTTTATATCTGGCCTCA
CCATCACCAGCCTCGCCATCCCTCAAGGCATTAGCTATGCCAAGCTTGCCAATCTTCCTCCCAT
TGTTGGCCTTTATTCCAGCTTTGTCCCACCTCTGGTGTATGCTATTTTTGGGAGTTCCAGGCAC
ATGGCCGTGGGGACGTTAGCAGCAGCATCATTGCTAATAGGCCAAACTATATCAACCGTCGCAA
GCCCAGAAACGGACCCAACTTTGTACCTTCATTTGATCTTCACTACCACTTTTATCACCGGAGT
TTTCCAGGCTTGCTTGGGTATTTTTAGGCTAGGGATATTGGTGGATTTCTTTTCCCATTCTACC
ATCACTGGATTCATGGGAGGGACAGCATTTATTCTCATTGCCCAACAGTTAAAGGGATTTTTTG
GTATGAAACATTTTTCAACCAAAACCAACCTTGTTGAAGTGGCCAAGAGCATCATCACCAATAG
ACATGAGATAAGGTGGGAAACCACAGTTCTTGGACTGGTCTTCCTTGCTTTTCTACAATTTACC
AGACACGTGAGGAACAAGAGACCAAAACTCTTTTGGGTATCAGCTATAGCTCCAATGACTGTAG
TGATAGTTGGTAGCATTTTTGTGTACCTCGTCCATGGCCAAAAACATGGAATTCCAATCGTGGG
CCATCTAGATAGAGGGTTAAATCCTTGGTCCATTCAGTATTTCAACTTTGACAGTAAATATTTG
CCAGCAGTGATGCAAGCTGCCCTTATCACTGGAGTCTTGTCATTGGCGGAAGGAATAGCAATTG
GAAGAAGCTTTTCTGTTACTGATAACACACCTCATGATGGGAACAAAGAAATGGTAGCTTTTGG
CCTCATGAACCTATTTGGTTCTTTCACTTCTTGTTACTTGACTAGTGGACCATTTTCCAAGACA
GCTGTGAATTACAATGCAGGGGGTAAAACTGCCATGACAAACGTGGTCCAAGCAGTCCTCATGG
CGTTGACACTGCAATTTTTGGCACCACTATTTGGCTTCACACCTCTTGTTGCACTATCAGCTAT
CATTACATCTGCAATGTTGGGACTGGTTAACTATACCGAAGTTATCTATCTCTACAAAGTTGAC
AAATTTGATTTCGTTATTTGCATGGCTGCCTTCTTGGGAGTTGCCTTCTTGGGCATGGATTATG
GTCTCATGATCTCTGTTGGACTTGGTGTGATTAGAGCACTGTTATATGTTGCTAGACCTGCAAC
ATGCAAGCTTGGAAAGTTAAATGAGTTTGGTATATACAGAGATGTTGAGCAATACCCTGCTTCA
ACTTTCCCAGGATTGATCATTGTTCAGCTTGGCTCCCCTGTTTACTTTTCAAATTCTGTGTACG
TCAAAGAAAGGGTTATGAGGTATATTAAAAGTCAGCAAAGATCTAATGAAGATGTTGTTGAGCA
AGTCATACTCGATATGTCAGGAGTGACATCCATTGACACAACTGCAATTGAAGGATTGTTGGAG
CTAAACAAATGTTGGAAAAGAATGGAATTGAGATGTTTTTGGTAAACCCAAGGCTGGAGGTTA
TGGAGAAACTAATAATATCCAAGTTTGTTGACAAGCTTGGGAAGGAATCGTTCTATCTAACATT
GGACGATGCAGTGAAGGCAAGCCAATACTCACTCAAGAAAAATGATAATGGAGACATAGTTCAT
GAAACTAGTCATGCTTAA
```

Domesticated sequence

```
cactctgtggtctcaaatgGGTACGATAGGCAACAATTCCCATGAGGGTGATCATCATGGGGTG
AACTTCACAGCCCAGAGAGGGTTTTACACAAAACTCAAATCCGGGTTGAAGGAAACGTTCTTCC
CTGATGATCCGTTCAGGCAGATCAAGGAGGAGGAGAACCGGTCGCGGAGGATTATTAAAGGGGT
CCAGTACTACGTCCCCATCTTCGAGTGGCTCCCAAATTACACACTGCGCCTCTTCATTTCCGAC
TTTATATCTGGCCTCACCATCACCAGCCTCGCCATCCCTCAAGGCATTAGCTATGCCAAGCTTG
CCAATCTTCCTCCCATTGTTGGCCTTTATTCCAGCTTTGTCCCACCTCTGGTGTATGCTATTTT
TGGGAGTTCCAGGCACATGGCCGTGGGGACGTTAGCAGCAGCATCATTGCTAATAGGCCAAACT
ATATCAACCGTCGCAAGCCCAGAAACGGACCCAACTTTGTACCTTCATTTGATCTTCACTACCA
CTTTTATCACCGGAGTTTTCCAGGCTTGCTTGGGTATTTTTAGGCTAGGGATATTGGTGGATTT
CTTTTCCCATTCTACCATCACTGGATTCATGGGAGGGACAGCATTTATTCTCATTGCCCAACAG
TTAAAGGGATTTTTTGGTATGAAACATTTTTCAACCAAAACCAACCTTGTTGAAGTGGCCAAGA
GCATCATCACCAATAGACATGAGATAAGGTGGGAAACCACAGTTCTTGGACTGGTGTTCCTTGC
TTTTCTACAATTTACCAGACACGTGAGGAACAAGAGGCCAAAACTCTTTTGGGTATCAGCTATA
GCTCCAATGACTGTAGTGATAGTTGGTAGCATTTTTGTGTACCTCGTCCATGGCCAAAAACATG
GAATTCCAATCGTGGGCCATCTAGATAGAGGGTTAAATCCTTGGTCCATTCAGTATTTCAACTT
TGACAGTAAATATTTGCCAGCAGTGATGCAAGCTGCCCTTATCACTGGAGTCTTGTCATTGGCG
```

```
GAAGGAATAGCAATTGGAAGAAGCTTTTCTGTTACTGATAACACACCTCATGATGGGAACAAAG
AAATGGTAGCTTTTGGCCTCATGAACCTATTTGGTTCTTTCACTTCTTGTTACTTGACTAGTGG
ACCATTTTCCAAGACAGCTGTGAATTACAATGCAGGGGGTAAAACTGCCATGACAAACGTGGTC
CAAGCAGTCCTCATGGCGTTGACACTGCAATTTTTGGCACCACTATTTGGCTTCACACCTCTTG
TTGCACTATCAGCTATCATTACATCTGCAATGTTGGGACTGGTTAACTATACCGAAGTTATCTA
TCTCTACAAAGTTGACAAATTTGATTTCGTTATTTGCATGGCTGCCTTCTTGGGAGTTGCCTTC
TTGGGCATGGATTATGGACTCATGATCTCTGTTGGACTTGGTGTGATTAGAGCACTGTTATATG
TTGCTAGACCTGCAACATGCAAGCTTGGAAAGTTAAATGAGTTTGGTATATACAGAGATGTTGA
GCAATACCCTGCTTCAACTTTCCCAGGATTGATCATTGTTCAGCTTGGCTCCCCTGTTTACTTT
TCAAATTCTGTGTACGTCAAAGAAAGGGTTATGAGGTATATTAAAAGTCAGCAAAGATCTAATG
AAGATGTTGTTGAGCAAGTCATACTCGATATGTCAGGAGTGACATCCATTGACACAACTGCAAT
TGAAGGATTGTTGGAGCTAAACAAAATGTTGGAAAAGAATGGAATTGAGATGTTTTTGGTAAAC
CCAAGGCTGGAGGTTATGGAGAAACTAATAATATCCAAGTTTGTTGACAAGCTTGGGAAGGAAT
CGTTCTATCTAACATTGGACGATGCAGTGAAGGCAAGCCAATACTCACTCAAGAAAAATGATAA
TGGAGACATAGTTCATGAAACTAGTCATGCTggtgtgagaccacgaagtg
```

EC64094

Native sequence

AAGAAGGGTAAGGTGATCCCCTCTTAGCGGTTCATGAAGTGCCCAAATTCTTCTCAACCCCTTT
GCTTCATGAAGTGTCAAAGTCTTCCATGATCATGTTATTAACATTGTGGATGTTGATGATGATG
GGAATTGCAGATATAAGAGTGTAGATGGGTATGGGGCAAGAAAGTTGGCCTGAAGTGCTAAGAA
CAATGATAAATCAGCTCACCATACGCAATTCTACCTATCTTAGATTGTATGGCACCAACAAATT
GTACCAACCTGTCCTACAAGCCTGATTATTGCTCCAAATAAGATTGCAATAGAGGACAAGTGGT
TGAGCATTTCGGATATAGACCACCTTGTTGCTTCGGTGTATCCGGCTGTGTTGGTGACACTATC
ACGTAAGAGATTCTCTATAGTGAGAGTTTTGGCTAACTGCAATAGTTAACAGACTTCTAACCAT
TAACCAACATCTCAGCTCTATGCAGTCCACTCAAGATTATGGGATGTTTGAATCTTTCATTCTA
AGAATATTTGTCTTGTCTATGCATCGAGTGGTCTCAACTCTCACGCTTGTGATAGCTGCTCCTA
ATATTAAGGAACTGTCCAATTTATTCTTTCAACATTAATCACTTTTAACCATTGCAACAGACAC
TTACTAAATTTCTTTAACATACCTTAAAATTACTTGATTTTTCATATTTTAAAAAGTGTGTCTC
AAATAGTGTTGTTGGTGTAATAAATTCTTAAATTTTTAAGATTATATTTTAACGTGTTTCAAAA
AAGATTATATTTTAACAATTCTAAATATTACTTTGTTATCAGAATTTTATAAATAATAATTTCT
CCAATTTAAGATCTTAAGAACAAATTCTTATTTTAAAGATGAAAATGAATTAAAGTATAATCTT
TAATGTTAAACTTTGACTCCGAGTAAATATTTGAAAATATCTTCATCCTCAGACCCAAGCCTAT
GTACTTGTTCACCGTGACGAATCAAGAAATTATATCTTGTTAAGGAAAAACATACATATAATGA
TATAAATTTGTAACATGTTACAATTAAAAGTTTTTTTTTTTGGTAAGTAGGAAAATTACAATTA
AAAGTTTGAATATACATCACAATTATTCTATACAAGATTTTATATTTGAGAAACAGTAGAAAGC
TTTGCTTAAACAAAATTCATTTTTTATATAATTGGACGTAGTGACAGACCCAATTTGCAAGGAA
ACAGGACAATCGCCCCTCATGAGAAAAATATTTCTTCATGTATTATTTGTATCGTATATAAATT
GAGAAATGAGAAGTGCACGGAAAATGTAAACTTGTTTTAAACATTCACCCCGATTTGTAACCA
CTTAATCTTATGTGACATACTGTCGTTTTACACTACGGTGTATATCCATTAATCTCATATAAAT
TCTCCAATTTTTCAACAAAAACCTTATCCTTTTTCCATTTTCCATTTTTAAAAAATACCAAAAA
CTAATTCAATATTGACTAACCGTATTAATTACATAGATTTGGATGACAAAGATCCCTCATATGA
TCCAACTCTTGTTGTTAGCCGTTAGGGTCAGATCCTAACCGATCCATGTAGCTGTTTTAATGTA
TTCATGGCAAAGATAAAGATCAGAGACAACGCATTGCTGAACTGGAGAAGACTCTCAAACATAT
AAATGGAAAGATTGTCAACAAAATCACATCCAATTTACTCTCCAATTCTTCGTTCTTCATAATT
TCTCTTTTGAGAAAAGAAAACATTAAGACCCCATACCCATAGACGAGTTGGACTCTATAAGGAA
AAAGCTACACCTATCCGCTACTAACAATCCTGTGTCAACCACACGTGTGTGAGAAATACCACTC
GACTATTCTTTATTAATAACCCTGTCTATAAATATACCCGCAACACAAGTACTATCCCTGCCTT
TTTGATTCTCTTGCAGCGTGTGTGTGCTGCTCTTGTCTATTCTATATAACCCATTTCATTAGTT
TGTTTCAAACACCATC

Domesticated sequence

cactctgtggtctcaggagAAGAAGGGTAAGGTGATCCCCTCTTAGCGGTTCATGAAGTGCCCA
AATTCTTCTCAACCCCTTTGCTTCATGAAGTGTCAAAGTCTTGCATGATCATGTTATTAACATT
GTGGATGTTGATGATGATGGGAATTGCAGATATAAGAGTGTAGATGGGTATGGGGCAAGAAAGT
TGGCCTGAAGTGCTAAGAACAATGATAAATCAGCTCACCATACGCAATTCTACCTATCTTAGAT
TGTATGGCACCAACAAATTGTACCAACCTGTCCTACAAGCCTGATTATTGCTCCAAATAAGATT
GCAATAGAGGACAAGTGGTTGAGCATTTCGGATATAGACCACCTTGTTGCTTCGGTGTATCCGG
CTGTGTTGGTGACACTATCACGTAAGAGATTCTCTATAGTGAGAGTTTTGGCTAACTGCAATAG
TTAACAGACTTCTAACCATTAACCAACATCTCAGCTCTATGCAGTCCACTCAAGATTATGGGAT
GTTTGAATCTTTCATTCTAAGAATATTTGTCTTGTCTATGCATCGAGTCGTCTGAACTCTCACG
CTTGTGATAGCTGCTCCTAATATTAAGGAACTGTCCAATTTATTCTTTCAACATTAATCACTTT
TAACCATTGCAACAGACACTTACTAAATTTCTTTAACATACCTTAAAATTACTTGATTTTTCAT
ATTTTAAAAAGTGTGTCTCAAATAGTGTTGTTGGTGTAATAAATTCTTAAATTTTTAAGATTAT
ATTTTAACGTGTTTCAAAAAAGATTATATTTTAACAATTCTAAATATTACTTTGTTATCAGAAT
TTTATAAATAATAATTTCTCCAATTTAAGATCTTAAGAACAAATTCTTATTTTAAAGATGAAAA

TGAATTAAAGTATAATCTTTAATGTTAAACTTTGACTCCGAGTAAATATTTGAAAATATCTTCA
TCCTCAGACCCAAGCCTATGTACTTGTTCACCGTGACGAATCAAGAAATTATATCTTGTTAAGG
AAAAACATACATATAATGATATAAATTTGTAACATGTTACAATTAAAAGTTTTTTTTTTTGGTA
AGTAGGAAAATTACAATTAAAAGTTTGAATATACATCACAATTATTCTATACAAGATTTTATAT
TTGAGAAACAGTAGAAAGCTTTGCTTAAACAAAATTCATTTTTTATATAATTGGACGTAGTGAC
AGACCCAATTTGCAAGGAAACAGGACAATCGCCCCTCATGAGAAAATATTTCTTCATGTATTA
TTTGTATCGTATATAAATTGAGAAAATGAGAAGTGCACGGAAATGTAAACTTGTTTTAAACAT
TCACCCCGATTTGTAACCACTTAATCTTATGTGACATACTGTCGTTTTACACTACGGTGTATAT
CCATTAATCTCATATAAATTCTCCAATTTTTCAACAAAAACCTTATCCTTTTTCCATTTTCCAT
TTTTAAAAAATACCAAAAACTAATTCAATATTGACTAACCGTATTAATTACATAGATTTGGATG
ACAAAGATCCCTCATATGATCCAACTCTTGTTGTTAGCCGTTAGGGTCAGATCCTAACCGATCC
ATGTAGCTGTTTTAATGTATTCATGGCAAAGATAAAGATCAGAGACAACGCATTGCTGAACTGG
AGAAGAGTCTCAAACATATAAATGGAAAGATTGTCAACAAAATCACATCCAATTTACTCTCCAA
TTCTTCGTTCTTCATAATTTCTCTTTTGAGAAAAGAAAACATTAAGACCCCATACCCATAGACG
AGTTGGACTCTATAAGGAAAAAGCTACACCTATCCGCTACTAACAATCCTGTGTCAACCACACG
TGTGTGAGAAATACCACTCGACTATTCTTTATTAATAACCCTGTCTATAAATATACCCGCAACA
CAAGTACTATCCCTGCCTTTTTGATTCTCTTGCAGCGTGTGTGTGCTGCTCTTGTCTATTCTAT
ATAACCCATTTCATTAGTTTGTTTCAAACACCATCaatgtgagaccacgaagtg

EC64095

Native sequence

```
CAAAAGTCGTAGGTGGATCAACCGACGATGCTTTAGGTACACAAGGAGGAGCCACGCATAATGA
CATTCATAACATTTGTGAAGGTACATTCTATCTTTGTTGGTCTGACCATCTTGATATTGACGAT
GACGGTTTATCCATATCATTGTCATGACTCTTTTAAGATCTGAGGTTCATACATCAAATATACT
CTGACCTAGAGAATCTGATTTTTCAGTGGTAAAGAGCTCATTGTCCCACTAGACATACCAAATT
TGTTGTTGAACTAGAGATTCAGCCAAGCTATTGGCTGATAGTCCTCAACAATGATTCCCTTTTT
TCAAACAATTATATGGAGGAAGAAATTTTTTCCCCTCAACTTTGATTTCATTAACTATTCCATT
AGAGTGATAAGTACGAATATTTGGATTTTAGCTACTAAACTAAAGATAAAGGCGGGAGGCTAAA
CCAATAAAATTGTGTAAATAACTTGAATTTGATTGCATGTATGCGTTAATTGCATTGTACATGA
TTAGTATTTATGGGAAAAATTATTCTTCCTTGCCAAACGACATGGCATATAGCATTTAGTCTTA
CTCGGCTATTTGCGCATGTAGTGCATAAATATTCCCTCATTGAGGGATGCATAGCGACCCACGT
TCATCCGACAACTTTCCACTTCACATCATAATTCGTTGACAAAGTCCAATGTACACGTCCTCTC
TTGAAACACGTGTTCATGTCAATTACTCCTTAAAGAGTTAGTTGTGTTGAAACCACGAATCATG
AGTTAAATGCATCATTGTCCAGTTCTACTGAACTTTGACTAAGTCTGATATCTTTAAGACTGAA
ATACTCTTCCGTAGACTCGAGAAGAATTGATTAACCCTTGTCCGGTTGCAATAATGCTCCCCT
AACATTTTTTTTTGTAGATAGAAGAGATGACAAAGCCATCATAAACTCACACACACAAAGTGGA
GGAATCGAGGTTCGAACCCCCGATCATGACGTCCGGTCTAACAATTTCAGCATTTTTGCCAGTT
GAGCTATGACTTATGAGACTGCTCCCCTAACATTCTATCGTCCGAAAAAAATATCTTTTATTCT
TAAATTGCATATTTGCCGATTTCAACTAATATAATAATTATTTTTATATATAAAATATATAACA
ATTAATTACTAATATTGACCGTTTAAAAAAAATGTGAAATGGAGTGGTGTGCTTTTTTTCAATT
AGTTGGAAACCCGCTATTTTCTTTTTGAGGAAAAACGAACATATAGTAAGCTGGACTTTCTGCC
CTCCTTTGGGATATACTCACACTTTCGTCTCTTTTGACTTGAGACTAATCCTCGGTGCGACCGC
ATAAACCAACTGAAATTGATCTTTTGAACCTTTACGAACCCTTCAATTAGGACATTTCGTTGAG
ATCTAAGTCCAACTCCAATGATTTGCAGAATATAATATTCTAATCTAAAATACATGATGCCTTG
TGTATAAATAAATCTAATTAATTAAGCTGCACAAATTAAACTTTTTTTGGCCCATTCACAATTC
CATACTCCACAACTACATAGTTTACCCCCTCGAAAAATAAATATTACATAGTTTACCATCTTTT
TTGCCCATTCACGTTGGCCTATGTATACTGTATAAATATTTTAACTATCATGAAAGGTCATTGA
CGAAATTGTTTATTTTCTTTTATTTACAGATTGTACTACCTGGGATTTAATTTGTCCTTGAAGT
ATACTATATGAAGATTAATCTATAATTTTTTAAAATACTCTTACATTTTATTATTGAATACATT
TTTATTAAGGTCGCATTAAATATACTTAATAAAAAAATATCCCATGATACAAAAAAATGTATCT
AGTGATTGCCATTTTATACGTGACGCCAGGTTATTAATTTTTTTCCTTTGCCCCTTCGAAATAT
CCAAACCATATATTTTCATTTTTACCAAATGCATTATTTGTATAAATAAACAGACCAATCACAC
ATTCATGATTTATCCCTAAGATCTCCAATTTTTTCTTATTAAGAAGCTACTTTTACTCTCTCTT
TTTTTTCGAGCTTTACCTTGTTTTCTTAATTTATTTTTTTTAGTACCA
```


Domesticated sequence

```
cactctgtggtctcaggagCAAAAGTCGTAGGTGGATCAACCGACGATGCTTTAGGTACACAAG
GAGGAGCCACGCATAATGACATTCATAACATTTGTGAAGGTACATTCTATCTTTGTTGGTCTGA
CCATCTTGATATTGACGATGACGGTTTATCCATATCATTGTCATGACTCTTTTAAGATCTGAGG
TTCATACATCAAATATACTCTGACCTAGAGAATCTGATTTTTCAGTGGTAAAGAGCTCATTGTC
CCACTAGACATACCAAATTTGTTGTTGAACTAGAGATTCAGCCAAGCTATTGGCTGATAGTCCT
CAACAATGATTCCCTTTTTTCAAACAATTATATGGAGGAAGAAATTTTTTCCCCTCAACTTTGA
TTTCATTAACTATTCCATTAGAGTGATAAGTACGAATATTTGGATTTTAGCTACTAAACTAAAG
ATAAAGGCGGGAGGCTAAACCAATAAAATTGTGTAAATAACTTGAATTTGATTGCATGTATGCG
TTAATTGCATTGTACATGATTAGTATTTATGGGAAAAATTATTCTTCCTTGCCAAACGACATGG
CATATAGCATTTAGTCTTACTCGGCTATTTGCGCATGTAGTGCATAAATATTCCCTCATTGAGG
GATGCATAGCGACCCACGTTCATCCGACAACTTTCCACTTCACATCATAATTCGTTGACAAAGT
```

```
CCAATGTACACGTCCTCTCTTGAAACACGTGTTCATGTCAATTACTCCTTAAAGAGTTAGTTGT
GTTGAAACCACGAATCATGAGTTAAATGCATCATTGTCCAGTTCTACTGAACTTTGACTAAGTC
TGATATCTTTAAGACTGAAATACTCTTCCGTAGACTCGAGAAAGAATTGATTAACCCTTGTCCG
GTTGCAATAATGCTCCCCTAACATTTTTTTTTGTAGATAGAAGAGATGACAAAGCCATCATAAA
CTCACACAGACAAAGTGGAGGAATCGAGGTTCGAACCCCCGATCATGACGTCCGGTCTAACAAT
TTCAGCATTTTTGCCAGTTGAGCTATGACTTATGAGACTGCTCCCCTAACATTCTATCGTCCGA
AAAAAATATCTTTTATTCTTAAATTGCATATTTGCCGATTTCAACTAATATAATAATTATTTTT
ATATATAAAATATATAACAATTAATTACTAATATTGACCGTTTAAAAAAAATGTGAAATGGAGT
GGTGTGCTTTTTTTCAATTAGTTGGAAACCCGCTATTTTCTTTTTGAGGAAAAACGAACATATA
GTAAGCTGGACTTTCTGCCCTCCTTTGGGATATACTCACACTTTGCTCTCTTTTGACTTGAGAC
TAATCCTCGGTGCGACCGCATAAACCAACTGAAATTGATCTTTTGAACCTTTACGAACCCTTCA
ATTAGGACATTTCGTTGAGATCTAAGTCCAACTCCAATGATTTGCAGAATATAATATTCTAATC
TAAAATACATGATGCCTTGTGTATAAATAAATCTAATTAATTAAGCTGCACAAATTAAACTTTT
TTTGGCCCATTCACAATTCCATACTCCACAACTACATAGTTTACCCCCTCGAAAAATAAATATT
ACATAGTTTACCATCTTTTTTGCCCATTCACGTTGGCCTATGTATACTGTATAAATATTTTAAC
TATCATGAAAGGTCATTGACGAAATTGTTTATTTTCTTTTATTTACAGATTGTACTACCTGGGA
TTTAATTTGTCCTTGAAGTATACTATATGAAGATTAATCTATAATTTTTTAAAATACTCTTACA
TTTTATTATTGAATACATTTTTATTAAGGTCGCATTAAATATACTTAATAAAAAAATATCCCAT
GATACAAAAAAATGTATCTAGTGATTGCCATTTTATACGTGACGCCAGGTTATTAATTTTTTTC
CTTTGCCCCTTCGAAATATCCAAACCATATATTTTCATTTTTACCAAATGCATTATTTGTATAA
ATAAACAGACCAATCACACATTCATGATTTATCCCTAAGATCTCCAATTTTTTCTTATTAAGAA
GCTACTTTTACTCTCTCTTTTTTTTCGAGCTTTACCTTGTTTTCTTAATTTATTTTTTTTAGTA
CCAaatgtgagaccacgaagtg
```

EC64096

Native sequence

```
CATCACGACTTTCGGTTCTTGCTTTTAGGTATACTCATGTATATTCTTTAAAATTATGAATTAT
TTTGTAACAATCGTGTTTAGTATATTATAAAAATCTTTCTTGCAAAGTTTAATTGTTTTAACA
AGAGATGATTAAATATGAATTTGATTAAATATGAATTTTCAAACTTTTGGCACATGTAAATTCA
AATTTAATTCATATTTTGTTAAAAAAAAATCAAAACATTTGTGGGAGGATTGTTATTATATTAT
AAACATTTCTAAAAAAATCATTAAAAAAATGTGAAAATTACATGTGAGATGACTTAAAAACAA
AGACCAAAAGTCGTGACGGAAAACATGTAGGGACTAAAAACTTGCCGAAAATTTTTAGAATGAT
TAAAATTGAAACTTAAGGTATTTATCGGGATCAAAAATTTATTTAACCCTTTAAACTATTAGTA
AAATGAAGATGACAGTTTTTTTTTTTTGTGGTCGAGGTTTGAACCTCAGACCTTACATATATTA
TGCATTGTCCTCGTCAACGAGTTAAGCTCACGAGGACATGAAGATGACAGTTTGCATAAACAAC
AAAAGAAAATAATAATGCAAAATAGTAAATAAAAATAAAATAGCTAACATAGAGGAAGAGAGG
ATTAAAGAAAGAAATGATGCAATAATTAATGTGCACAAGAGAAATAATTGAATAAGATTTACAT
ATTATAATGTGAAACAGTAATAAAATAAAATAAAAACAATCACCCTTTGTAAACAATGGTTAA
ATTTAGAGCTAATTACATTCTCCTCTCTTCAAAGATGCTTGAATTATATTCTCATCTCCTCTTA
TATTAAAATATACACTTCTCTCTTCTCTTATTCAAAACATATTAGAAAATATTTCACTCCCTTC
CCTTGAGAGAAGTTAGAATTACATTCTCCTCATCTCTTATGTTAAAATCTACGCTTTAGTCCCT
CGATAGTTTTTTTTTCCTAATAATCTAGCAAAAAAATAATATAACATACTTCAAATTTTGAATA
CATATCATTGTCTATTTTTTATTCACAATTTCATTAACATATAGTTTTAAACTCAATTATAAAA
TATGAAACAACCAGAATAAAATTAAAATATGTGAAAAATTACTAAAGTCGACTAAGTATGGTTA
TTAAAAATTGAATTTTATACTCTAAATTATAAAATTAATAACAAAATATTTAAATTTAATTGTC
ATTGACATTTAGATTATGAGTTTAATAGTGGTGGGTTGACAGTCCAAAACAGTTTTTCACAAAC
AAACAATCAAAACATTCTGAATTGACACATCAATTAATGATTTTAATTTATGTGGGCCAAATGT
GGCAATTCAGAATGTTTTGATTGGTTGTGTGTGCAATATTGTTTTGAACTGTCAAAAGAAATGA
ATGAATATTTTTTAACTGGTGATTCAAAATTAATATATAATATAAAAATATTTATTTATATTAA
AATAGATTAAAGTGTAGTCCATATTTAATTATTAAGGGAGAGTTCGTAATTCAAGCATCTTATA
GAAGAGGGGAGTATAAATTTTACTTTTTAAGGAAGGAGGTGTATATTATAACATAAGAGGGAGA
CGTGTAATTCAAGCATCTTTGAGAGGAGGAGAATGTGGTTTACTCTTTAATTTATTCATGTAAC
CTTTGGTTTTCTTATTGGAGGAGTAAGAACTAAGAAGAATTATCATTTGTCACTATCAACAAAG
TGACTGGAAATTGGGGCGTAAATCGGTGAGAACTATATTTTGATCCTTATAGTTCTCACAAGCC
AAGTCAAACAAACTTTTTATTTGAATGGCTCATAAAAAATGAAACGGTCCGATAATATTCTCTC
TATGCAAGTGACATATTTATATTTATATTCTCTCTTTTATTATTAGCTTTTTTATAGGGAATT
TTTATTGGTTTTTTTTTTTGTGAGGTGCTATTTTATAATCAAGAGAGTTTATTTATTTTCATCCG
ATTTCCTTTGCATCTCATTTCTCTAGGATTATGATTAATTTCATGTTAACTAAACAGAGAGTTA
TAGCACACTAAGAAACTAGCTAGAATGCAGCACCTACTTTCACCACCAGAAATGAGAAAGAGCA
AACTAGTTTATCACAACAGCTACCATTCTATACTTAGATAACGAGGTAAACCACACTAAACAAA
TTGACACCGATCGGATTGACGTATTTGAGTTTATCTACTGATGTAAACATTTGTGAAATTGTTT
TGGAGAGTTTATGGAAATAGCTTATGATATATTTATACGATTTTTTTAGATTATTTTCACAAGT
TGTCTAGGATAGCTTATGAAATAATGTATACCTTTTATGAAAAGAATTTGACTTATTGTATCT
CTTGCTATAGAAATAGCTTATAAATAAGTACTTATTTGACAAGCGCTTAATTAAGTTGTTTATC
GAAACGTGCTCAAAGAAGTTAATTCATTGTGGATTGAATTTTTCAGGTATTATTCAGAGGTGGC
CAACTTCAAAGTGCTCTCAACACTATCATCAACATTACCTAACTTTGATCATACTTCTATATAA
TAATCC
```

Domesticated sequence

```
cactctgtggtctcaggagCATCACGACTTTCGGTTCTTGCTTTTAGGTATACTCATGTATATT
CTTTAAAATTATGAATTATTTTGTAACAATCGTGTTTAGTATATTATAAAAATCTTTCTTGCAA
AAGTTTAATTGTTTTAACAAGAGATGATTAAATATGAATTTGATTAAATATGAATTTTCAAACT
TTTGGCACATGTAAATTCAAATTTAATTCATATTTTGTTAAAAAAAAATCAAAACATTTGTGGG
AGGATTGTTATTATATTATAAACATTTCTAAAAAAATCATTAAAAAAATGTGAAAATTACATG
```

```
TGAGATGACTTAAAAACAAAGACCAAAAGTCGTGACGGAAAACATGTAGGGACTAAAAACTTGC
CGAAAATTTTTAGAATGATTAAAATTGAAACTTAAGGTATTTATCGGGATCAAAAATTTATTTA
ACCCTTTAAACTATTAGTAAAATGAAGATGACAGTTTTTTTTTTTTGTGGTCGAGGTTTGAACC
TCAGACCTTACATATATTATGCATTGTCCTCGTCAACGAGTTAAGCTCACGAGGACATGAAGAT
GACAGTTTGCATAAACAACAAAGAAAATAATAATGCAAAATAGTAAATAAAAATAAAATAGC
TAACATAGAGGAAGAGAGGATTAAAGAAAGAAATGATGCAATAATTAATGTGCACAAGAGAAAT
AATTGAATAAGATTTACATATTATAATGTGAAACAGTAATAAAATAAAATAAAAACAATCACC
CTTTGTAAACAATGGTTAAATTTAGAGCTAATTACATTCTCCTCTCTTCAAAGATGCTTGAATT
ATATTCTCATCTCCTCTTATATTAAAATATACACTTCTCTCTTCTCTTATTCAAAACATATTAG
AAAATATTTCACTCCCTTCCCTTGAGAGAAGTTAGAATTACATTCTCCTCATCTCTTATGTTAA
AATCTACGCTTTAGTCCCTCGATAGTTTTTTTTTCCTAATAATCTAGCAAAAAAATAATATAAC
ATACTTCAAATTTTGAATACATATCATTGTCTATTTTTTATTCACAATTTCATTAACATATAGT
TTTAAACTCAATTATAAAATATGAAACAACCAGAATAAAATTAAAATATGTGAAAAATTACTAA
AGTCGACTAAGTATGGTTATTAAAAATTGAATTTTATACTCTAAATTATAAAATTAATAACAAA
ATATTTAAATTTAATTGTCATTGACATTTAGATTATGAGTTTAATAGTGGTGGGTTGACAGTCC
AAAACAGTTTTTCACAAACAAACAATCAAACATTCTGAATTGACACATCAATTAATGATTTTA
ATTTATGTGGGCCAAATGTGGCAATTCAGAATGTTTTGATTGGTTGTGTGTGCAATATTGTTTT
GAACTGTCAAAAGAAATGAATGAATATTTTTTAACTGGTGATTCAAAATTAATATATAATATAA
AAATATTTATTTATATTAAAATAGATTAAAGTGTAGTCCATATTTAATTATTAAGGGAGAGTTC
GTAATTCAAGCATCTTATAGAAGAGGGGAGTATAAATTTTACTTTTTAAGGAAGGAGGTGTATA
TTATAACATAAGAGGGACACCTGTAATTCAAGCATCTTTGAGAGGAGGAGAATGTGGTTTACTC
TTTAATTTATTCATGTAACCTTTGGTTTTCTTATTGGAGGAGTAAGAACTAAGAAGAATTATCA
TTTGTCACTATCAACAAAGTGACTGGAAATTGGGGCGTAAATCGGTGAGAACTATATTTTGATC
CTTATAGTTCTCACAAGCCAAGTCAAACAAACTTTTTATTTGAATGGCTCATAAAAAATGAAAC
GGTCCGATAATATTCTCTCTATGCAAGTGACATATTTATATTTATATTCTCTCTTTTTATTATT
AGCTTTTTTATAGGGAATTTTTATTGGTTTTTTTTTTGTGAGGTGCTATTTTATAATCAAGAGA
GTTTATTTATTTTCATCCGATTTCCTTTGCATCTCATTTCTCTAGGATTATGATTAATTTCATG
TTAACTAAACAGAGAGTTATAGCACACTAAGAAACTAGCTAGAATGCAGCACCTACTTTCACCA
CCAGAAATGAGAAAGAGCAAACTAGTTTATCACAACAGCTACCATTCTATACTTAGATAACGAG
GTAAACCACACTAAACAAATTGACACCGATCGGATTGACGTATTTGAGTTTATCTACTGATGTA
AACATTTGTGAAATTGTTTTGGAGAGTTTATGGAAATAGCTTATGATATATTTATACGATTTTT
TTAGATTATTTTCACAAGTTGTCTAGGATAGCTTATGAAATAATGTATACCTTTTATGAAAGA
ATTTGACTTTATTGTATCTCTTGCTATAGAAATAGCTTATAAATAAGTACTTATTTGACAAGCG
CTTAATTAAGTTGTTTATCGAAACGTGCTCAAAGAAGTTAATTCATTGTGGATTGAATTTTTCA
GGTATTATTCAGAGGTGGCCAACTTCAAAGTGCTCTCAACACTATCATCAACATTACCTAACTT
TGATCATACTTCTATATAATAATCCaatgtgagaccacgaagtg
```

There are some days when science makes sense,

When results don't just sit on the fence.

But those days are so rare,

That I often despair

And think maybe I'm just a bit dense.