Characterization Of GmSAT1 And Related Proteins From Legume Nodules

Submitted by

David Michael Chiasson

This thesis is submitted in fulfillment of the requirements
for the degree of Doctor of Philosophy

School of Agriculture, Food, and Wine
The University of Adelaide, Waite Campus
Urrbrae, SA, 5064, Australia

August, 2012

Table of Contents

List of Figures

List of Tables

I. Abstract

Nitrogen is an essential nutrient for plant growth and is normally obtained from the soil medium. Legumes are a unique group of plants that acquired the ability to form a symbiosis with nitrogen-fixing bacteria, called rhizobia, enabling growth in nitrogen-poor soils. *GmSAT1,* a predicted bHLH transcription factor from soybean, is essential for nitrogen fixation; however the role of this protein remains elusive (Kaiser et al., 1998; Loughlin, 2007).

In this work, a further functional characterization of GmSAT1 was undertaken. Using the promoters of known upregulated genes in yeast upon expression of *GmSAT1*, it was found that purified GmSAT1 directly interacts with DNA. Further, GFP-fusion analysis in onion epidermal cells, found that GmSAT1 localizes to the nucleus, as well as peripheral vesicles, demonstrating that GmSAT1 is a likely a transcription factor. Residues from both the N- and C-termini required for GmSAT1 activity were also identified by exchanging domains with GmSAT2, a protein that arose during the relatively recent whole-genome duplication in soybean.

Recently, GmSAT1 was shown to be essential for proper nodule development in soybean (Loughlin, 2007). Therefore, a DNA microarray analysis was conducted to identify transcripts that are differentially expressed after silencing of *GmSAT1* by RNA interference (RNAi) in soybean nodules. Of the ninety-five genes downregulated, twelve were associated with the circadian clock, potentially explaining the *GmSAT1* RNAi phenotype. Investigations were also initiated in the model legume *Medicago truncatula* to identify and characterize *GmSAT1* orthologues. Two genes, *MtSAT1* and *MtSAT2* were cloned and analyzed. *MtSAT1* and *MtSAT2* are expressed in roots and the inner cortex of nodules, similar to *GmSAT1*. By *in planta* GFP-fusion analysis, both MtSAT1 and MtSAT2 were found to associate with vesicles and the nucleus. Insertional mutations in either gene alone did not render a phenotype, however downregulating both genes by RNAi disrupted nodule formation.

Using the sequence of a newly discovered ammonium channel protein from yeast (ScAMF1), which is activated by GmSAT1, a novel subfamily of major facilitator transporter proteins (MFSs) from plants was identified. Interestingly, members of the

MFS gene family are found linked with *GmSAT1* loci in soybean, as well as *M. truncatula* and many sequenced dicots. *GmMFS1.3,* a representative from soybean, was cloned and characterized. *GmMFS1.3* expression was localized to the root stele and the inner cortex of nodules. Further, expression of GmMFS1.3 in yeast induced the uptake of methylammonium. Interestingly, *GmMFS1.5* was found to be downregulated in *GmSAT1* RNAi nodules. The link between GmSAT1 and the MFS transporters *in planta* will be the focus of future experiments.

A novel receptor-like kinase protein was also characterized from soybean nodules. GmCaMK1 was identified in a protein interaction screen using conserved calmodulin as bait. The calmodulin-binding domain overlaps the GmCaMK1 kinase subdomain XI, however it was found that GmCaMK1 is able to auto-phosphorylate independent of calmodulin. Therefore, calmodulin binding may influence the interaction of GmCaMK1 with its phosphorylation targets. Taken together, these studies have enriched our knowledge of nitrogen fixation, a critically important component of agricultural practice.

II. Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

The author acknowledges that copyright of published works contained within this thesis (Chapter 5, GmCaMK1 manuscript) resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

…….. ……..
David Chiasson Date

III. Acknowledgements

I would like to acknowledge those people that have contributed to my research. First, I must thank my supervisor, Dr. Brent Kaiser, for supporting my project and allowing me to explore new ideas and possibilities. Brent was always available for discussions on new data and enthusiastic about the potential implications of our discoveries. I am also grateful for the support I received to conduct research overseas and attend conferences in some great locations.

I am also indebted to those who helped me conduct my research. I would like to thank Thomas DeFalco and Dr. Wayne Snedden for the opportunity to collaborate on the GmCamK1 manuscript. It was a great boost early in my PhD to be involved in a project that was leading to an immediate publication. I would also like to recognize Sergey Ivanov, Elena Fedorova, Eric Limpens, and Dr. Ton Bisseling from Wageningen University for help with *Medicago* experiments. I must also acknowledge Manijeh Mohammadidehcheshmeh and Danielle Mazurkiewicz for help with soybean transformations and yeast uptake experiments, respectively.

I would like to thank my wife, Louise Gillis, who lovingly supported me throughout my PhD. She had to cope with my irregular schedule and frequent weekend disappearances to the lab, not to mention living through the ups and downs of my results. She even had to move to The Netherlands for four months with me for research.

Finally, I must also acknowledge the generous support I received through scholarships from the Natural Sciences and Engineering Research Council of Canada (NSERC), The University of Adelaide, the Australian Government.

Cheers!

III. Abbreviations

AA	Amino Acid
bHLH	Basic Helix-Loop-Helix
BLAST	Basic Local Alignment Search Tool
cDNA	Complementary DNA
CaMV 35S	Cauliflower Mosaic Virus Constitutive Promoter
DNA	Deoxyribonucleic Acid
EMSA	Electromobility Shift Assay
ER	Endoplasmic Reticulum
EST	Expressed Sequence Tag
GFP	Green fluorescent protein
GUS	ß-Glucuronidase
kB	Kilobase
kD	Kilodalton
LB	Luria Broth
MA	Methylammonium
OD	Optical Density
PBM	Peribacteroid Membrane
PCR	Polymerase Chain Reaction
PEG	Polyethylene Glycol
RACE	Rapid Amplification of cDNA Ends
RNA	Ribonucleic Acid
RNAi	RNA Interference
SDS	Sodium Dodecyl Sulfate
SDS-PAGE	SDS Polyacrylamide Gel Electrophoresis
TBST	Tris Buffered Saline (with Tween 20)
TMD	Transmembrane Domain
Tris	Tris(hydroxymethyl)aminomethane
UTR	Untranslated Region
v/v	Volume/Volume
w/v	Weight/Volume
X-gluc	5-bromo-4-chloro-3-indolyl-beta-D-glucuronic acid
YNB	Yeast Nitrogen Base

1 Literature Review

1.1 Nitrogen Fixation

Nitrogen is the most abundant mineral element in plant tissues (2% of total dry matter) and is a component of such molecules as DNA, RNA, and proteins (Miller and Cramer, 2005). Nitrogen is present in the soil in such forms as NO_3^-, NH_4^+, peptides, amino acids, and organic compounds. The majority of $NO_3^-$and NH_4^+ is generated by the breakdown of organic matter by soil microorganisms (Miller and Cramer, 2005). In agricultural soils, it has been estimated that NO_3^- is present at a concentration of 1-5 mM and ammonium at 20-200 μM (Owen and Jones, 2001). Plants mainly acquire mineral nitrogen from the soil via uptake of $NO_3^-$and NH_4^+ by high and low affinity transport systems (Glass et al., 2001). As the concentration of available nitrogen is variable, plants must coordinate these transport systems, as well as modify the root architecture (Forde and Lorenzo, 2001).

The earth's atmosphere contains 78% nitrogen in the form of N_2, which can only be utilized by diazotrophs. The ability to "fix" N_2 is restricted to bacteria from both Bacteria and Archaea kingdoms. The process of fixing nitrogen via the enzyme nitrogenase requires 16 ATP, 8 H^+, and 8 e^- to produce 2 NH_3 from one N_2. Importantly, nitrogenase is oxygen-labile, and thus requires an anaerobic environment. Nitrogen-fixing bacteria inhabit a wide variety of habitats, however the soil species belonging to *Rhizobia* and *Frankia* are able to form symbiosis with plants (Masson-Boivin et al., 2009). *Frankia* is a filamentous bacterium that associates with actinorhizal trees and shrubs, such as *Alnus* sp. (alder) and *Casuarina*, in nodules that resemble lateral roots. This symbiosis has enabled actinorhizal plants to colonize nitrogen-poor soils across the globe (Swensen, 1996).

Legumes, including such plants as soybean, alfalfa, and pea, are able to form symbiosis with *Rhizobium* sp. The bacteria gain entry into the roots and are housed within nodules. The plant provides reduced carbon and essential nutrients to the bacteria in exchange for ammonia (Udvardi and Day, 1997). The nodule provides an ideal environment for bacterial nitrogen fixation, having a low oxygen tension, which is maintained by leghemoglobin (Appleby, 1984). Leghemoglobin also provides bound oxygen for bacterial

respiration and is essential for the nitrogen-fixing symbiosis (Ott et al., 2009). A few recent studies have demonstrated that the symbiosis in not entirely mutualistic in all cases. Prell et al. (2009) found that in pea nodules, the plant limits the supply of branched chain amino acids to bacteroids. By controlling the supply, the plant is able to regulate the symbiosis, thus making the bacteroids symbiotic auxotrophs. In *Lotus japonicus*, a homocitrate synthase (FEN1) required for nodulation was recently discovered (Hakoyama et al., 2009). The authors demonstrated that homocitrate is essential for nitrogenase activity, but that rhizobia have lost the enzyme required for its synthesis (NifV). Thus, the rhizobia have become dependent on the host plant in order to fix nitrogen. Further, *Medicago truncatula* produces nodule-specific cysteine-rich peptides that target the housed bacteria (Van de Velde et al., 2010). The peptides are transported into the symbiosome space, causing the bacteroids to become terminally differentiated. Therefore, in some cases the plant is able to dictate the relationship once the bacteria have been successfully recruited from the soil.

1.2 Early Signaling Events in Nodulation

Working mainly with two model legumes, *Lotus japonicus* and *Medicago truncatula*, researchers have begun to dissect the early signaling events in the root epidermis (**Figure 1.1**). Initially, the plant secretes flavonoid compounds in the surrounding soil that are detected by rhizobia. These flavonoids induce the expression of *Nod* genes in the bacteria, leading to the production of Nod factors (Jones et al., 2007). These Nod factors are then perceived by the plant via members of the LysM family on the plasma membrane (Amor et al., 2003; Limpens et al., 2003; Madsen et al., 2003; Radutoiu et al., 2003), leading to such events as Ca^{2+} oscillations, root hair curling, and initiation of cortical cell divisions (Oldroyd and Downie, 2008). The Ca^{2+} oscillations observed in the cell nucleus after the addition of Nod factors are mediated by an ATPase pump, called MCA8 (Capoen et al., 2011). These oscillations are also dependent on potassium channels (CASTOR, POLLUX in *Lotus*, DMI1 in *Medicago*) on the nuclear membrane (Ane et al., 2004; Imaizumi-Anraku et al., 2005; Charpentier et al., 2008). Expression of these potassium channels in kidney cells induces Ca^{2+} oscillations in the nucleus (Venkateshwaran et al., 2012). A calcium- and calmodulin-dependent protein kinase (CCamK) then perceives the Ca^{2+} changes (Lévy et al., 2004; Mitra et al., 2004). CCamK interacts with Cyclops (IPD3 in

Medicago), a phosphorylation target, triggering nodule organogenesis (Messinese et al., 2007; Yano et al., 2008). Lastly, components of the nuclear pore complex are also required for calcium spiking (Kanamori et al., 2006; Saito et al., 2007; Groth et al., 2010).

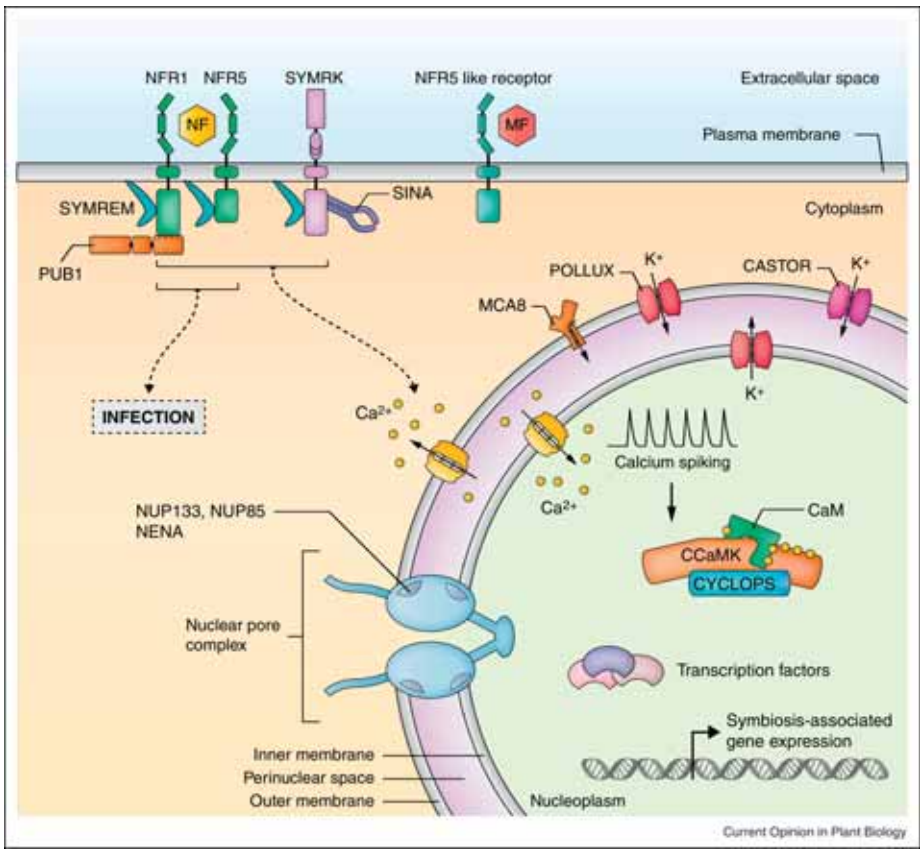


Figure 1.1. Early signaling events in the epidermis of legume roots.
Overview of signal transduction events after Nod factor perception at the plasma membrane by LysM-receptor-like kinases. Nuclear calcium spiking is initiated after perception, a process requiring nuclear membrane-localized potassium and calcium channels, as well as the nuclear pore complex. The calcium signature is detected by the CCamK/Cyclops complex, which is required for bacterial infection and nodule organogenesis (Taken from Singh and Parniske, 2012).

1.3 Nodule Organogenesis

The nodulation process can be divided into two distinct events: root epidermal infection by rhizobia and cortical cell divisions leading to nodule primordia. Rhizobia enter the root via curled root hairs and migrate to the root cortex through infection threads. Mutant screens in *Medicago truncatula* and *Lotus japonicus* have identified genes required for bacterial entry and infection thread growth (Veereshlingam et al., 2004; Yano et al., 2008; Xie et al.,

2012), cortical cell divisions (Limpens et al., 2005; Marsh et al., 2007), and cortical cell differentiation upon bacterial entry (Cheon et al., 1993; Vinardell et al., 2003; Combier et al., 2006). A number of transcription factors of the GRAS ((Kaló et al., 2005; Smit et al., 2005), bZIP (Nishimura et al., 2002), homeobox (Grønlund et al., 2003), NF-YC (Zanetti et al., 2010) and ERF (Middleton et al., 2007; Asamizu et al., 2008) families are also required for nodulation. Recently, Benedito and colleagues (2008) identified 1354 genes that were differentially activated in the nodule, with 473 being nodule specific. Therefore, it is likely that a number of important genes have yet to be characterized.

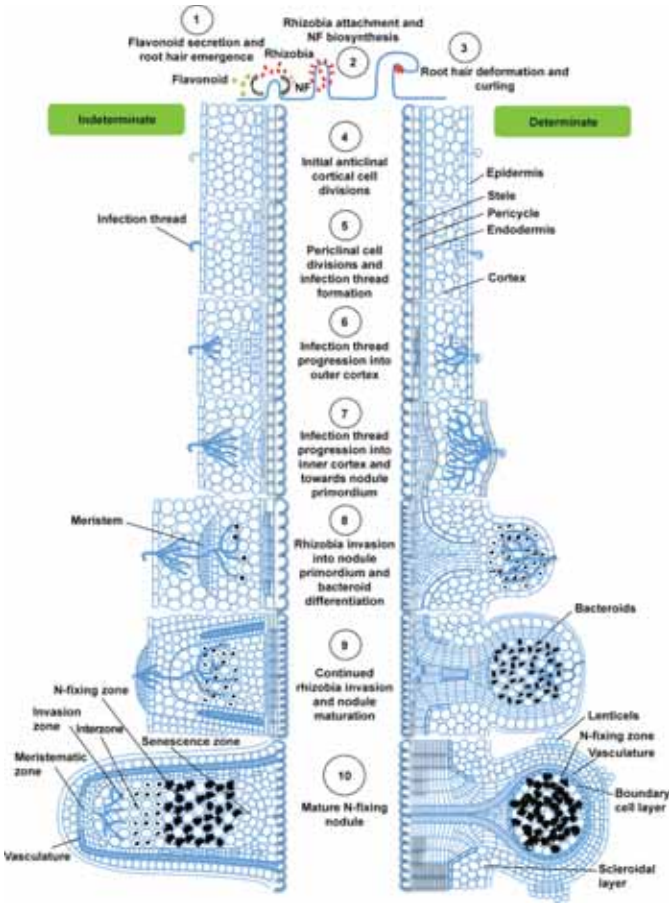


Figure 1.2. Development of determinate and indeterminate nodules.
The left side shows indeterminate (e.g., *Medicago*) nodule formation from root cortical cells. The mature nodule demonstrates distinct zonation (meristem, infection, interzone, nitrogen-fixation, and senescence zones). The right side follows indeterminate nodule formation (e.g., *Lotus* and Soybean) initiated from root pericycle cells. (Taken from Ferguson et al., 2010).

Upon reaching the nodule primordium, the bacteria gain entry via endocytosis, becoming surrounded by the symbiosome (or peribacteroid) membrane. Upon successful entry, the bacteria then differentiate, becoming nitrogen-fixing bacteroids. There are two modes of subsequent nodule cell differentiation. In plants such as pea and *Medicago*, the nodule meristem is persistent, leading to zones of differentiation and the formation of indeterminate nodules (Figure 1.2). Additionally, the bacteroids become terminally differentiated, a process mediated by cysteine-rich peptides (Van de Velde et al., 2010). In contrast, the nodule cells of soybean and *Lotus* divide and differentiate synchronously, forming determinate nodules (Ferguson et al., 2010). The bacteroids from determinate nodules can be re-isolated and remain viable (Mergaert et al., 2006).

1.4 Symbiosome Development

Entry of rhizobia into cells from the infection thread requires an invagination of the plant plasma membrane, which surrounds the bacteria forming an unwalled droplet (Verma and Hong, 1996). This new droplet membrane is called the symbiosome (or peribacteroid, PBM) membrane (SM) and the interior named the symbiosome space. The formation of the symbiosome superficially resembles endocytosis, however a recent study demonstrated that uptake of rhizobia actually depends on the exocytotic pathway (Ivanov et al., 2012). The authors demonstrated that proteins from the VAMP721 family were localized to sites of droplet release and that VAMP721d/e were necessary for SM formation.

A few studies have shed light on the identity of the SM. Cheon et al. (1993) originally identified Rab7, a small endosome-associated GTP-binding protein, as being import for symbiosome formation. Downregulation of *Rab7* in soybean impaired symbiosome development, and led to the accumulation of late endosomes. Recently, Rab7 was shown to populate endosomes, the tonoplast, and the SM in the *Medicago* nitrogen fixation zone through to the senescence zone (Limpens et al., 2009). Further, downregulation of *Rab7* induced early senescence of symbiosomes. The same authors also demonstrated that SYP22 and VTI11 (vacuolar SNAREs) are located on the SM of senescencing symbiosomes. Lastly, the plasma membrane marker SYP132 was located on the SM of symbiosomes at all stages of development. Therefore, it would appear that symbiosomes are marked as late endosomes until senescence, where they acquire a vacuolar identity.

1.5 Symbiosome Membrane Proteins

Proteomics has proved to be a useful tool for studying the SM. Since symbiosomes are very abundant in nodules and relatively dense, they can be purified in a two step process (Day et al., 1989). In infected cells, the amount of SM has been estimated to be 20-40 fold greater than the plasma membrane (Verma and Hong, 1996). The SMs of soybean, *Medicago*, pea, and *Lotus* have thus far been analyzed (Panter et al., 2000; Saalbach et al., 2002; Wienkoop and Saalbach, 2003; Catalano et al., 2004). These studies found hundreds of proteins such as chaperones, metabolic enzymes, receptor kinases, and transporters on the SM. The SM transporters are responsible for the delivery of micronutrients, metabolites, and compounds to and from the bacteroids. To date, a number of transporters have been kinetically characterized on the SM. These include transporters of protons (Blumwald et al., 1985), dicarboxylates (Udvardi et al., 1988), iron (Kaiser et al., 2003), zinc (Moreau et al., 2002), sulfate (Krusell et al., 2005), and nitrate (Vincill et al., 2005). The most abundant protein on the SM is an aquaporin (Nod26), first discovered over two decades ago (Fortin et al., 1987; Miao et al., 1992). This interesting protein has been since been characterized by numerous groups. It is now believed that Nod26 may be responsible for moving NH_3 across the SM, based on its transport capabilities (Weaver et al., 1994; Dean et al., 1999; Guenther et al., 2003) and its interaction with cytosolic glutamine synthetase (Masalkar et al., 2010).

1.6 Nitrogen Assimilation and Transport out of Nodules

After passage of ammonia across the SM, it is incorporated into glutamate by glutamine synthetase (Morey et al., 2002; Masalkar et al., 2010). In temperate legumes such as pea and *Medicago*, the nitrogen is primarily exported from the nodule as asparagine and glutamine (Atkins and Smith, 2007). In tropical legumes such as soybean and common bean, glutamine is converted into ureides for export. In infected cells of soybean nodules, glutamine is transported to mitochondria and plastids for *de novo* purine biosynthesis (Smith and Atkins, 2002). The purines are then degraded to xanthine, which diffuses to nearby uninfected cells for conversion to uric acid in the cytosol, then allantoin in peroxisomes (Hanks et al., 1981; Van den Bosch and Newcom, 1986; Datta et al., 1991). Allantoin is transported into the endoplasmic reticulum, where it is converted to allantoic

acid (Werner et al., 2008). Both allantoin and allantoic acid (collectively called ureides) diffuse out of the infected region via either a symplastic or apoplastic route to vascular bundles in the nodule inner cortex for export to aerial organs (Walsh et al., 1989; Brown et al., 1995).

2 Assessing the *in planta* function of GmSAT1

2.1 Introduction

2.1.1 GmSAT1 activity in yeast

GmSAT1 (*Glycine max* symbiotic ammonium transporter 1) was originally identified in a yeast complementation screen (Kaiser et al., 1998). GmSAT1 was able to restore growth to an ammonium transport mutant strain (26972c, *mep1-1 mep2-1*). In 26972c, two of the yeast high affinity transporters (Mep1 and Mep2) are inactive, while the third (Mep3) is inhibited by the mutant Mep1-1 (Marini et al., 2000). GmSAT1 was found to promote growth on low ammonium (1 mM) in 26972c, as well as promote uptake of toxic ammonium analog, methylammonium (MA). GmSAT1 contains a predicted C-terminal transmembrane domain and was found associated with the plasma membrane in yeast, as well as the peribacteroid membrane (PBM) in soybean nodules (Kaiser et al., 1998). The transcript for *GmSAT1* was also enhanced in the nodule by northern blotting. Collectively, the data suggested that GmSAT1 was an unusual ammonium transporter.

Work by Marini and colleagues (2000) demonstrated that GmSAT1 was unable to complement a yeast strain (31019b) with deletions of the three high affinity ammonium transporters (Δmep1/Δmep2/Δmep3). Therefore the authors questioned the original annotation as an ammonium transporter. In addition to a predicted transmembrane domain, GmSAT1 also contains a basic helix-loop-helix (bHLH) DNA-binding domain. It was suggested that GmSAT1 might be complementing 26972c by either relieving the inhibition of Mep3 or influencing gene transcription.

Recent work by Loughlin (2007) has demonstrated that GmSAT1 is likely a bHLH transcription factor. GmSAT1 was found localized to the nucleus in yeast by fusion to GFP and the nucleus of nodule cells by immunogold labeling. Further, microarray analysis from 26972c overexpressing *GmSAT1* found that a number of genes were upregulated, including *YOR378W*, a predicted major facilitator transporter. Subsequent experiments

have demonstrated that overexpression of YOR378W induces the uptake of MA (D. Mazurkiewicz, unpublished results). Mutating conserved residues in the DNA-binding domain of GmSAT1 abolished its ability to induce the expression of *YOR378W*. The same microarray also found a slight upregulation of *Mep3*, potentially explaining the ammonium complementation phenotype induced by GmSAT1.

2.1.2 The role of GmSAT1 *in planta*

As mentioned, GmSAT1 antiserum cross-reacted with a protein in the PBM fraction from soybean nodules (Kaiser et al., 1998), which was further supported by immunogold labeling of the PBM with the same serum (Loughlin, 2007). Immunogold labeling also revealed a signal in endoplasmic reticulum and or golgi vesicles and the nucleus. Therefore, GmSAT1 could be behaving in a similar manner as in yeast, being associated both with membranes and the nucleus and acting as a transcription factor. Loughlin (2007) used RNA interference (RNAi) to disrupt the expression of *GmSAT1*. It was found that downregulating *GmSAT1* led to a Fix- phenotype. When the plants were deprived of external nitrogen, in the presence of rhizobia, nodules eventually underwent premature senescence and the leaves became chlorotic. Inside the infected cells, there were disorders in symbiosome development and vacuole breakdown. Therefore, it was concluded that symbiosome biogenesis was disrupted and nitrogen fixation had not commenced. Taken together, these results demonstrate that GmSAT1 is likely a novel membrane-bound transcription factor required for symbiotic nitrogen fixation (Loughlin, 2007).

2.1.3 Membrane-bound transcription factors

GmSAT1 is a member a newly identified group of membrane-bound transcription factors (MTFs) that could comprise up to 10% of all transcription factors (Kim et al., 2007). These transcription factors are hypothesized to be inactive in the membrane-bound state and activated by proteolytic cleavage (Seo et al., 2008). This would allow the cell to rapidly respond to environmental signals without *de novo* protein synthesis. MTFs are activated by either regulated ubiquitin/proteosome-dependent processing (RUP) or regulated intramembreane proteolysis (RIP). In plants, members of the bZIP and NAC transcription factor families have been identified as MTFs. The bZIPs are all localized to

the ER, while the majority of NACs are associated with the PM (Seo et al., 2008). AtbZIP28 (Liu et al., 2007) and AtbZIP60 (Iwata and Koizumi, 2005) are involved in the ER stress response, as the processed forms are induced after treatment with tunicamycin. AtbZIP17 is processed in response to high salinity, activating genes involved in the salt stress response (Liu et al., 2007). It was also demonstrated that AtbZIP17 is released by RIP from the ER by a subtilisin-like serine protease (AtS1P). The NAC family members have been overexpressed in plants without their transmembrane domains, leading to diverse phenotypes. NTM1 regulates cell division (Kim et al., 2006), NTL6 pathogenesis (Kim et al., 2007), NTL8 flowering (Kim et al., 2007; Kim and Park, 2007), and NTL9 regulates leaf senescence (Yoon et al., 2008). Seo et al. (2010) recently demonstrated that NTL6 proteolytic processing is influenced by membrane fluidity. It was shown that NTL6 is cleaved by a metalloprotease via RIP in response to cold temperatures, then activating pathogenesis-related (PR) gene expression. As approximately 190 transcription factors in *Arabidopsis* are predicted to be membrane-anchored, this area of research is still in the early stages (Seo et al., 2008).

2.1.4 *GmSAT1* orthologues in *Arabidopsis*

An analysis of the *Arabidopsis* genome has identified 162 bHLH transcription factors (Bailey et al., 2003). There have been no reports demonstrating membrane association for any of these transcription factors. AT2G22750, AT2G22760, AT2G22770 (NAI1), and AT4G37850 belong to a subfamily of bHLH TFs that shows homology to GmSAT1. NAI1 is necessary for ER body formation in *Arabidopsis* (Matsushima et al., 2004). ER bodies are found in seedlings, roots, and wounded leaves (induced by JA) and are unique to Brassicales plants (Matsushima et al., 2002; Hayashi et al., 2010). In the *nai1* mutant, ER bodies do not form in seedlings or roots, but accumulate abnormal ER bodies after wounding. Through transcriptomic analysis, *PYK10* and *PBP1* were identified as targets of NAI1. PYK10 is a ß-glucosidase found in ER bodies and contains an ER retention signal (Matsushima et al., 2004). PBP1 is a cytosolic protein that binds to PYK10 and influences its catalytic activity (Nagano et al., 2005). It has now been determined that inducible ER bodies contain a different set of ß-glucosidases (such as BGLU18) and are distinct from constitutive ER bodies (Ogasawara et al., 2009). To date, it is believed that ER bodies are involved in the plant defense response, but many questions still remain

unanswered (Yamada et al., 2011). The other three *Arabidopsis SAT*-like genes have yet to be assigned a function. Of note, a recent paper demonstrated that, unlike GmSAT1, NAI1-GFP might be located exclusively in the nucleus (Tominaga-Wada et al., 2011), although only one image of root tip cells was presented. Since GmSAT1 is membrane-bound, and legumes do not contain ER bodies, it may have a function that is unique relative to the *Arabidopsis* proteins.

2.2 Results

2.2.1 GmSAT1 binds DNA *in vitro*

As *GmSAT1* induces dramatic changes in gene expression when overexpressed in yeast, experiments were carried out to test its DNA-binding ability. Initial attempts to express recombinant GmSAT1ΔTMD (residues 1-266) in *E. coli* (strains BL21-CodonPlus-RIL and –RP, Stratagene) with a 6 x His tag (plasmid pTrcHis B, Invitrogen) resulted in accumulation in the insoluble fraction. A significant amount of protein could be purified when inclusion bodies were extracted with 6M urea, however the fusion protein precipitated upon dialysis. To circumvent the solubility issue, the GmSAT1 secondary sequence was analyzed to locate potentially troublesome stretches of amino acids using the program psipred (http://bioinf.cs.ucl.ac.uk/psipred/). GmSAT1 is predicted to contain mostly random coil secondary structure N-terminal to the bHLH domain (Figure 2.1). Therefore, GmSAT1 was further truncated to residues 128-270 ($GmSAT1_{128\text{-}270}$) and fused to the maltose-binding protein (MBP) to aid solubility and purification using the plasmid pMAL-c5X (NEB). Expression of the construct in *E coli* (NEB Express strain) produced a large protein band in the soluble fraction. MBP- $GmSAT1_{128\text{-}270}$ was then purified using amylose resin (Figure 2.2). To serve as a control, MBP alone was also purified in a similar manner.

Having purified MBP-$GmSAT1_{128\text{-}270}$, experiments to test DNA binding activity could proceed. A microarray conducted on yeast expressing *GmSAT1* identified a 60-fold induction of an unknown transporter called *YOR378W* as well as a 38-fold induction of the high affinity phosphate transporter, *PHO84* (Mazukiwicz and Loughlin, unpublished results). Inspection of the upstream promoter region (1564 bp) of *YOR378W* found 8 E-

boxes (CANNTG) known to be binding targets of bHLH transcription factors (Figure 2.3A). The promoter region of *PHO84* (1584 bp) contains 11 E-boxes and 3 G-boxes (CACGTG), high affinity bHLH binding sites (Figure 2.3B). A 284 bp portion of the *YOR378W* promoter sequence (called Y3) containing three E-boxes was cloned and labeled with ^{32}P-dCTP. Similarly, a 594 bp portion of the *PHO84* promoter (called P3) containing three E-boxes and two G-boxes was cloned and labeled. These sequences were chosen because they were directly upstream of the start codon, contained a relative enrichment of bHLH binding sites, and were not too large in size for the mobility shift assay. When either Y3 or P3 was incubated with MBP-GmSAT1 a noticeable shift in the migration of Y3 and P3 was evident by native PAGE (Figure 2.4). In contrast, no such shift occurred when Y3 or P3 was incubated with MBP alone. This result demonstrates that GmSAT1 is able to interact with DNA *in vitro*.

2.2.2 Identification of *GmSAT2*

Using the resources available (NCBI, Plant Genome Database) at the beginning of my PhD, I was able to identify a gene homologous to *GmSAT1* in soybean, termed *GmSAT2*. Subsequently, the soybean genomic sequence was published in 2010 and reaffirmed these observations. As the soybean genome is known to have undergone a recent duplication event (Schmutz et al., 2010), *GmSAT2* (Glyma13g32650, located on chromosome 13) and *GmSAT1* (Glyma15g06680, located on chromosome 15) likely resulted from this duplication. This is enforced by the high degree of synteny between the two genomic regions (Figure 2.5). Since GmSAT2 is 93.4% identical to GmSAT1 at the amino acid level (Figure 2.6A), it may have retained a similar function.

2.2.3 Assessing the Function of GmSAT2

GmSAT1 expression in the yeast strain 26972c (an ammonium transport mutant) induces growth on low ammonium (1mM), as well as the uptake of toxic methylammonium (0.1 M MA). Previous experiments have shown that the two phenotypes are possibly independent (Loughlin, 2007). Therefore, this system was chosen to determine if GmSAT2 is able to mimic the function of GmSAT1. *GmSAT2* was therefore cloned by nested PCR (because of the high similarity to *GmSAT1*) and inserted into the yeast expression vector pYES3-

DEST (Figure 2.10B) and transformed into 26972c. Here, it was found that GmSAT2 was able to complement the ammonium phenotype, but not the MA phenotype (Figure 2.7).

As GmSAT1 and GmSAT2 are very similar, a strategy was developed to determine which residues were responsible for the loss of MA uptake. Both *GmSAT1* and *GmSAT2* contain sites for the restriction enzyme *Bam*HI, *Spe*I, and *Hin*dIII (Figure 2.6). Therefore, the two genes were digested from parent plasmids at these sites (along with a restriction site from the vector) and swapped. Swapping the N-terminal *Spe*I fragment from *GmSAT2* to *GmSAT1* resulted in the loss of MA uptake induction by GmSAT1, but retention of growth on low ammonium (Figure 2.7). The *Bam*HI site is located 21 amino acids from the N-terminus, where the two proteins differ by just two residues. Exchanging the *Bam*HI-*GmSAT2* domain into *GmSAT1* lead to a loss of MA uptake, but wild-type growth on low ammonium (Figure 2.8). Therefore, the combination of mutating a serine to threonine at position 3 (S3T) and leucine to methionine at position 17 (L17M) disrupted the ability of GmSAT1 to induce MA uptake. Mutagenesis was then carried out to substitute each of these residues individually, but GmSAT1 retained its wild-type function (data not shown). The third swap involved exchanging the *Hin*dIII fragment. Here, the two proteins only differ by five amino acids (Q245R, G260D, V263D, W272Y, and GmSAT2 contains an extra serine at position 256). Surprisingly, substitution of these amino acids also led to a loss of MA uptake induced by GmSAT1, but the retention of growth on low ammonium (Figure 2.8). Subsequently, three residues in GmSAT1 were chosen for mutagenesis (G260D, V263D, W272Y). Exchanging any of these amino acids individually did not lead to a loss of MA uptake (data not shown).

2.2.4 The N-terminus of GmSAT1 is important for activity in yeast

A N-terminal GFP fusion to GmSAT1 was initially constructed in an attempt to co-immunoprecipitate interacting proteins from yeast. To test for expression of the desired protein, a western blot was conducted on 26972 expressing *GFP:GmSAT1* (Figure 2.9). Upon probing the total cell extract with an anti-GFP antibody, three distinct bands were visualized. Full-length GFP:GmSAT1 is expected to be 66.4 kDa and GFP alone is 27.3 kDa. The largest band visualized is of the correct size for full-length GFP:GmSAT1, and the second largest band of 58 kDa likely represents GFP:GmSAT1 cleaved at the C-

terminal RXXL site (Loughlin, 2007). The smallest band recognized by the anti-GFP antibody is 32 kDa, approximately 5 kDa larger than free GFP. Therefore, it is possible that a cleavage event also occurred in the N-terminus of GmSAT1, yielding a 4-5 kDa fragment (35-45 amino acids) that remained attached to GFP (Figure 2.9B). Of note, these bands were not present in the empty vector fraction or GFP-alone samples.

Having observed a potential cleavage event in the N-terminus of GmSAT1, various truncations were made to assess their effect on the function of the protein. Initially, 4 N-terminal truncations were made: T1 (24 AA), T2 (51 AA), T3 (75 AA), and T4 (100 AA) as shown in Figure 2.10. T2, T3, T4 all resulted in a loss of growth on low ammonium and uptake of MA (T3 and T4 not shown, T1 and T2 shown in Figure 2.11). Interestingly, the T1 truncation yielded an intermediate phenotype, where the yeast grew on low ammonium, but did not uptake MA. Further truncations were made between the N-terminus of GmSAT1 and GmSAT1-T1. GmSAT1-N1 (5 AA), -N2 (12 AA), -N3 (17 AA) were constructed by PCR (Figure 2.10). All truncated proteins were able to induce growth on low ammonium (Figure 2.12). The toxic MA phenotype was restored by GmSAT1-N1, but not by GmSAT1-N2 or –N3. GFP was then fused to N1 and N2 in an effort to generate an intact, functional tagged protein for purification purposes. It was found that adding GFP to both proteins resulted in a partial MA sensitivity phenotype (Figure 2.13). Finally, the GFP-GmSAT1-N2 construct showed growth on low ammonium, while GFP-GmSAT1-N1 did not. Due to the inconsistencies, co-immunoprecipitation experiments were not carried out.

2.2.5 Alternative Transcript Abundance from the *GmSAT1* and *GmSAT2* loci

While performing BLAST searches against soybean ESTs to distinguish *GmSAT1* and *GmSAT2*, a novel set of ESTs were identified that mapped further upstream relative to the *GmSAT1* mRNA sequence deposited in GenBank (Figure 2.14). These short ESTs originated from 454 high-throughput sequencing of mRNA from soybean seeds containing globular stage embryos. Having identified a potentially new transcript, experiments were initiated to determine if the *GmSAT1* loci expressed alternative splice variants. Initially, a forward primer designed to bind to the 5' portion of the *GmSAT1* 5'UTR was used in conjunction with a reverse primer that bound to the 3' end of the *GmSAT1* coding

sequence. Upon separating this PCR product, it was noted that two bands were present (Figure 2.15A). Subsequently, the bands were excised and sequencing and found to represent two alternative 5'UTR sequences, one with a long 5'UTR (termed "*GmSAT1-intron*" and identical to the GenBank sequence of *GmSAT1*, AF069738) and a shorter version (termed "*GmSAT1-splice*", since an intron had been removed by alternative splicing). Based on the ESTs from Figure 2.14, a primer was designed and a third transcript was cloned and sequenced (termed "*GmSAT1.2*"). After translating the transcripts, it was found that *GmSAT1-intron* and *GmSAT1.2* encoded for identical proteins, whereas *GmSAT1-splice* encoded an extra 23 amino acid, hydrophobic N-terminal extension (Figure 2.16B). The same cloning strategy was then conducted with the *GmSAT2* locus, and it was determined that it also produces three transcripts (*GmSAT2-intron*, *GmSAT2-splice*, and *GmSAT2.2*), similar to *GmSAT1* (Figure 2.16A).

qPCR primers were then designed to selectively amplify the three unique transcripts for both *GmSAT1* and *GmSAT2*. For *GmSAT1*, it was found that all three transcripts were present in soybean nodules and roots. *GmSAT1*-intron was the most abundant transcript in nodules, and the *GmSAT1*-spliced transcript was preferentially expressed in nodule tissue (Figure 2.17A). In contrast, all *GmSAT2* transcripts were present at a relatively lower level in nodules, with a noticeable enrichment of *GmSAT2*-intron transcript in lateral roots (Figure 2.17B). Therefore, there is overlap in expression for *GmSAT1* and *GmSAT2*, however *GmSAT1* is present at a higher level in soybean nodules.

2.2.6 Localization of *GmSAT1* and *GmSAT2* promoter activity

The promoters (2kb upstream sequence of start codon) of both *GmSAT1* and *GmSAT2* were cloned and inserted into a GUS/GFP promoter-reporter construct (pKGWFS7, Figure 2.18A) and transformed into *Agrobacterium rhizogenes* K599. Hairy roots were generated, nodulated with *Bradyrhizobium japonicum* USDA110, and stained for GUS activity. As shown in Figure 2.19, the expression of *GmSAT1* was associated with the root vasculature and young nodules. As the nodules developed further, *GmSAT1* expression was confined to nodule tissues, particularly the outer cortex and the infected region. Sectioned nodules showed strong GUS activity in the inner cortex surrounding infected cells (Figure 2.20). GUS expression was also detected in uninfected cells, some in a ray

pattern, and in infected cells. Similarly, the *GmSAT2* promoter was active in the root vasculature, as well as young developing nodules (Figure 2.21). In contrast to *GmSAT1*, there was residual expression in root tissue as nodules aged. Sectioned nodules revealed that *GmSAT2* expression was confined to the inner nodule cortex and uninfected cells, with no staining observed in infected cells.

2.2.7 Subcellular localization of GmSAT1 in onion cells

Previous immunolabeling experiments have demonstrated that GmSAT1 localizes to the symbiosome membrane, ER vesicles and or Golgi, and the nucleus. In order to complement these findings, various GFP fusions to GmSAT1 were generated and bombarded into onion epidermal cells. A N-terminal fusion, GFP:GmSAT1, was found to associated with small vesicles (Figure 2.22A). These vesicles were often found at the cell periphery and surrounding the nucleus. A C-terminal fusion, GmSAT1:GFP, also localized to peripheral small vesicles. However, the vesicles were of a slightly different morphology (elongated) and less abundant. Surprisingly, neither fusion localized to the cell nucleus. As studies in yeast (section 2.2.4) indicated that GmSAT1 contains a cleaved N-terminal peptide, GFP was fused to a truncated version of GmSAT1. GFP-T1-GmSAT1 (missing the first 23 amino acids) in contrast was found to enter the nucleus and could be visualized in vesicles (Figure 2.22D). A construct was also made to assess the function of the added hydrophobic sequence that is encoded by the GmSAT1-splice transcript. A fusion of the signal peptide of GmSAT1-splice N-terminal to GFP (S1:GFP) localized to a distinct population of larger vesicles that were in close association with the cell periphery, as well as similar vesicles to GFP:GmSAT1 (Figure 2.22F). A fusion of the signal peptide of Nodulin-25 to GFP (N25:GFP) was mainly cytoplasmic, similar to free GFP.

2.2.8 Microarray of *GmSAT1* RNAi in soybean

GmSAT1 was previously silenced in soybean hairy roots by RNA interference (RNAi). Reducing the *GmSAT1* transcript abundance led to a Fix- phenotype when the plants were grown in the absence of external nitrogen (Loughlin, 2007). To identify differentially regulated transcripts after RNAi of *GmSAT1*, a microarray analysis was conducted using

the Soybean 1.0 ST microarray. Nodules were picked from *GmSAT1* RNAi and empty vector control hairy roots at 24 days post-inoculation (Figure 2.23). In total, RNA was extracted from 19 RNAi and 8 vector control samples and analyzed by qPCR for *GmSAT1* expression. Four samples from the RNAi pool showed similar down-regulation levels of *GmSAT1* and were chosen for further analysis (Table 2.1). In addition, four control samples that showed similar basal expression levels were chosen.

The RNA samples were then sent to the Ramaciotti Centre at the University of New South Wales. The RNA was checked and quantified by a bioanalyzer and all samples were deemed acceptable (Figure 2.24). The Soybean 1.0 ST microarrays (Whole Transcript, Affymetrix) chosen for the experiment were designed based upon the recently released soybean genome sequence. The Soybean 1.0 ST array contains 12-14 unique 24 bp probes that target all exons of each predicted gene (66,659 in total). After labeling, hybridizing, and scanning, the array images were then aligned to positional cues and .CEL files were generated. Each array was then subjected to a quality control analysis by the Ramaciotti staff. All scanned arrays were within deemed satisfactory (by monitoring signal intensities, internal controls, etc.) as shown in Figure 2.25. The .CEL files were imported into a trial version of the Parktek Genomics Suite software using default settings. The library and annotation files were obtained from the gene networks in seed development website (http://seedgenenetwork.net/annotate#soybeanWT).

Using the four replicates of the RNAi and control, a list was generated containing those genes that showed significant downregulation (≥ 2-fold, p-value ≤ 0.05) in the *GmSAT1* RNAi (Table 2.2). There were 95 genes significantly downregulated with representation from such classes as circadian clock proteins, leucine-rich repeat kinases, calcium/calmodulin dependent kinases, and transporters. The circadian clock is represented by such genes as *GIGANTEA (GI,* Glyma20g30980 and Glyma09g07240*), PSEUDO-RESPONSE REGULATOR 5 (PRR5,* Glyma04g40640*), PSEUDO-RESPONSE REGULATOR 7 (PRR7,* Glyma12g07860*) COLD, CIRCADIAN RHYTHM AND RNA BINDING 2 (CCR2*, Glyma06g01470*)*. Significant transporters included *ZINC TRANSPORTER 1 (ZIP1*, Glyma13g10790*)*, ABC transporter (*ABCG37*, Glyma17g03860), carbohydrate transmembrane transporter (*MSS1,* Glyma20g28230), and a major facilitator orthologue to YOR378W from yeast (*GmMFS1.5,* Glyma09g33680).

Other interesting downregulated transcripts included *ULTRAPETELA* (Glyma06g45180), *Uricase* (Glyma20g17440), and *SNF1-RELATED PROTEIN KINASE 2.4 (SNFK2.4, Glyma08g13380)*. qPCR was then carried out on a selection of genes to confirm the microarray results. All genes checked were in agreement, showing at least a 2-fold reduction in transcript levels (Table 2.3). In addition, 90 genes were found to be significantly upregulated upon silencing of *GmSAT1* (Table 2.4). Included are PLASMA MEMBRANE INTRINSIC PROTEINs (*PIP1;4*, Glyma11g35030 and *PIP2B*, Glyma10g35520), TONOPLASTIC INTRINIC PROTEIN (Glyma03g34310), bHLH transcription factor (Glyma20g39220), bZIP transcription factor (Glyma18g51250), ETHYLENE RESPONSE FACTOR (*ERF-1*, Glyma17g15480), and O-METHYLTRANSFERASE1 (OMT1, Glyma20g31700 and Glyma20g31610).

2.2.9 Analysis of *B. japonicum* fixation genes in *GmSAT1* RNAi

Because the *GmSAT1* RNAi plants generally have smaller nodules and eventually become Fix-, it was surprising not to find significant differences in expression in known nodulin genes in the array results. In order to further probe the phenotype, qPCR primers were designed to amplify known *Bradyrhizobium japonicum* USDA110 gene transcripts related to nitrogen fixation. From the same pool of RNA used for the microarray, cDNA was generated using random decamer primers. Two genes from *B. japonium* (*SigA* and *GapA*) were used as controls. *SigA* (*bll7349*) is a sigma factor required for initiation of transcription and *GapA* (*bll1523*) is a glyceraldehyde 3-phosphate dehydrogenase. To probe the status of nitrogen fixation, two genes, *NifH* and *FixU* were chosen. *NifH* (*blr1769*, Dinitrogenase reductase protein) and *FixU* (*bsr1757*, Nitrogen fixation protein) are both specifically expressed in bacteroids (Pessi et al., 2007). If the anaerobic environment required for nitrogen fixation were disrupted, then these two genes would not be expressed (Sciotti et al., 2003). Surprisingly, both genes were found to be expressed at similar levels in both control and *GmSAT1* RNAi samples (Figure 2.26).

2.2.10 Expression of *GmSAT1* is increased at night

Having identified genes belonging to the circadian clock pathway, soybean plants were grown and sampled every four hours for a 24-hour period. As shown in Figure 2.27A, *GmSAT1* was found to have an increased expression level at night relative to daylight hours. Two known components of the plant circadian clock were also analyzed. *GI* (*Glyma09g07240*) showed increased expression at 6 pm, which continued until dawn. *LHY* (Glyma16g01980), on the other hand, was expressed at its highest level during daylight hours. *GmMFS1.5* transcript levels were more stable, but showed a decreased expression at 6 am relative to other times. Finally, the three transcripts from the *GmSAT1* locus were analyzed by qPCR. Similar to the total *GmSAT1* transcript levels, the individual variants showed a trend of highest expression at the 10 pm and 2 am time points (Figure 2.27E).

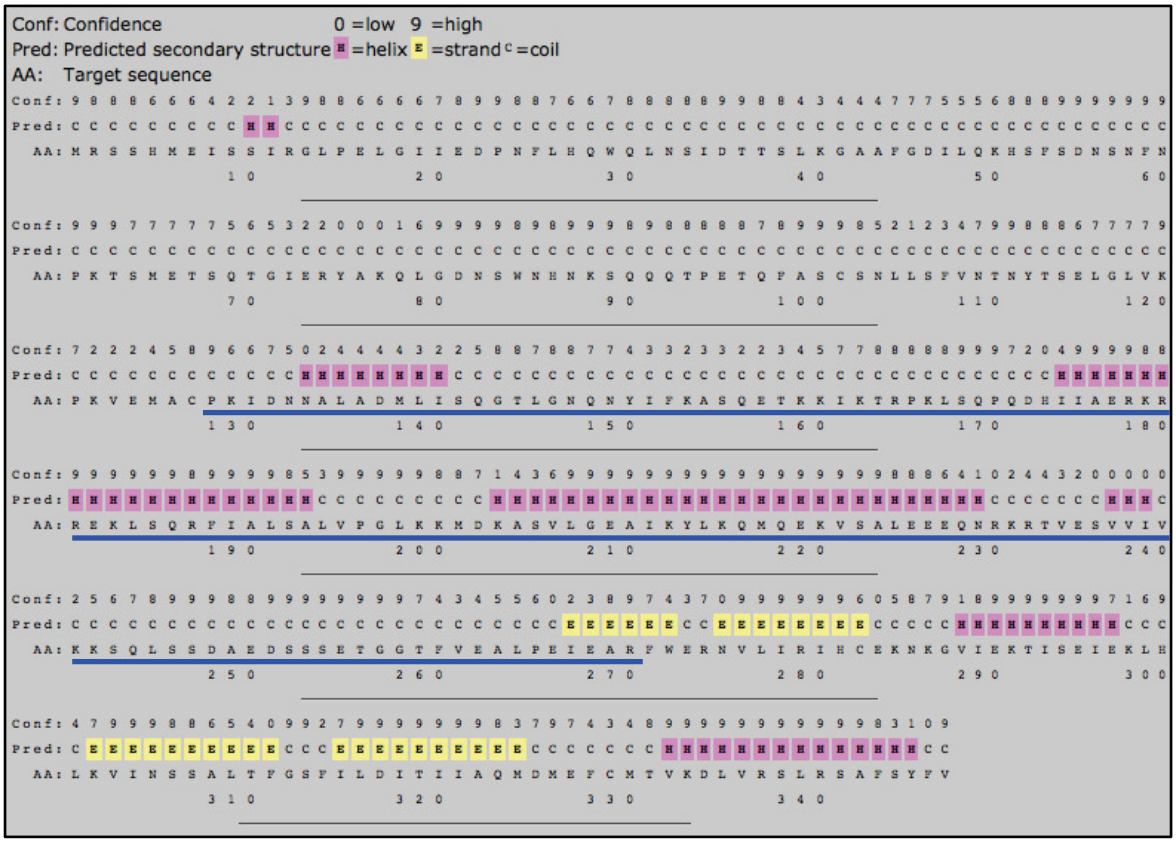


Figure 2.1. Secondary structure prediction of GmSAT1.
The GmSAT1 protein secondary structure was predicted using the PSI-Pred prediction program (http://bioinf.cs.ucl.ac.uk/psipred/). Helices are annotated in pink, strands in yellow, and coils are uncolored. GmSAT1 was truncated to residues 128-270 (underlined in blue), removing the N-terminal coiled region and C-terminal transmembrane domain (residues 303-322), for fusion to the maltose-binding protein. The bHLH domain spans from residues 170-218.

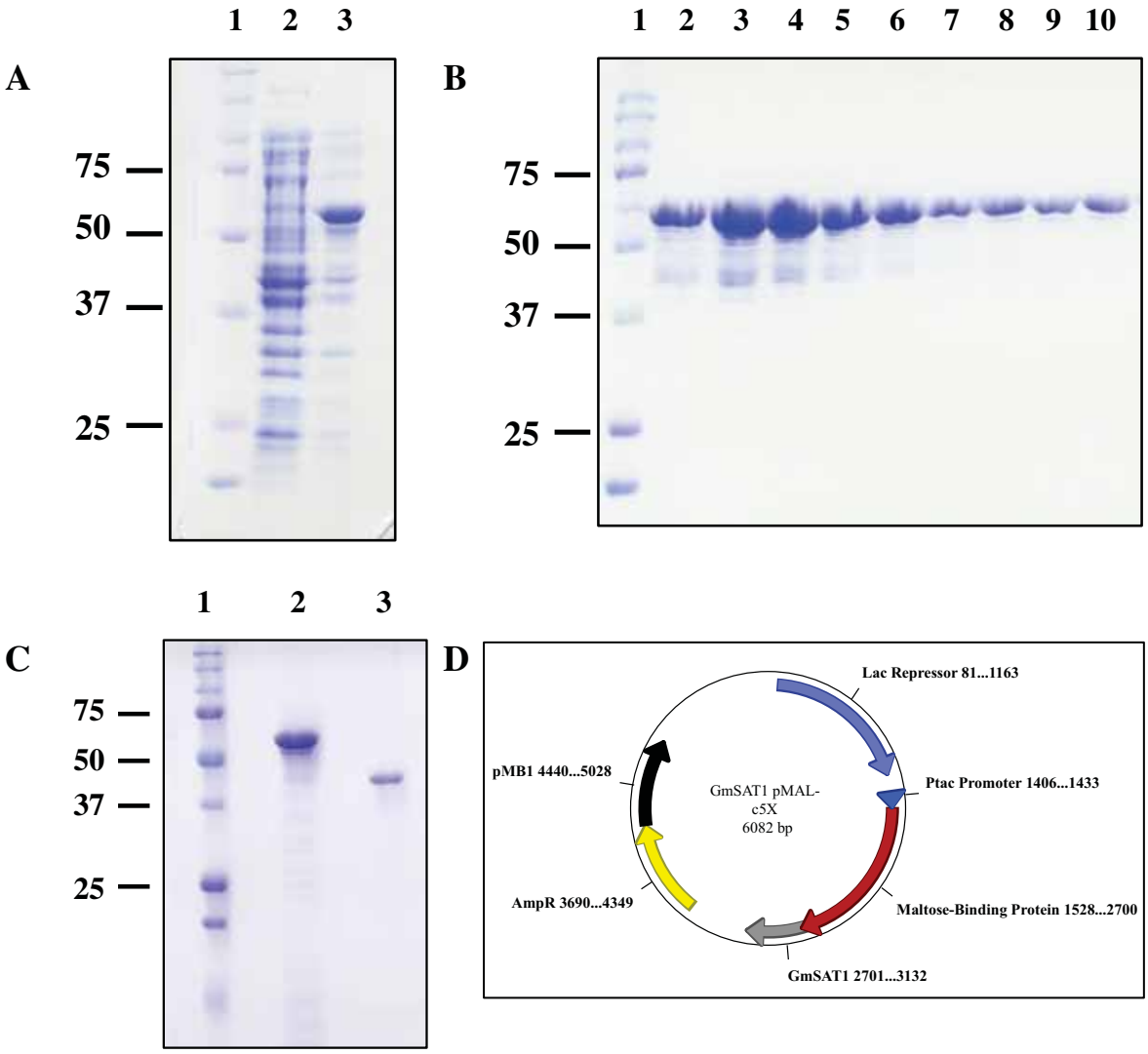


Figure 2.2. Expression and purification of MBP:GmSAT1 from *E. coli*.
A, SDS-PAGE gel showing MBP:GmSAT1 soluble expression in NEB Express before (lane 2) and after induction (lane 3) with IPTG. **B**, SDS-PAGE gel analysis of fractions 1-9 (lanes 2-10) obtained after elution from the amylose column. **C**, Purified proteins used for EMSA analysis. Lane 2, MBP:GmSAT1. Lane 3, maltose-binding protein. **D**, *GmSAT1*-pMAL-c5X plasmid used for protein expression. The plasmid contains a copy of the lac repressor (to reduce basal expression) and the Ptac promoter (for high-level expression upon IPTG induction). Protein ladder sizes for SDS-PAGE gels are indicated in kDa.

A

CGAAGTAAATAACGTGCAGAACATTGCATATGATAGATAACAACCTATAGTATATATTTCAAAATCGCGACAA
ATCTGAAATATTAATTGATTTGATTAGCTGCAGAGATCTTCTCGATGTAAGGAGCTATTTAAATAGAAAGGGT
CAGAATCGTCGTCCTGTGGGAGCTGGTATATACTAAATTGATGAACAACTAAAATAGCACTCTCATTGTCATT
TTGGGTGTGAGACCTTAAGCAAGGAGAGATATTTTTAATGGATGTATTAGCGCAGCCGTTGCTTACGGTATTC
CTTGGCATATTGTCCCTGAAAAAAAACCGTTGTACACTCGGATACGCATTTTTCGGACTTCTCCCACCCAAGA
AACGCAACTTTTGAAGGGCGCCTACACCACTAGACCAACAAGGCCAATGCAAAGCTACAAACAAATGCGTGGG
CGCATGGGCATAAGGAGGTGTAAAAAGCTTTTTTTTTTTATGTGAGGAATTACAGGCGGAAGATACTCAAATA
AAGCACTGATTTGATGTTACCGAACTTTTCCCCAGGCATAAAATTTTATTTTCATATTTCATCAATTCGATAT
TATTGATATCGCACCTAATGTAGTCATTTCCTTTCTCTCTTTGCTAGTAGATTTTAAAAAGGTTTCTTGCTGT
ATTTTTTTATACTTTTATAGACGCCGCTTTCCAAAGTTCATCTAAAATAAGTCGCTACAATGATTTGATCTTG
ATTCGGTGCTTCAGGAGCCTGCTTGCTTCTTTGCAGCTTCGGCAATTTGGAACCTTGTGTGCATATCCTCGTT
TGTAAGCCCAGAAGCACACAGACATCGATCACATCTATAATAGGAGCATGCTATCATAAAAGATTCACAGTAT
TCGTATAATTACATGGCAGACTTTTTGAAAAGAATGGTATGGAACCAAGTGAGCGCCTAGAAAAAAAGCTGAA
AAGCTGAAATCTTAATTGGATAAGAAACTTCAATAACATTTTTCACGCGTTTCTTTTTCAACTAGATTAGCCT
TAACAAACGACAATCATCGAGTACCCACGCTCATTACGCTGAACGTGCAAATGTAACGGGGTGGAAACCTAAA
ACGTTGAATGAACTAGCCAGAACGATGACAATTGCGGCAAGCTTTCCATACGTGTTCTGTTGTTGCGGCTTAG
AGTGGTAGGTAACCGCGCCAAGCAAAATTACCTACCACTTTAAGCCCAAAAAGGTGATGTGTAGCTATTGCGG
CTGTGGCGGCTATTGCGGTTGTCGCCACGCTTGTCGCCA**GGGTTGTCGCTACGAATGTTTCAGCTTTTTTTTC**
AATGGGCCACTCTTTTCACGTGAAGCGCTCCACGAAATGTGCGGATATTGCGGAAGCTCTCAATGATGAGCTA
TCGGTAGGGATTAAAATAAACGAGATTTGGCAGCAGTTTTAAATACCAAGGGAGTAGTATATCTAGCCTCCAA
TAACTTCATTTGTGGAAACTTACTTCTATCTCACTCTCTACCTGAAAAGAAAACAAATATTGTACATTCATCA
ATTGCTGGAACATAAAGAACATAAAATCAAT**ATG**

B

CCGAAGACGTTTCACTCAGGTTTTTCATCATGATTTGGAACAGCTCCTCTCTCTTATCCAACTGAACAAGATC
CATCTTATGCAGAAGAACAAAAATTTTGGCGTCGGGAGAGTACTTCCTTAATTGCTTCAAAGCTTTTGCAAAT
ATTTCAATATCCTTGAGAACTTCAGTTGACTCTACATCAAAAACGTGAATTAACACCTGCACCATCTGGAAAA
TGTGGTCTTTTTGCTTGGTGAAATAATTCTCCATAAACACGTCCTGCCCACCACAGTCCCACAGATTTAGAGT
CATATTCCCAAGAAATCTCAAATGGGAGTGCTCTACATCAATGGTGGCACCCAATCTCCTAGTGTCAAAAGCG
GAGTAGTTACTAAAGATGATCGACCTCATTGACGATTTACCGGAGCCGGACCGGCCCATCAGAAGCAGTTTCT
TCCTATTATTTGACGACATTACTAAATTGTCGATTGATAAACGTGATTTTGATGGCCTTCACTCCTGTGTAAG
ATACCTATTGTATTTATTCAAAAAGAAATTATTACCATAATCACAATTTAGGGTAGGGTCCCACATTGAACTT
TTCACTTCGTTTTTTTACCGTTTAGTAGACAGAATGCGAGAGTGATAAAGAAGAGGCGGTTAATCAATGAAAA
AAAAAAAAAAAAATTTAAAAAAGAAAAGAGAAAAGGAATAAAAAAGTGTCACGTGATAAAAATCACTACCCGG
AGATGACTTCAAACGACTCGGTATACTCTGCCTAATAAACCTTAATTTTCTTACAAAAAAAAAAGATTCAATA
AAAAAAGAAATGAGATCAAAAAAAAAAAAAATTAAAAAAAAAAAGAAACTAATTTATCAGCCGCTCGTTTATC
AACCGTTATTACCAAATTATGAATAAAAAAACCATATTATTATGAAAAGACACAACCGGAAGGGGAGATCACA
GACCTTGACCAAGAAAACATGCCAAGAAATGACAGCAATCAG**TATTACGCACGTTGGTGCTGTTATAGGCGCC**
CTATACGTGCAGCATTTGCTCGTAAGGGCCCTTTCAACTCATCTAGCGGCTATGAAGAAAATGTTGCCCGGCT
GAAAAACACCCGTTCCTCTCACTGCCGCACCGCCCGATGCCAATTTAATAGTTCCACGTGGACGTGTTATTTC
CAGCACGTGGGGCGGAAATTAGCGACGGCAATTGATTATGGTTCGCCGCAGTCCATCGAAATCAGTGAGATCG
GTGCAGTTATGCACCAAATGTCGTGTGAAAGGCTTTCCTTATCCCTCTTCTCCCGTTTTGCCTGCTTATTAGC
TAGATTAAAAACGTGCGTATTACTCATTAATTAACCGACCTCATCTATGAGCTAATTATTATTCCTTTTTGGC
AGCATGATGCAACCACATTGCACACCGGTAATGCCAACTTAGATCCACTTACTATTGTGGCTCGTATACGTAT
ATATATAAGCTCATCCTCATCTCTTGTATAAAGTAAAGTTCTAAGTTCACTTCTAAATTTTATCTTTCCTCAT
CTCGTAGATCACCAGGGCACACAACAAACAAAACTCCACGAATACAATCCAAATG

Figure 2.3. Overview of *YOR378W* and *PHO84* promoter sequences.
A, 1564 bp upstream sequence of *YOR378W*. **B**, 1584 bp upstream sequence of *PHO84*. E-boxes (blue, CANNTG), G-boxes (green, CACGTG), and start codons (red) are indicated. Bold and underlined sequences indicate the cloned fragments used for EMSA analysis (**A**, Y3 and **B**, P3).

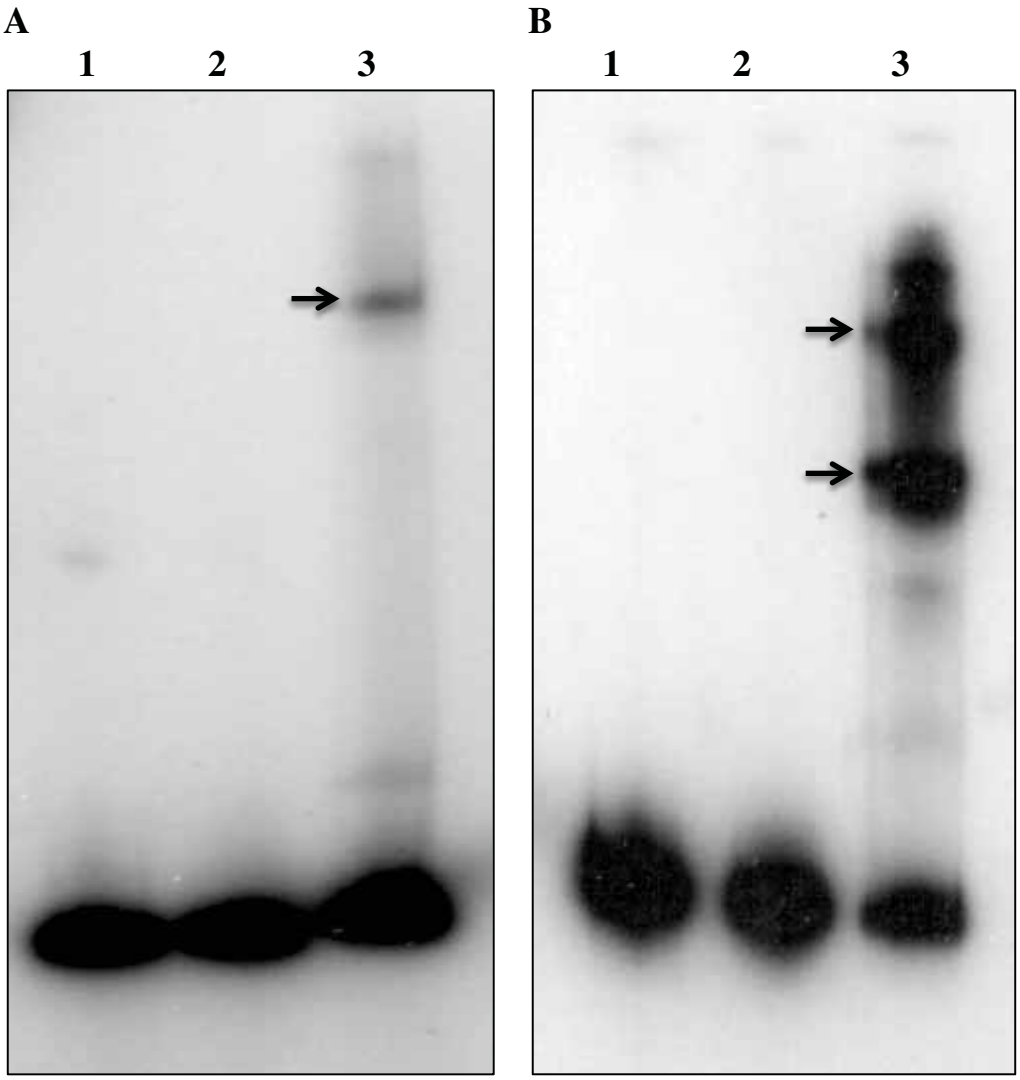
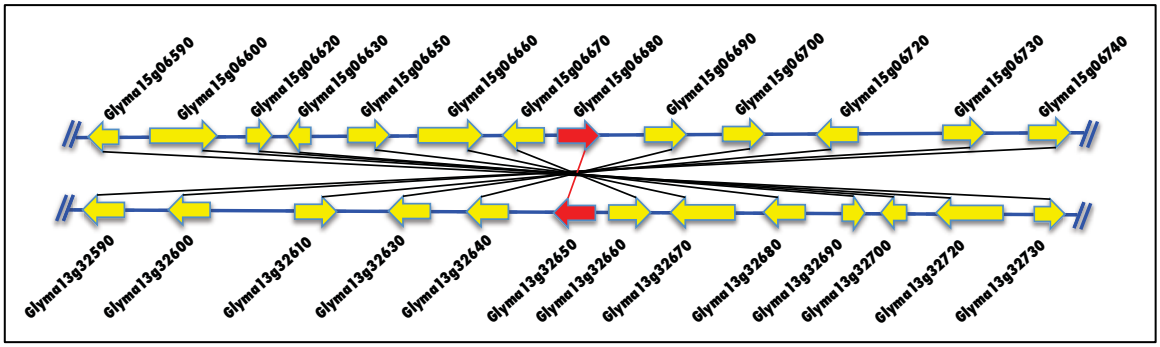


Figure 2.4. GmSAT1 binds to the *YOR378W* and *PHO84* promoters *in vitro*.
Electromobility shift assays were carried out with purified MBP:GmSAT1 and *YOR378W* (**A,** 288 bp), and *PHO84* (**B,** 594 bp) promoter fragments labeled with ^{32}P. Samples were incubated at room temperature for 25 mins, separated by 6% native-PAGE, dried, and exposed to film. Probe alone (lane 1), empty MBP + probe (lane 2), MBP:GmSAT1 + probe (lane 3). Arrows indicated shifts in mobility of the respective probes.

A

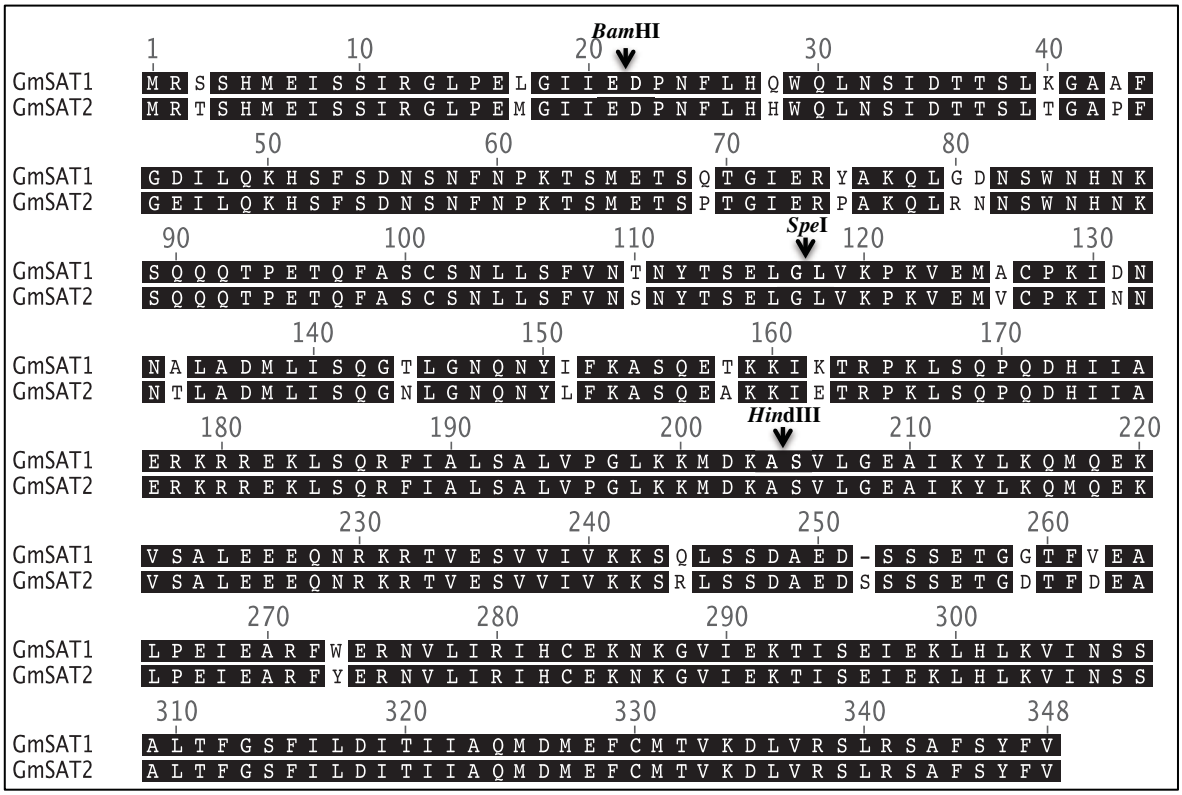


B

Chromosome 13	Chromosome 15	Annotation
Glyma13g32590	Glyma15g06740	Unknown protein
Glyma13g32600	Glyma15g06730	Squalene monooxygenase
Glyma13g32610	Glyma15g06720	Unknown protein
Glyma13g32630	Glyma15g06700	Unknown protein
Glyma13g32640	Glyma15g06690	ER stress-inducible protein
Glyma13g32650	Glyma15g06680	*GmSAT2*/*GmSAT1*
Glyma13g32660	Glyma15g06670	Inorganic pyrophosphatase
Glyma13g32670	Glyma15g06660	Major facilitator transporter
Glyma13g32680	Glyma15g06650	Methionine sulfoxide reductase B 2 (
Glyma13g32690	Glyma15g06630	LisH/CRA/RING-U-box domains-containing
Glyma13g32700	Glyma15g06620	Unknown protein
Glyma13g32720	Glyma15g06600	AtFAAH (fatty acid amide hydrolase)
Glyma13g32730	Glyma15g06590	Octicosapeptide/Phox/Bem1p family protein

Figure 2.5 Genomic synteny of the *GmSAT1* and *GmSAT2* loci.
A, Overview of a 50 kb syntenic region between soybean chromosome 13 (bottom) and 15 (top) surrounding the *GmSAT1* and *GmSAT2* loci (shown in red). The region is part of a large block (score 15818.0, *E*-value 0.0, with 397 anchors) identified by the plant genome duplication database (http://chibba.agtec.uga.edu/duplication/). **B**, List of syntenic loci between Chromosome 13 and Chromosome 15 depicted in (A).

A



B

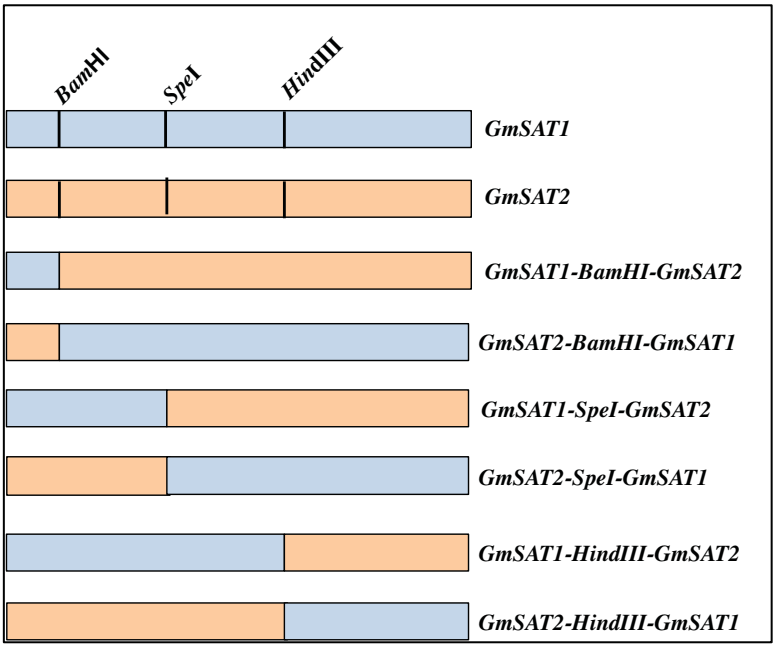


Figure 2.6. Overview of *GmSAT1* and *GmSAT2* and domain swaps.
A, Pairwise alignment of GmSAT1 and GmSAT2 proteins using Geneious software. The two proteins are 93.4% identical (325/348 residues). Indicated are the *Bam*HI, *Spe*I, and *Hin*dIII sites used for swapping domains of *GmSAT1* and *GmSAT2*. **B**, Overview of constructs generated by exchanging domains of *GmSAT1* and *GmSAT2* using the highlighted restriction digest sites.

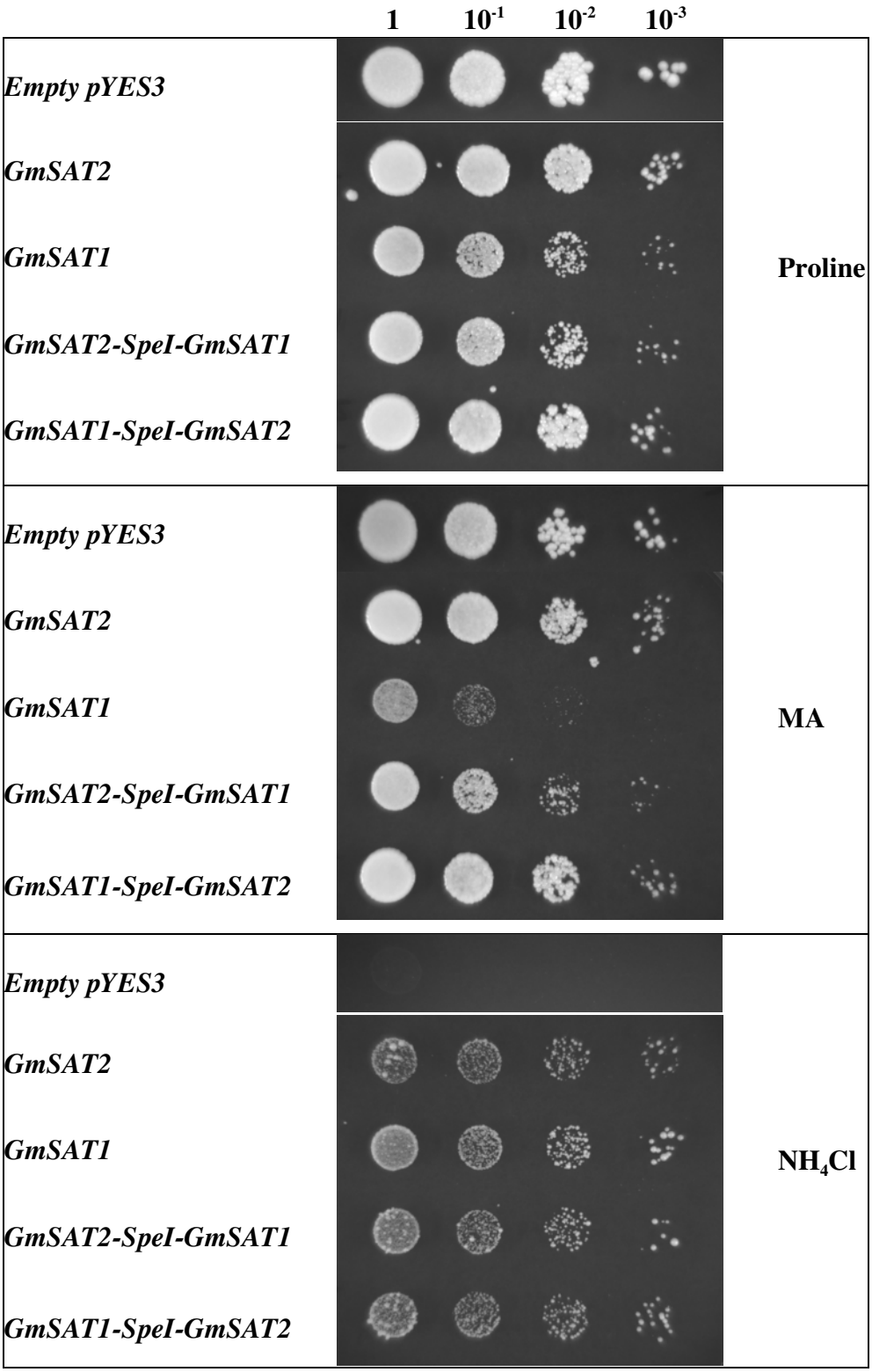


Figure 2.7. Growth of yeast strain 26972c expressing *GmSAT1* and *GmSAT2*.
Growth analysis of *GmSAT1* and *GmSAT2* along with *Spe*I domain swaps (see Figure 2.6) between *GmSAT1* and *GmSAT2*. All sequences were expressed from the plasmid pYES3-DEST. Yeast were grown, diluted to OD_{600} of 0.5, and spot inoculated onto Grenson's medium (with 2% (w/v) galactose) containing either 0.1% proline (w/v), 0.1% proline (w/v) + 100 mM MA, or 1 mM NH_4Cl. The leftmost spot contains an OD_{600} of 0.5, followed by three 10-fold serial diutions (indicated on top).

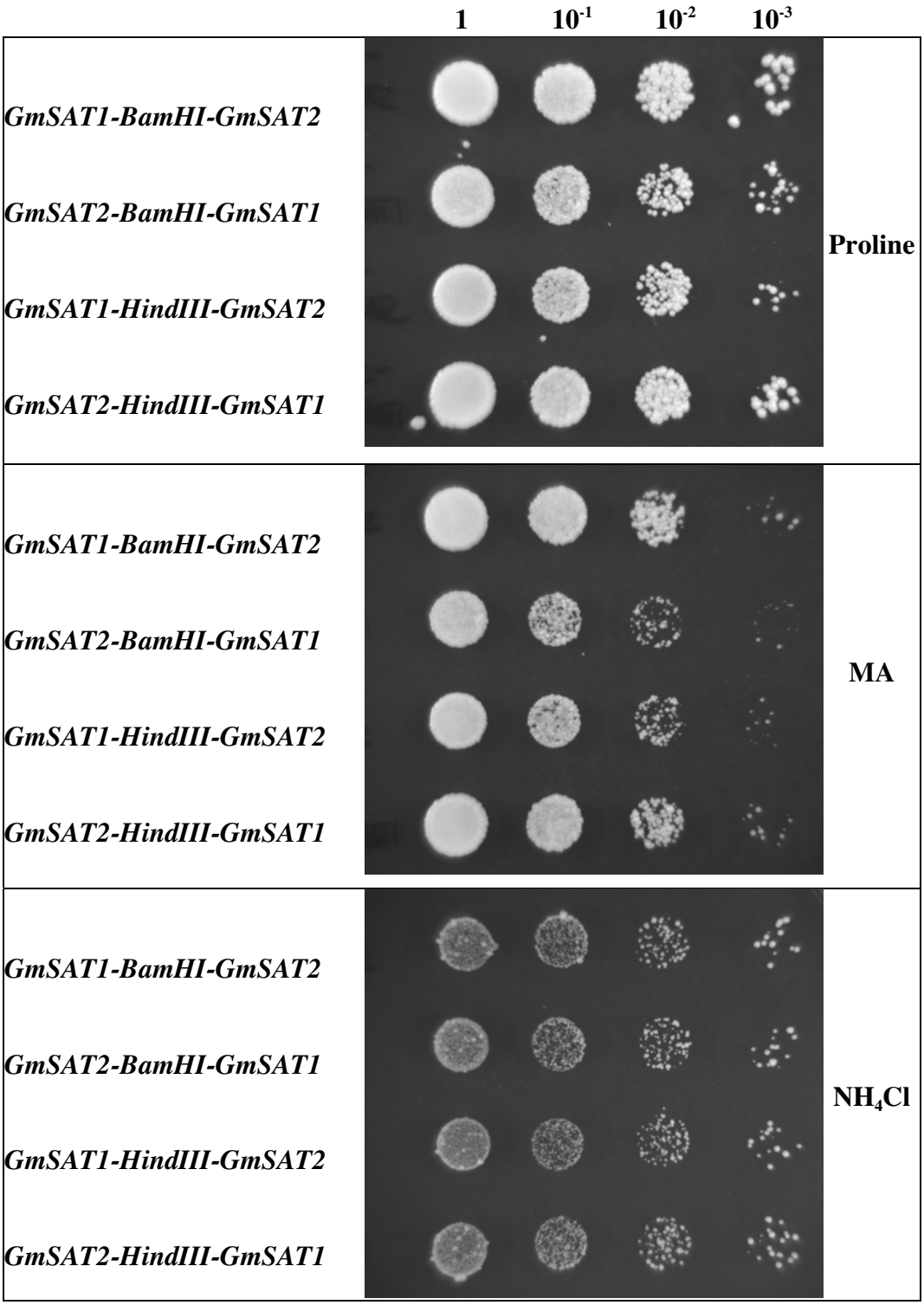


Figure 2.8. Growth of 26972c with *GmSAT1* and *GmSAT2* domain exchanges.
GmSAT1 and *GmSAT2* were digested with either *Bam*HI or *Hin*dIII and the domains swapped to generate hybrid proteins (see Figure 2.6). Yeast were grown, diluted to OD_{600} of 0.5, and spot inoculated onto Grenson's medium (with 2% (w/v) galactose) containing either 0.1% proline (w/v), 0.1% proline (w/v) + 100 mM MA, or 1 mM NH_4Cl. The leftmost spot contains an OD_{600} of 0.5, followed by three 10-fold serial diutions (indicated on top).

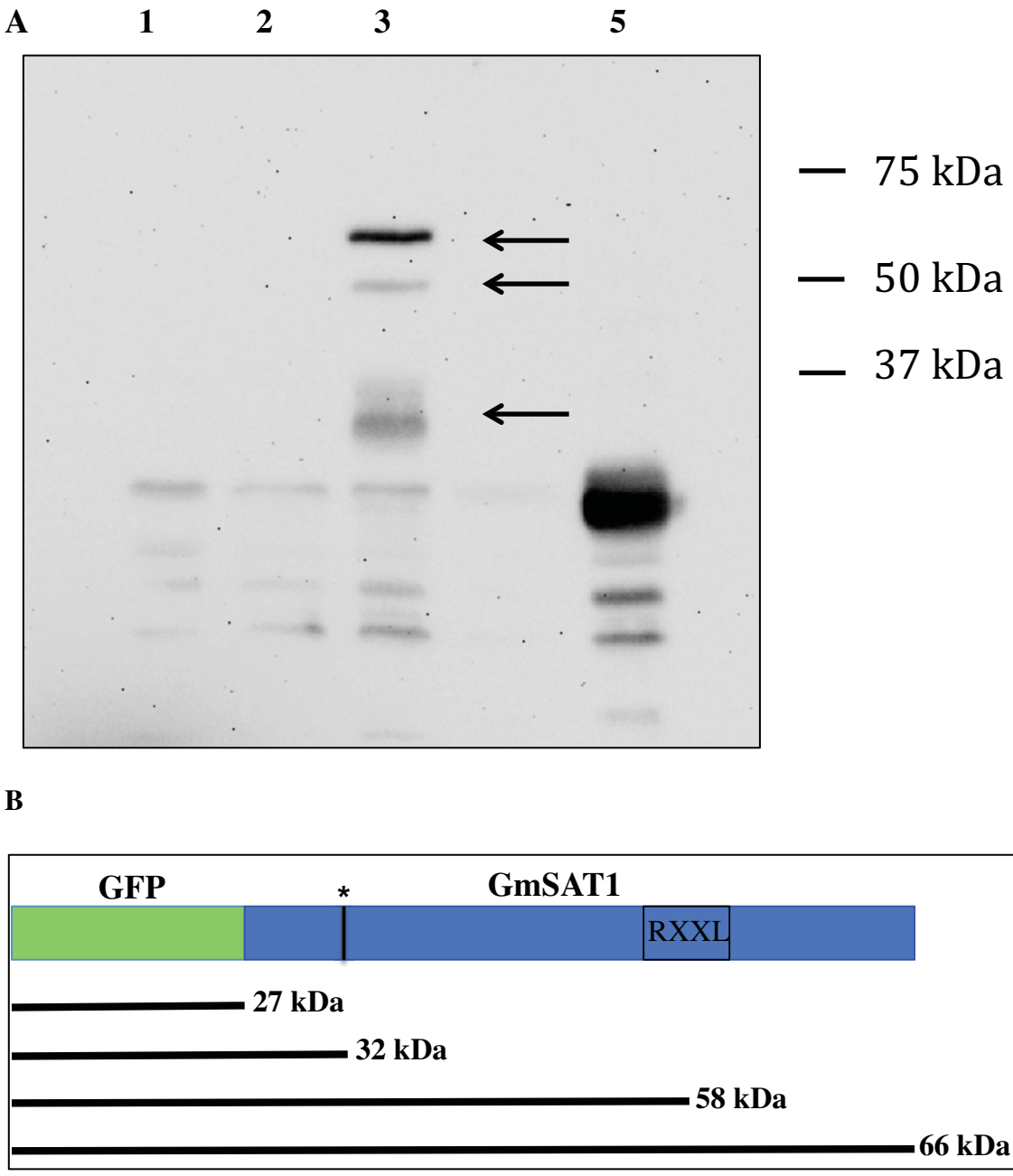
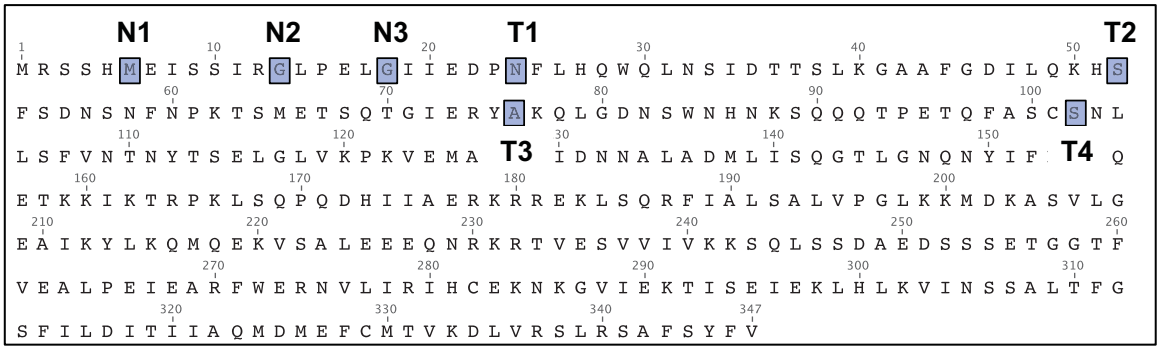


Figure 2.9. Processing of the GFP:GmSAT1 in yeast.
A, Western blot analysis of 26972c expressing GFP:GmSAT1, incubated with an anti-GFP antibody. Total protein was isolated from yeast carrying empty pYES3-DEST (lane 1), GmSAT1-pYES3 (lane 2), GFP:GmSAT1-pYES3 (lane 3), and GFP-pYES3 (lane 5). Each lane contains 20 µg of protein. MW marker sizes are indicated on the right. Arrows in lane 3 indicate the 3 unique bands (66, 58, 32 kDa) detected after expression of GFP:GmSAT1. B, Graphic summary of GFP:GmSAT1 fusion relative to protein bands sizes observed in (A). Indicated is a site-1 protease recognition sequence (RXXL) as well as a potential N-terminal cleavage site (*).

A



B

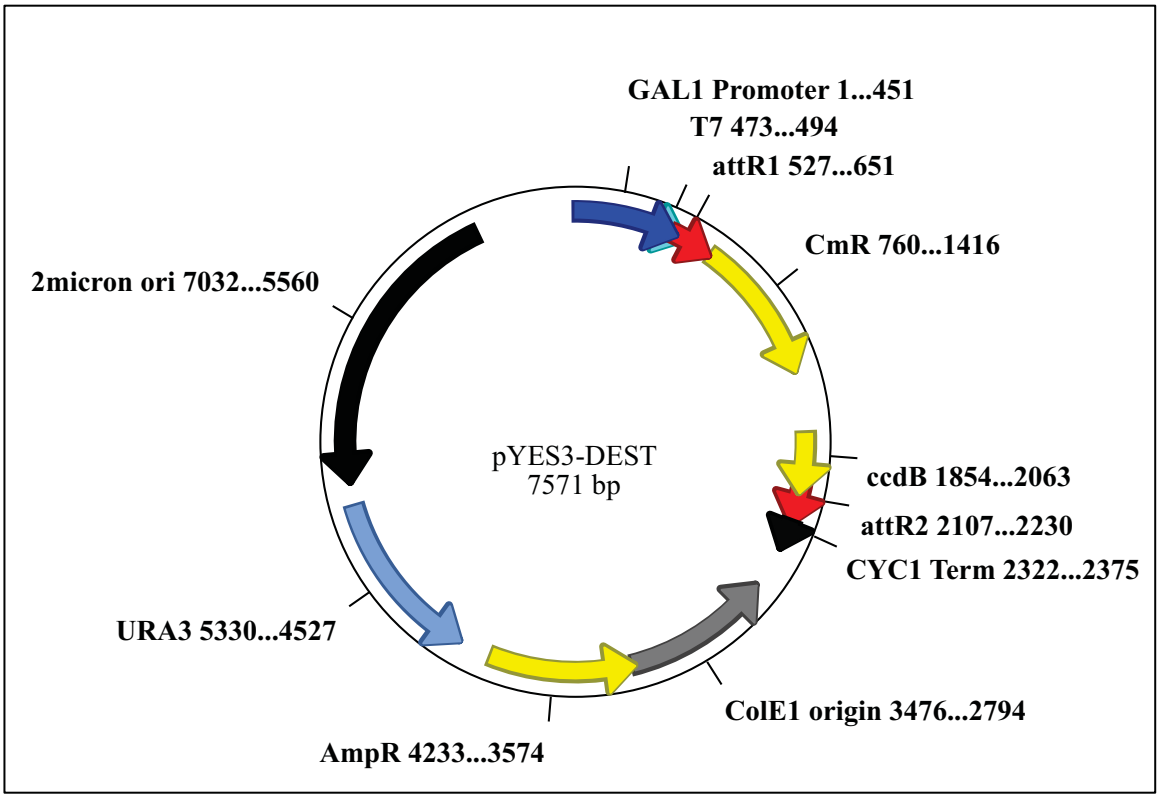


Figure 2.10. Overview of *GmSAT1* N-terminal deletions.
A, Primers were designed to remove portions of the *GmSAT1* N-terminus and insert a start codon at the indicated positions in blue (N1, N2, N3, T1, T2, T3, and T4). **B**, Map of pYES3-DEST plasmid used for yeast protein expression. pYES3-DEST contains a gateway cloning cassette driven by the inducible GAL1 promoter. The plasmid is selected in yeast by uracil auxotrophy.

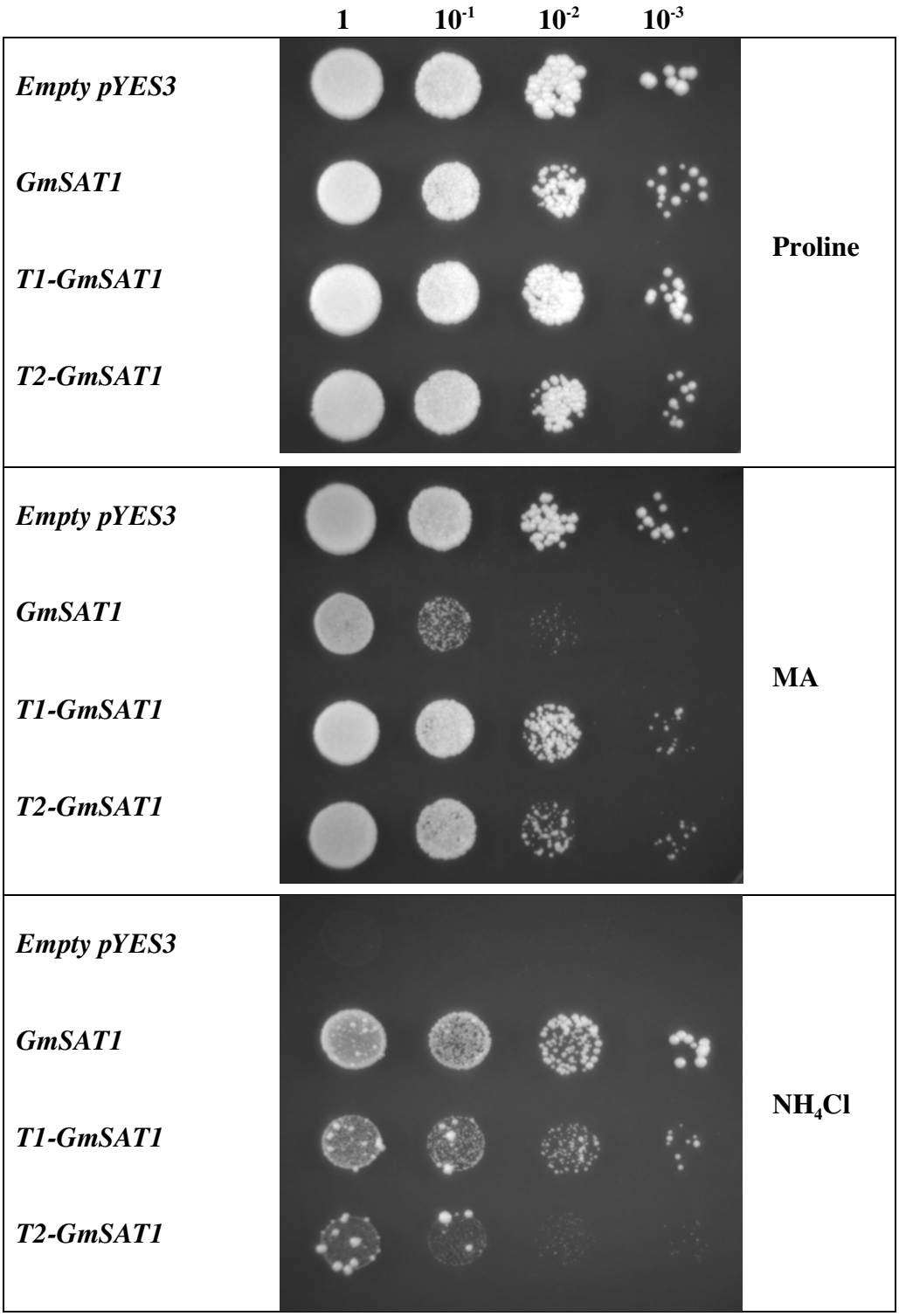


Figure 2.11. Growth of 26972c expressing truncated N-termini of GmSAT1.
The N-terminus of GmSAT1 was truncated by PCR as shown in Figure 2.10A. The T1 and T2 truncations remove 24 and 51 amino acids, respectively. Yeast were grown, diluted to OD_{600} of 0.5, and spot inoculated onto Grenson's medium (with 2% (w/v) galactose) containing either 0.1% proline (w/v), 0.1% proline (w/v) + 100 mM MA, or 1 mM NH_4Cl. The leftmost spot contains an OD_{600} of 0.5, followed by three 10-fold serial diutions (indicated on top).

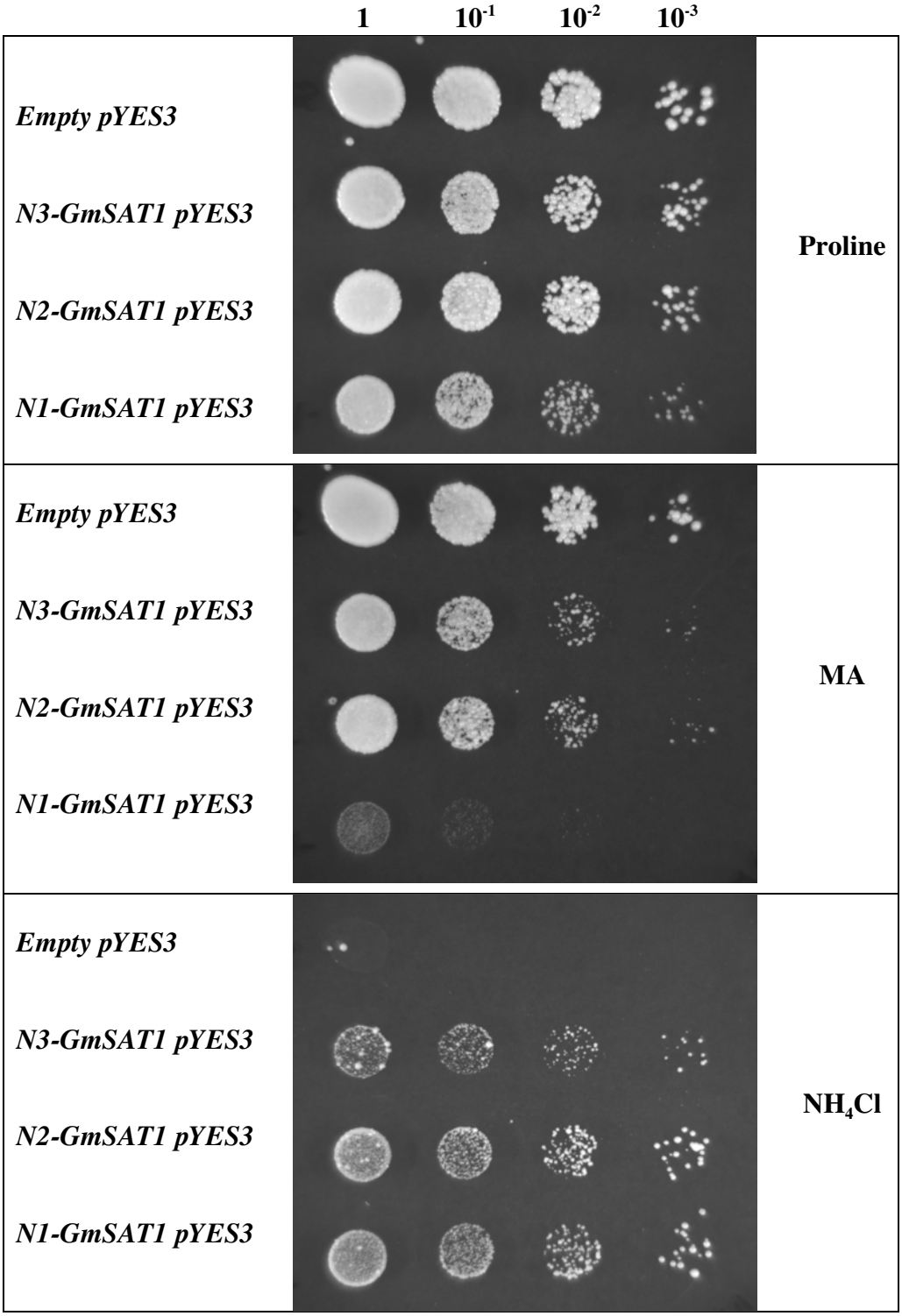


Figure 2.12. Growth of 26972c expressing further truncated N-termini of GmSAT1. Additional truncations were made to the N-terminus of GmSAT1 to test their effect on activity. The N1, N2, and N3 truncations removed 5, 12, and 17 amino acids, respectively (see Figure 2.10A). Yeast were grown, diluted to OD_{600} of 0.5, and spot inoculated onto Grenson's medium (with 2% (w/v) galactose) containing either 0.1% proline (w/v), 0.1% proline (w/v) + 100 mM MA, or 1 mM NH_4Cl. The leftmost spot contains an OD_{600} of 0.5, followed by three 10-fold serial diutions (indicated on top).

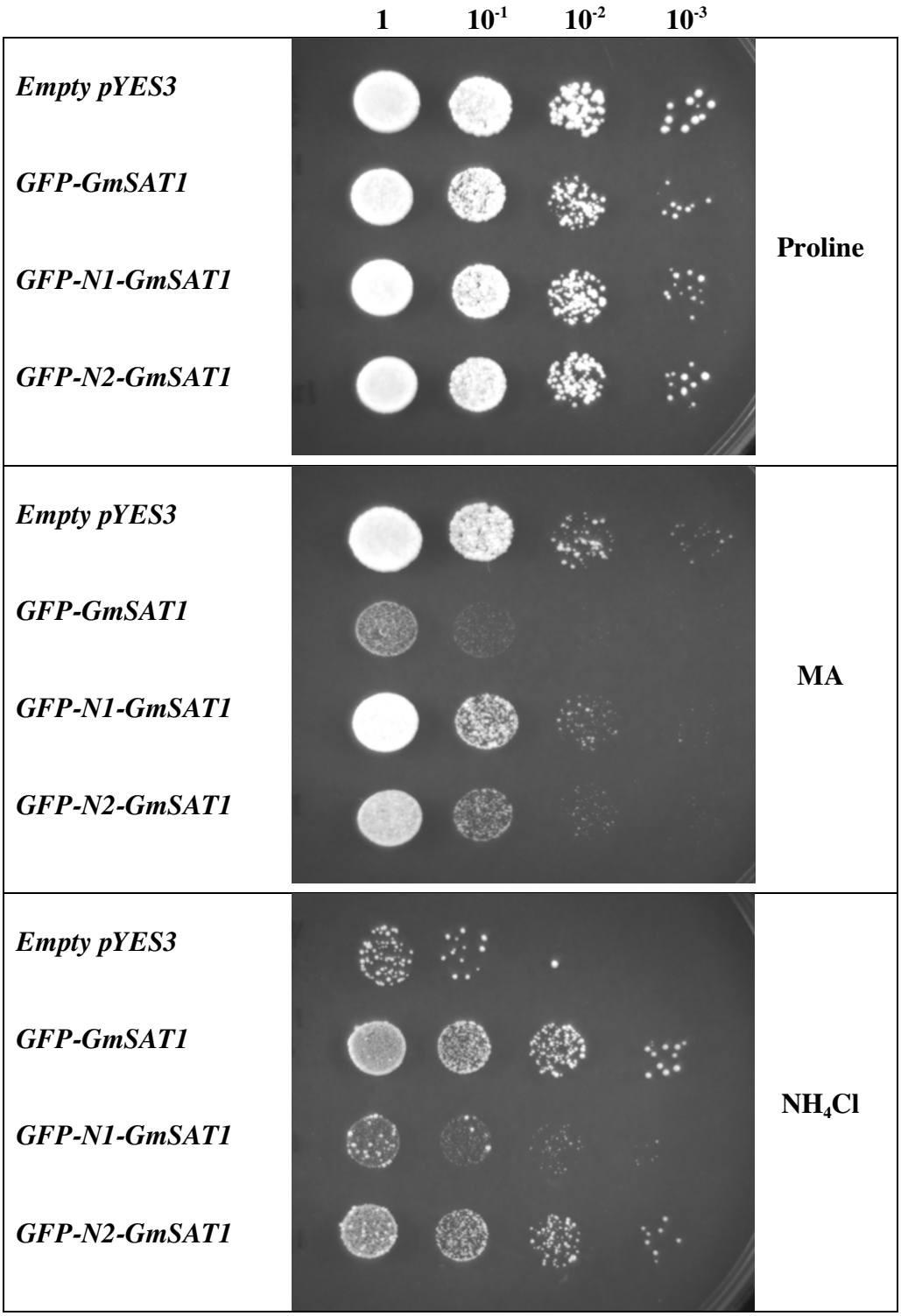


Figure 2.13. Growth of 26972c expressing GFP fusions to truncations of GmSAT1. GFP was fused to the N-terminus of GmSAT1, N1-GmSAT1, N2-GmSAT1 to test its effect on the methylammonium and ammonium phenotypes of 26972c. Yeast were grown, diluted to OD_{600} of 0.5, and spot inoculated onto Grenson's medium (with 2% (w/v) galactose) containing either 0.1% proline (w/v), 0.1% proline (w/v) + 100 mM MA, or 1 mM NH_4Cl. The leftmost spot contains an OD_{600} of 0.5, followed by three 10-fold serial diutions (indicated on top).

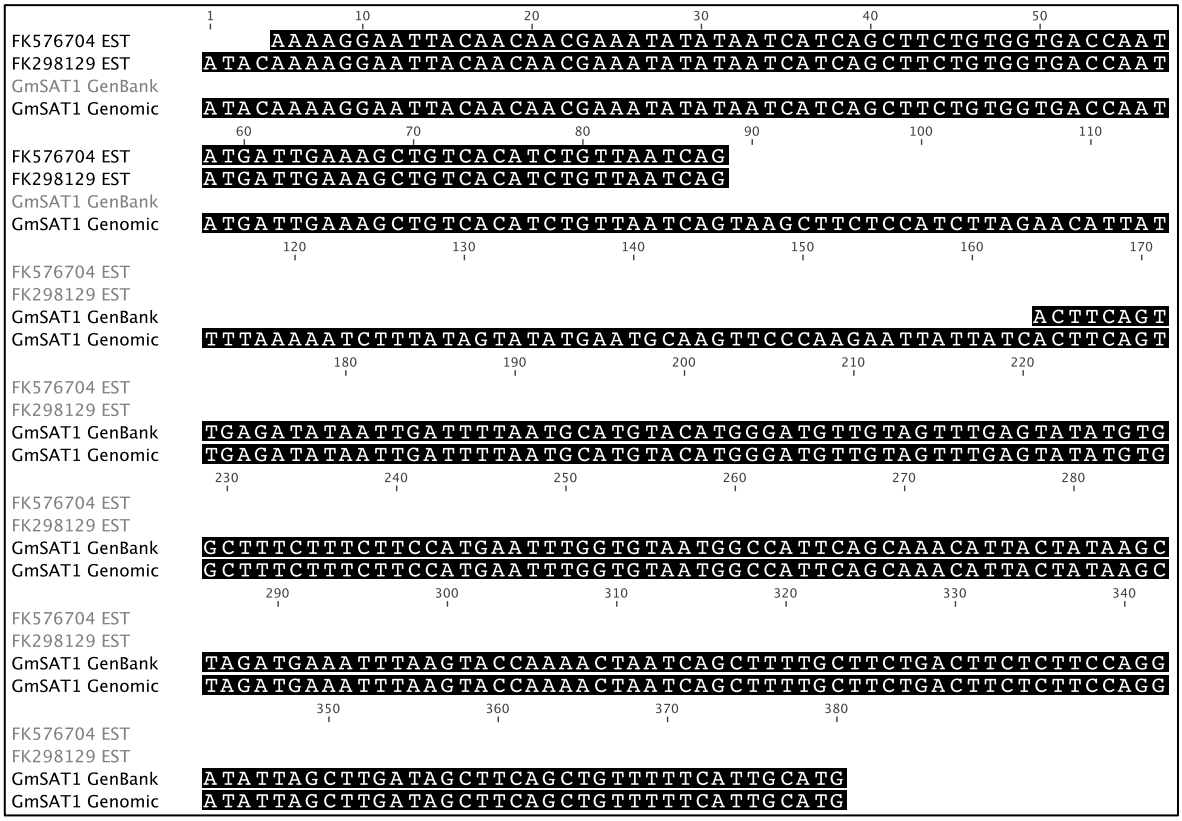


Figure 2.14. Alternative ESTs mapped to the *GmSAT1* locus.
Two ESTs were identified from GenBank (FK576704 and FK298129), which mapped upstream of the *GmSAT1* GenBank mRNA (AF069738) sequence (Kaiser et al., 1998). Also shown is the genomic sequence of *GmSAT1*, ending with the ATG start codon.

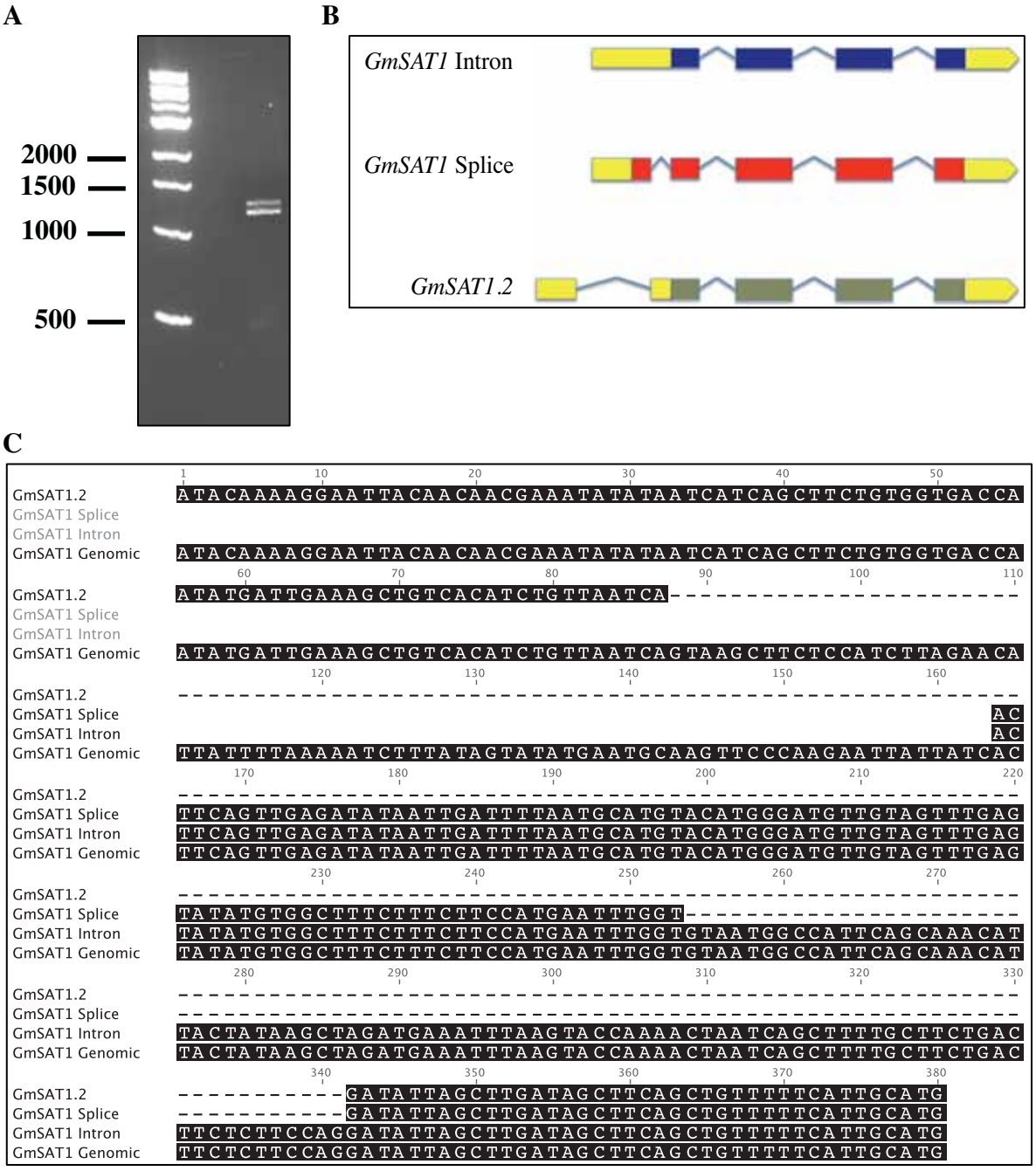
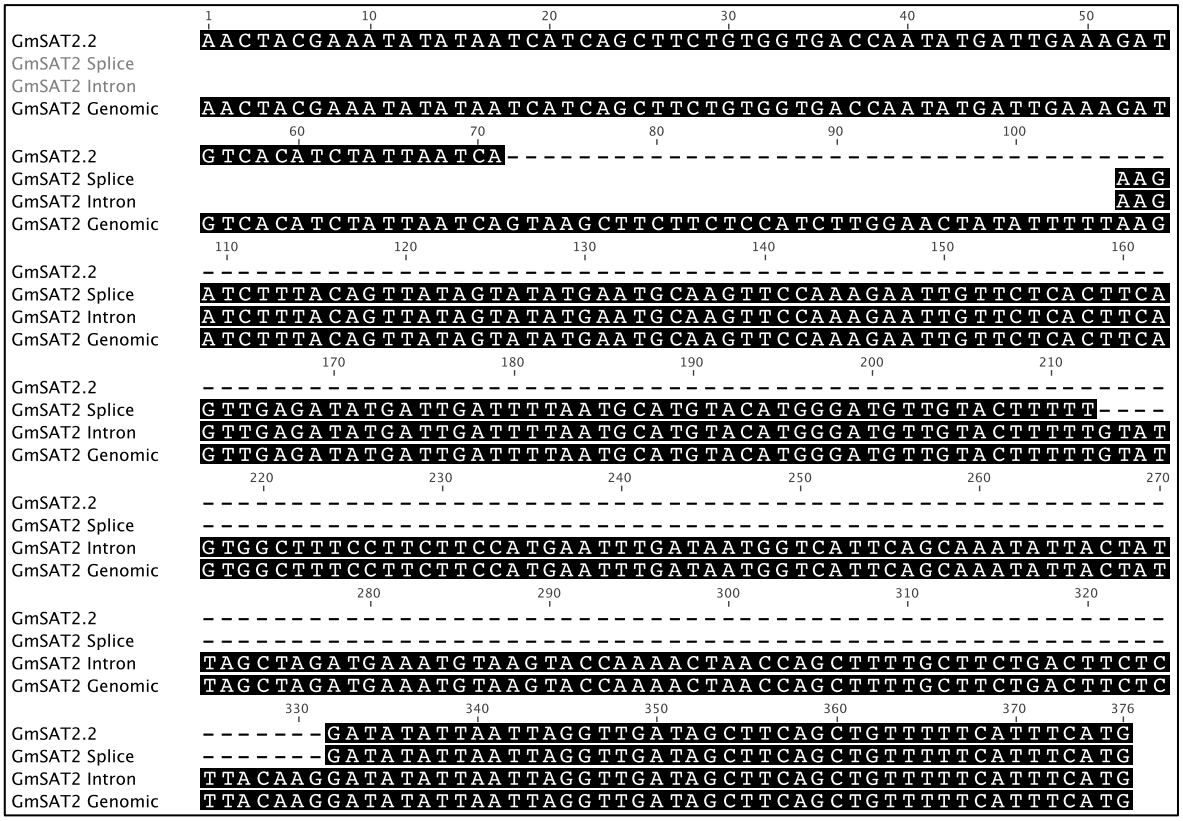


Figure 2.15. Overview of *GmSAT1* alternative transcripts.
A, Agarose gel showing the amplicon produced using with primers GmSAT1 5'UTR F and GmSAT1 CDS R (Table 2.5) with soybean nodule cDNA as the template. The size of the ladder fragments is shown on the left (bp). **B**, Overview of the three unique transcripts derived from the *GmSAT1* locus. Exons are shown as blocks, introns as joining lines. Yellow represents untranslated regions. **C**, Comparison of *GmSAT1* alternative 5'UTRs obtained from sequencing. Shown are the alternative transcript sequences against the *GmSAT1* genomic sequence, ending with the original start codon. The full-length of each transcript from the 5' end to the GmSAT1 stop codon is 1258 bp (*GmSAT1-intron*), 1170 bp (*GmSAT1-splice*), and 1167 (*GmSAT1.2*)

A



B

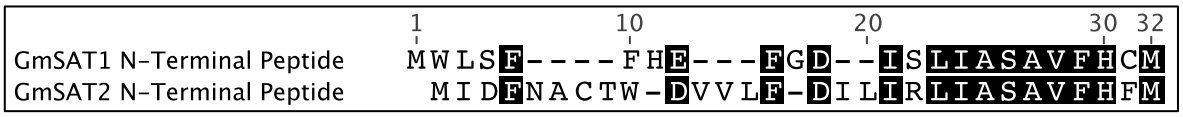
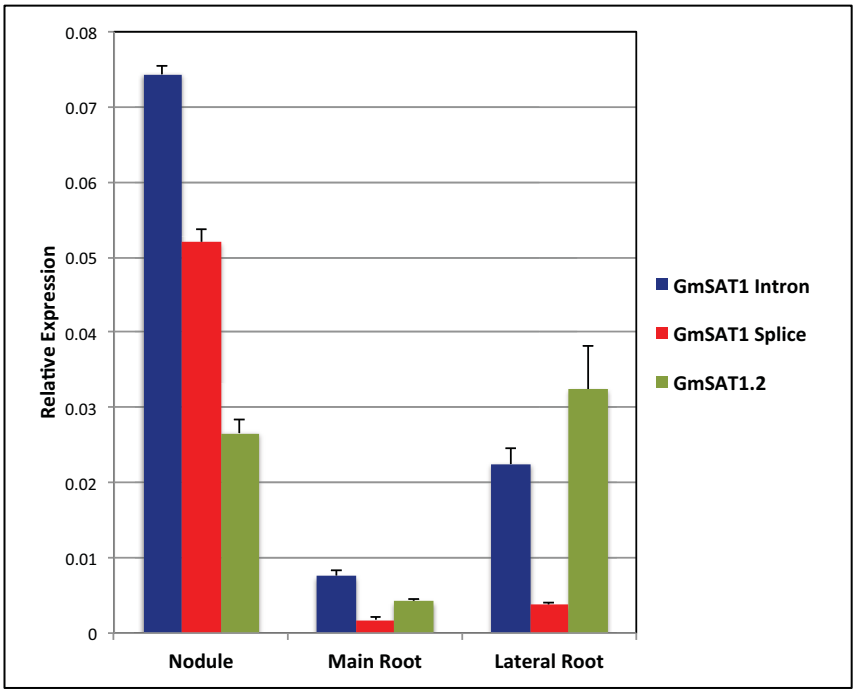


Figure 2.16. Summary of *GmSAT2* alternative transcripts.
A, Comparison of *GmSAT2* alternative 5'UTRs obtained from sequencing. Shown are the alternative transcript sequences (*GmSAT2-intron, GmSAT2-splice, GmSAT2.2*) against the *GmSAT2* genomic sequence end with the original start codon, as created with Geneious software. **B**, Comparison of the additional N-terminal peptide sequence encoded by *GmSAT1-splice* (22 amino acids) and *GmSAT2-splice* (28 amino acids).

A



B

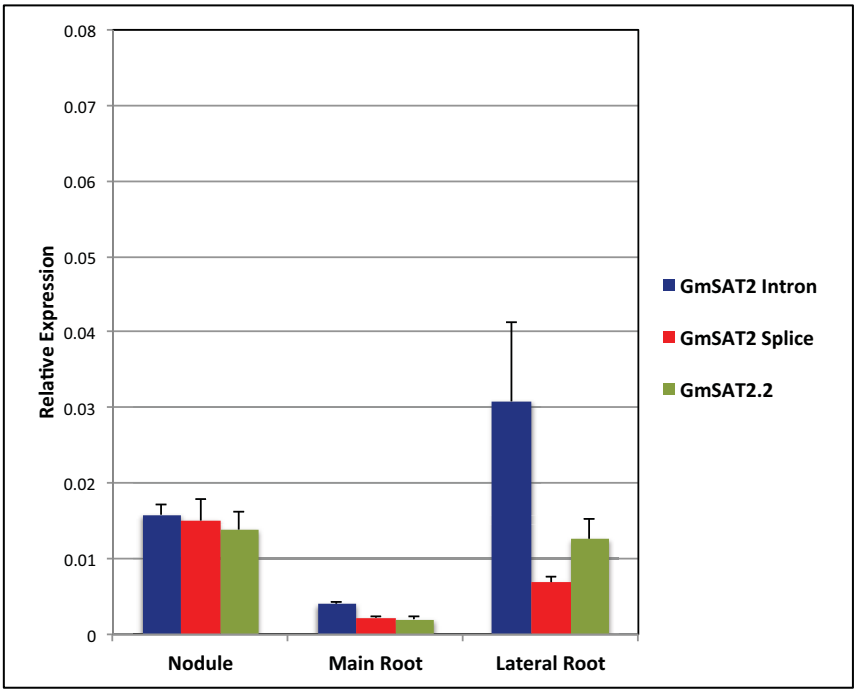
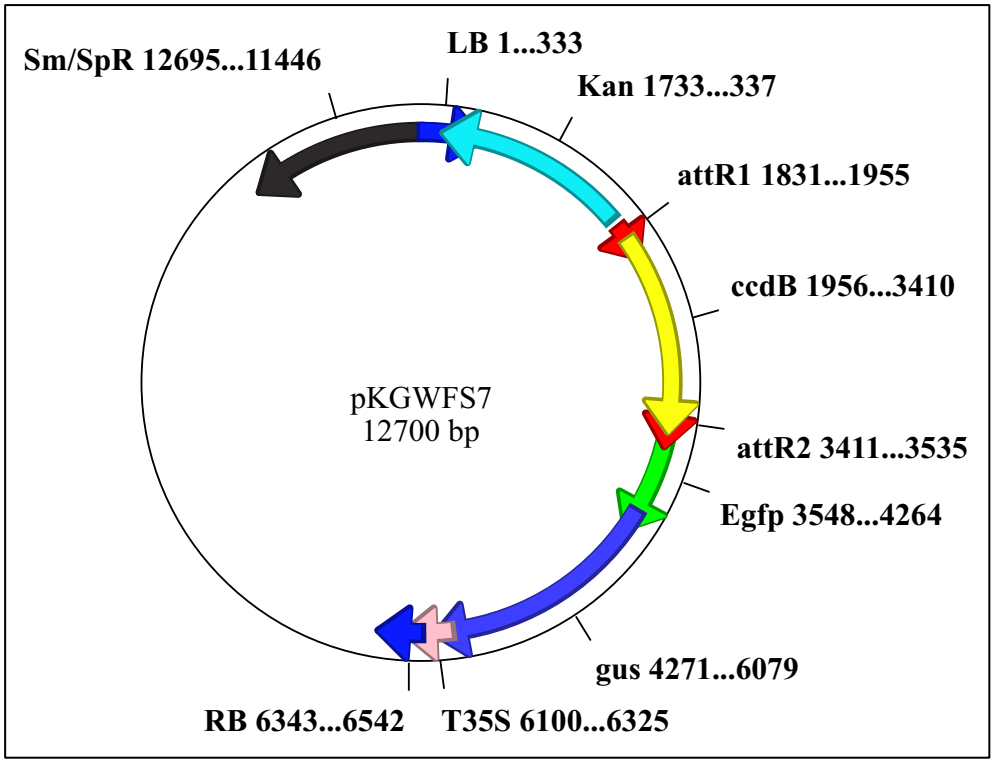


Figure 2.17. Expression of *GmSAT1* and *GmSAT2* splice variants.
qPCR expression analysis of the three splice variants identified for *GmSAT1* (**A**) and *GmSAT2* (**B**) in 36 day-old soybean tissues inoculated with *Bradyrhizobium japonicum* USDA110 at planting. RNA was extracted from ten pooled plants grown under identical conditions. Data values represent the means of three independent technical replicates ±SD relative *cons6* (Libault et al., 2008), as calculated by the ΔCT method (Livak and Schmittgen, 2001).

A



B

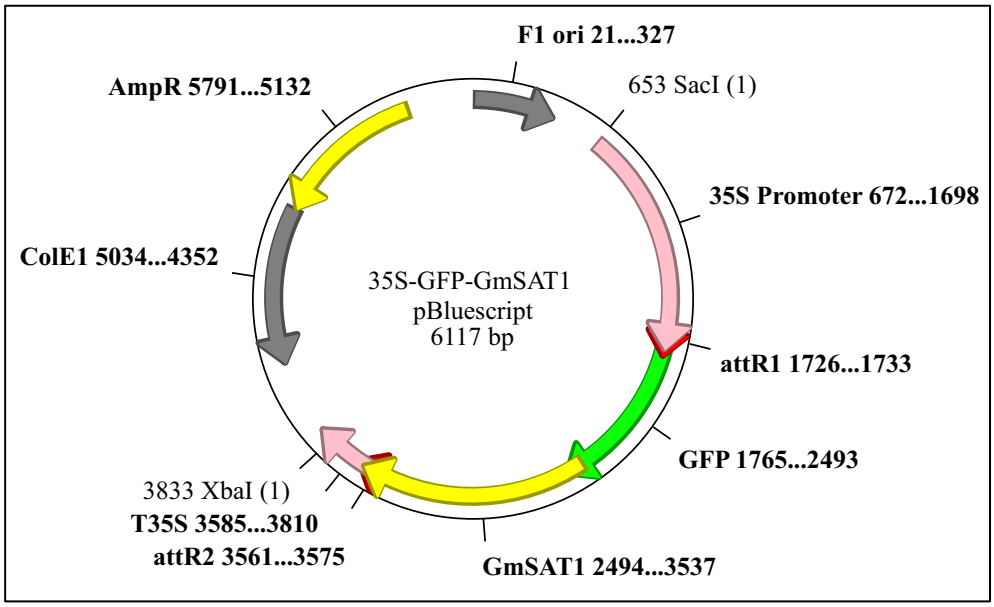


Figure 2.18. Plasmids used for GUS analysis and onion expression.
A, Plasmid pKGWFS7 used to monitor activity of the *GmSAT1* and *GmSAT2* promoters. pKGWFS7 contains a gateway cloning site to insert a promoter sequence to drive GUS/GFP expression (Karimi et al., 2002). **B**, Modified pBluescript plasmid used for expressing GFP fusions in onion epidermal peels by particle bombardment. Genes were first cloned into either pK7WG2D (to obtain a 35S promoter and 35S terminator) or pP7WGF2 (to obtain a 35S promoter, N-terminal GFP, and 35S terminator). The fragments were then excised by digestion with *Spe*I and *Xba*I and ligated into pBluescript.

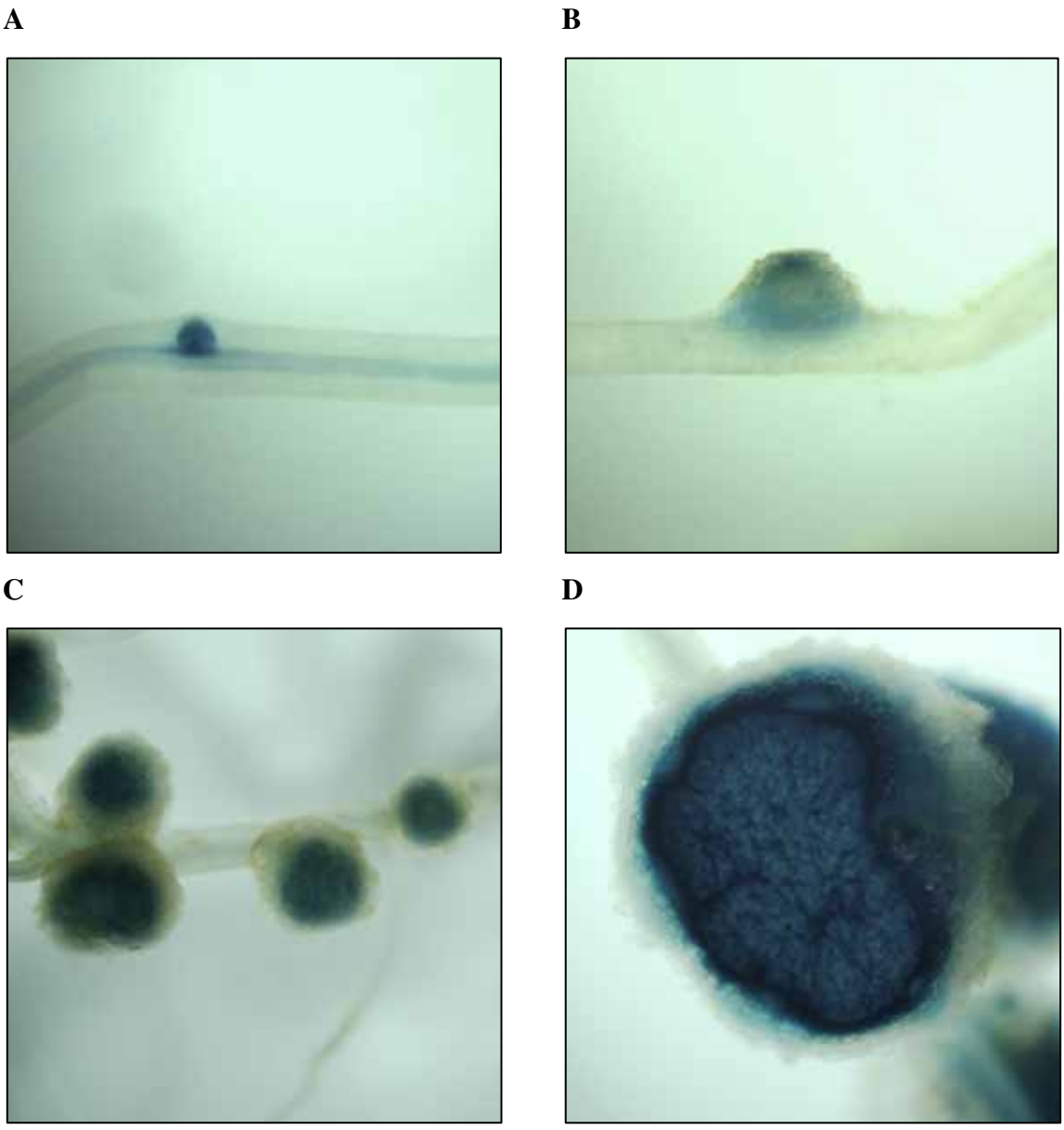


Figure 2.19. Promoter analysis of *GmSAT1* in soybean hairy roots.
Soybean hairy roots were generated by *A. rhizogenes* K599 carrying pKGWFS7 (Figure 2.18A) containing a copy of the *GmSAT1* promoter (1926 bp upstream of start codon) driving GUS/GFP. The plants were grown in sand and inoculated with *Bradyrhizobium japonicum* USDA110. **A**, **B**, The promoter was active in the root vasculature and nodules 12 days after inoculation. **C**, Expression was limited to nodules 36 days after inoculation. **D**, Image of 36 day-old nodule cross-sectioned by hand before performing the GUS assay. Images were obtained using a stereomicroscope.

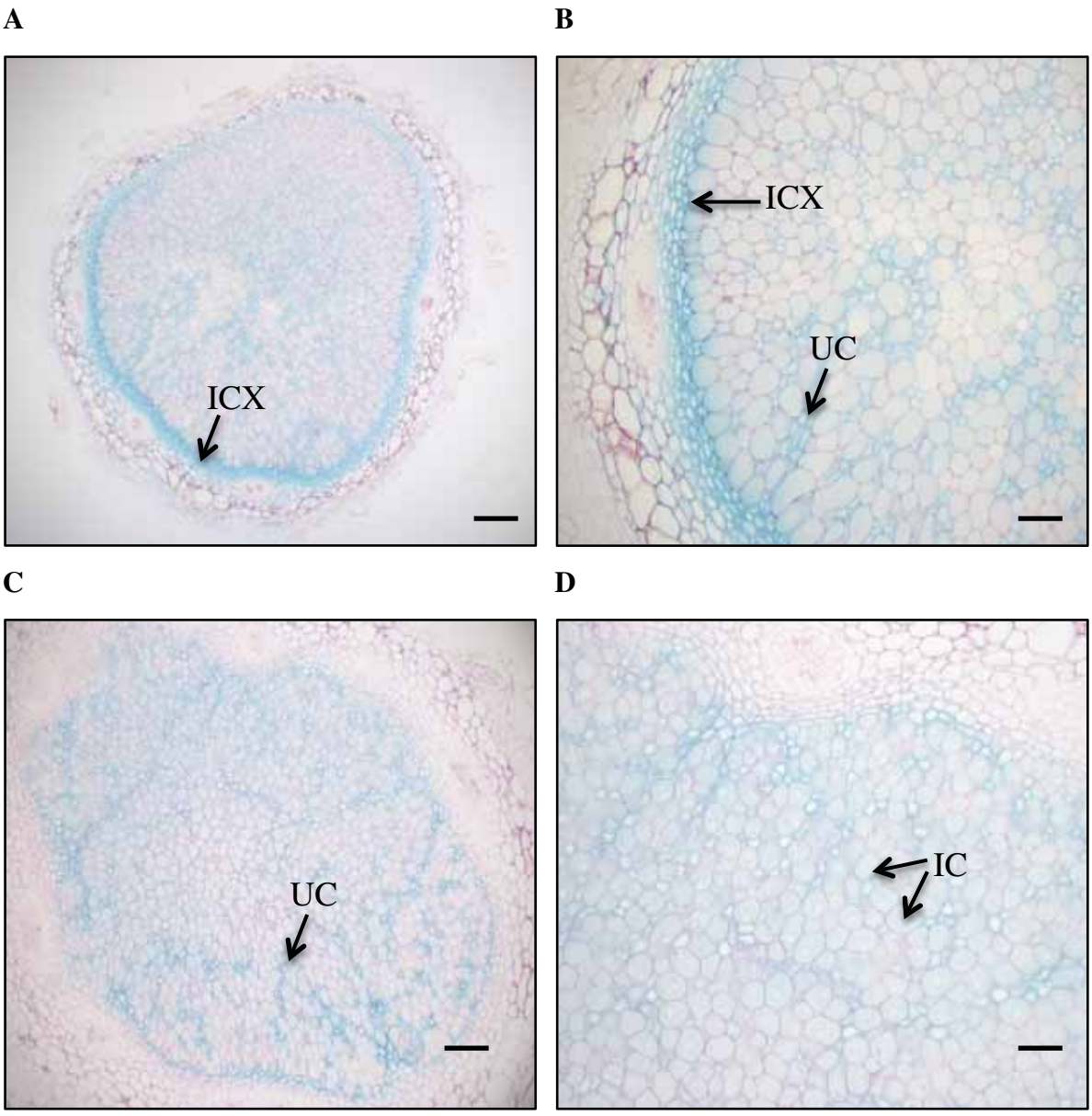


Figure 2.20. Promoter analysis of *GmSAT1* in soybean nodules.
Soybean nodules were hand-sectioned, then stained for GUS activity (8 hours). Positively stained nodules were then embedded in Technovit 7100, sectioned to 8 μm, and counterstained with ruthenium red. **A**, 28 day old nodule showing expression in the inner cortex (ICX) as well as uninfected cells. **B**, Close-up view of (**A**), highlighting uninfected cell (UC) and inner cortex (ICX) expression. **C**, 36 day-old nodule showing expression in uninfected cells (UC) in a ray-like pattern. **D**, Close-up view of (**C**), with signal visible in infected cells (IC). Bar = 200 μm (A and C), 80 μm (B and D).

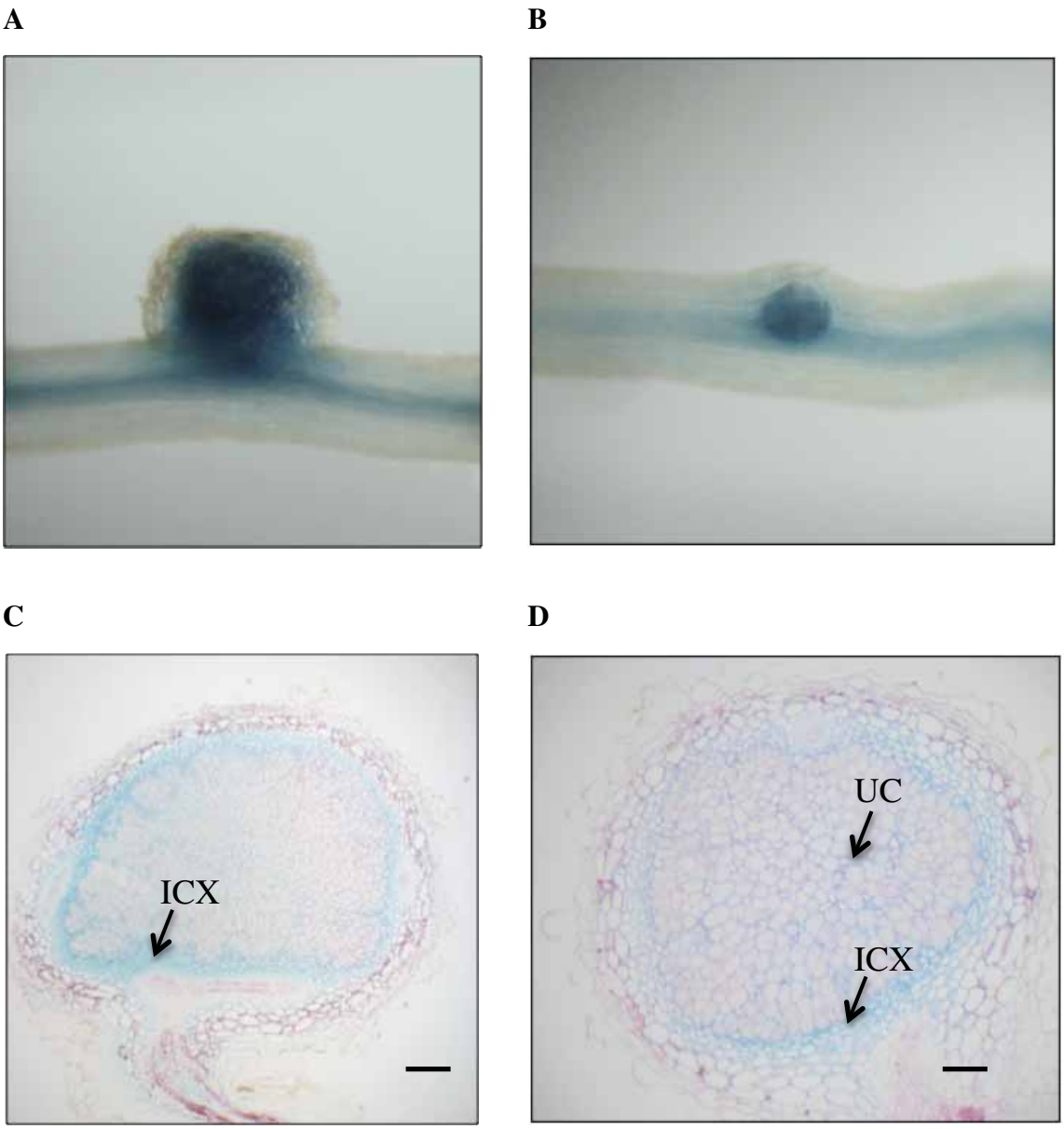


Figure 2.21. Promoter analysis of *GmSAT2* in soybean roots and nodules.
Soybean hairy roots were generated by *A. rhizogenes* K599 carrying pKGWFS7 (Figure 2.18A) containing a copy of the *GmSAT2* promoter (1989 bp upstream of start codon) driving GUS/GFP. The plants were grown in sand and inoculated with *Bradyrhizobium japonicum* USDA110. Signal was detected in the vasculature as well as developing nodules (**A**) and (**B**) 12 days after inoculation. 18 and 28 day-old soybean nodules were hand-sectioned, then stained for GUS activity (8 hours). Positively stained nodules were then embedded in Technovit 7100, sectioned to 8 μm, and counterstained with ruthenium red. Cross-sections of a 28 day-old (**C**) and 18 day-old (**D**) nodule showed expression in the inner cortex (ICX) and uninfected cells (UC). Bar = 400 μm (C) and 200 μm (D).

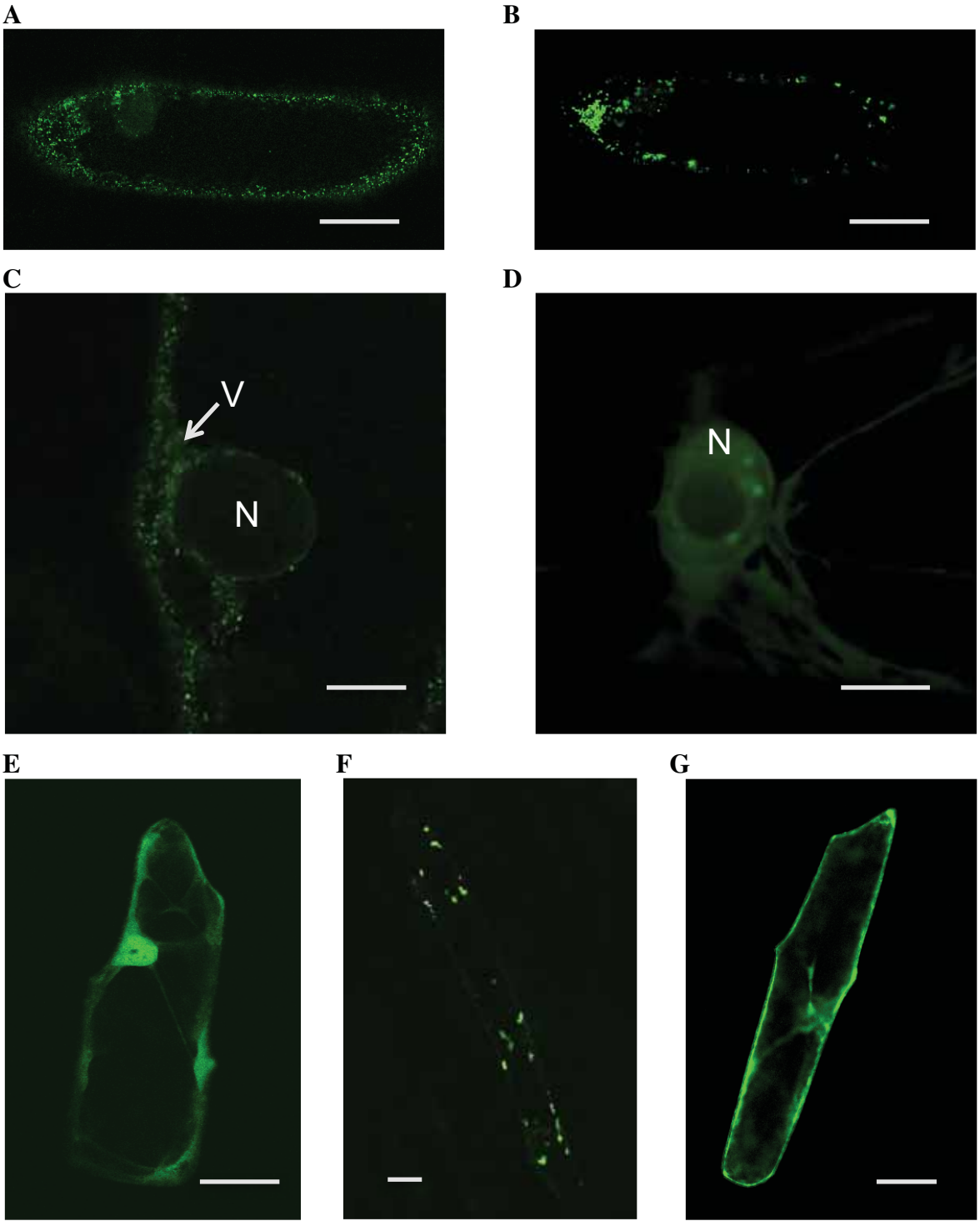
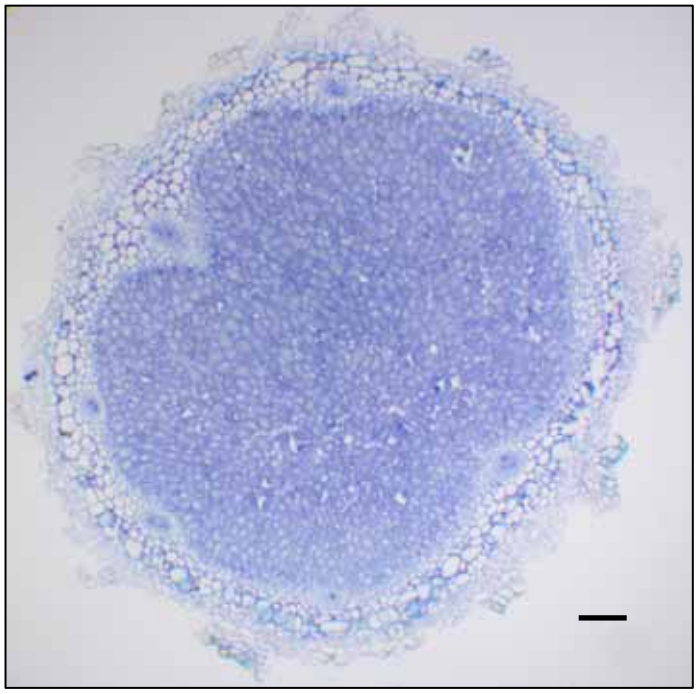


Figure 2.22. Expression of GFP fusions to GmSAT1 in onion epidermal cells.
Onion epidermal peels were bombarded with 5 μg of plasmid DNA-coated gold particles. Epidermal peels were incubated for 24 hours, then viewed by confocal microscopy. **A,** N-terminal GFP fusion to GmSAT1. **B,** C-terminal GFP fusion to GmSAT1. **C**, Close-up of GFP:GmSAT1 nucleus (N) showing expression in vesicles (V). **D**, Close-up of nucleus (N) showing localization of the GFP:T1-GmSAT1 construct. **E**, GFP alone. **F**, *GmSAT1-splice* signal peptide fused N-terminal to GFP. **G**, Nodulin-25 signal peptide fused N-terminal to GFP. Bar = 100 μm (A, B, E, F, G), 30 μm (C, D).

A



B

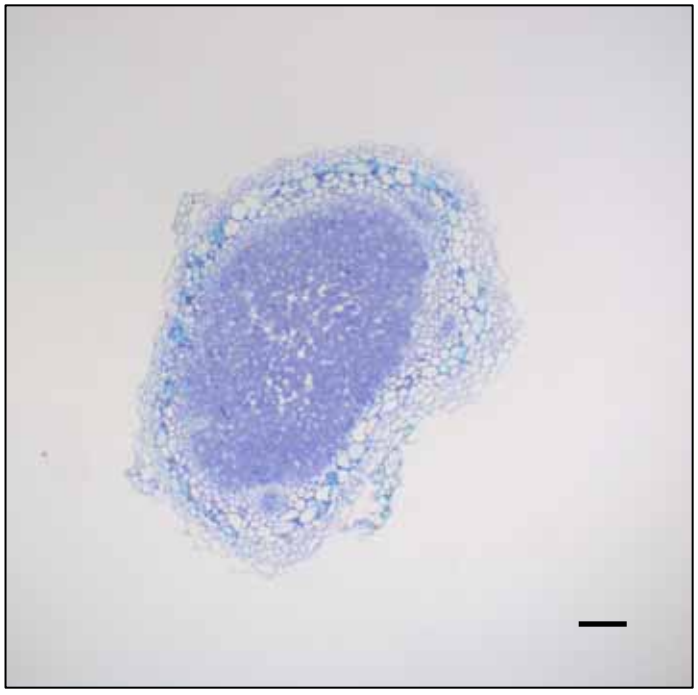
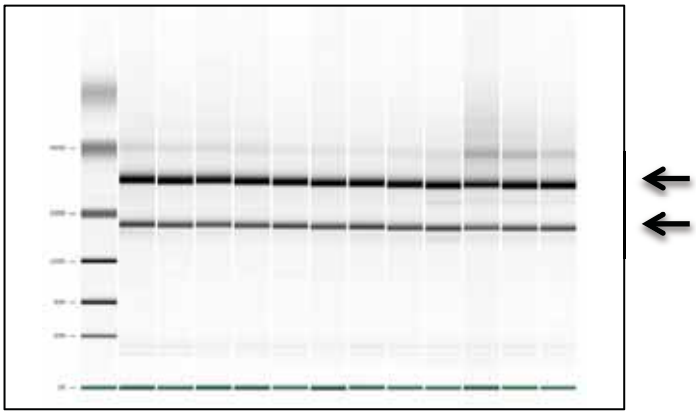


Figure 2.23. Effect of silencing *GmSAT1* on nodule development.
Sections of soybean hairy root nodules expressing empty vector control (**A**) or *GmSAT1* RNAi (**B**). These samples represent the tissue chosen for microarray analysis. Nodules were isolated 24 days after inoculation with *Bradyrhizobium japonicum* USDA110, and embedded in Technovit 7100. 8μm sections were cut and stained with toluidine blue. Bar = 200 μm.

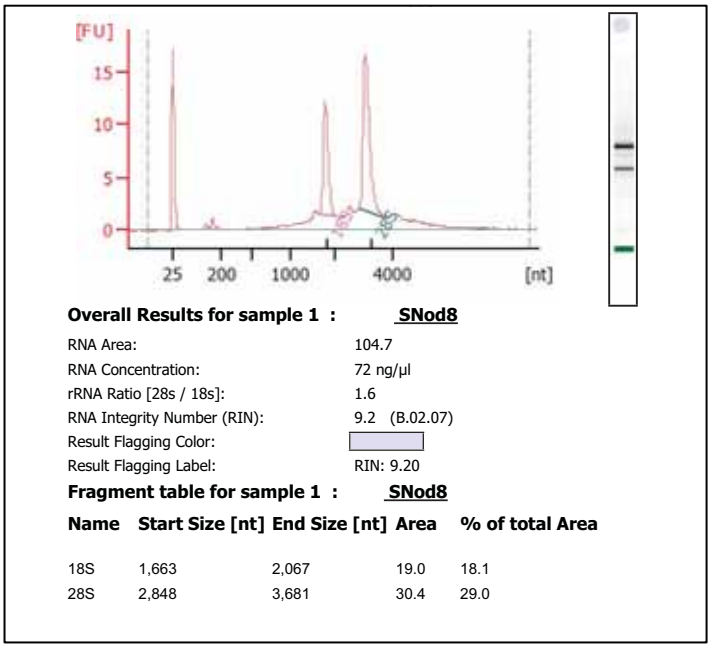
Sample Name	*GmSAT1* C_T	*cons6* C_T	Fold Change
GmSAT1 Vector Nodule 2	25.1	22.7	
GmSAT1 Vector Nodule 6	24.5	22.0	
GmSAT1 Vector Nodule 8	24.8	22.5	
GmSAT1 Vector Nodule 10	24.3	21.7	
GmSAT1 RNAi Nodule 8	29.1	22.9	-13.9
GmSAT1 RNAi Nodule 9	27.6	21.9	-9.6
GmSAT1 RNAi Nodule 11	28.4	22.8	-9.2
GmSAT1 RNAi Nodule 14	27.7	21.8	-9.6

Table 2.1. Expression level of *GmSAT1* in RNAi versus empty vector control.
qPCR was performed with GmSAT1 qPCR F + R and cons6 qPCR F + R primers (Libault et al., 2008). The C_T values shown are the average of three replicates. The fold change value (relative to vector controls) was generated using the calculation $2^{-\Delta\Delta CT}$ (Schmittgen and Livak, 2008). The nodule samples shown were chosen for microarray analysis.

A



B



Overall Results for sample 1 : SNod8

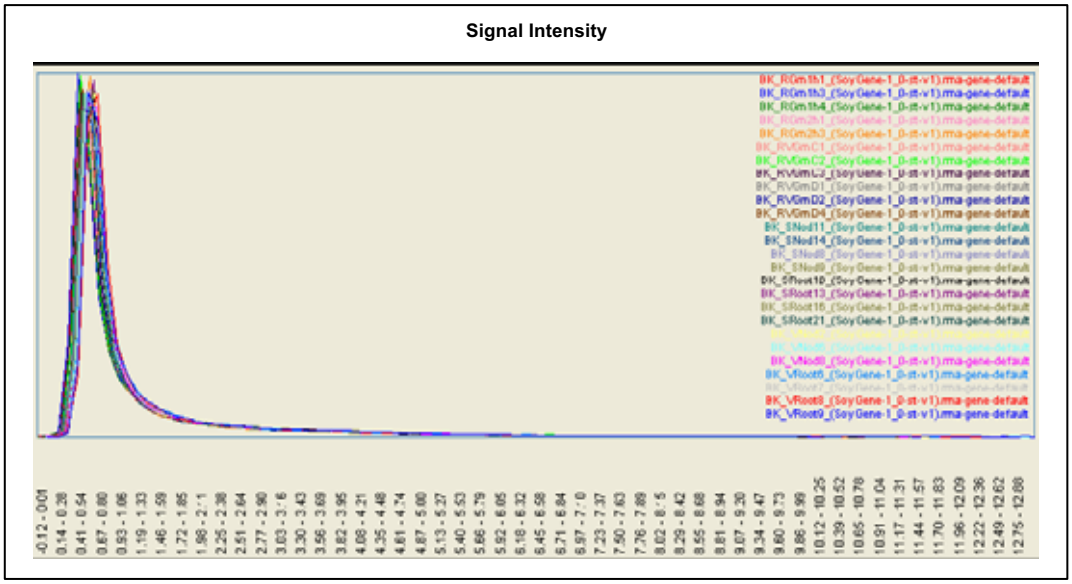
RNA Area:	104.7
RNA Concentration:	72 ng/µl
rRNA Ratio [28s / 18s]:	1.6
RNA Integrity Number (RIN):	9.2 (B.02.07)
Result Flagging Color:	
Result Flagging Label:	RIN: 9.20

Fragment table for sample 1 : SNod8

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	1,663	2,067	19.0	18.1
28S	2,848	3,681	30.4	29.0

Figure 2.24. Assessment of the RNA quality for microarray analysis.
RNA was sent to the Ramaciotti Centre and checked for quality before performing microarray analysis. **A**, Agilent bioanalzyer virtual gel showing RNA purity. Ribosomal bands (5S, 5.8S, 18S and 25S) are clearly visible and intact, with minimal genomic DNA contamination. Arrows indicate major ribosomal bands. **B**, Electropherogram summary of a representative *GmSAT1* RNAi sample showing quantitation of bands, as well as the RNA integrity number (RIN= 9.2, >8 indicates high quality).

A



B

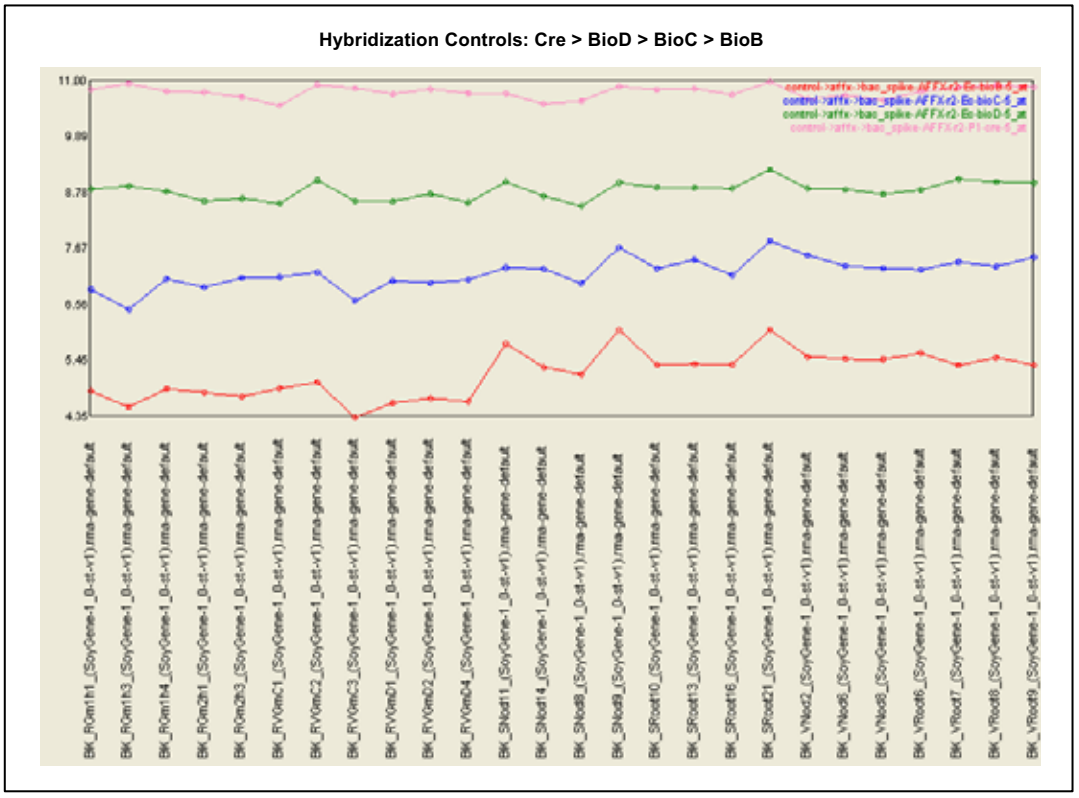


Figure 2.25. Quality control analysis of scanned soybean microarrays.
A, Signal intensity plot for all microarrays analyzed. B, Summary of signal from four hybridization controls present on each array. All arrays passed the post-scanning quality control. Included in the diagrams are additional samples from other experiments.

Gene Description	*Arabidopsis* Gene Name	Reference Sequence	Fold Change	p-value
Phosphatidylinositol transfer family protein	SEC14	Glyma10g20390	-5.77	0.0302
Nuclear RNA polymerase C2	NRPC2	Glyma01g00980	-4.93	0.0407
Zinc transporter 1 precursor	ZIP1	Glyma13g10790	-4.10	0.0384
Scpl31 serine carboxypeptidase-like 31	Scpl31	Glyma13g31690	-3.85	0.0263
bHLH transcription factor	(GmSAT1)	Glyma15g06680	-3.76	0.0001
Pseudo-response regulator 5	PRR5	Glyma04g40640	-3.52	0.0046
Senescence-associated gene 21	SAG21	Glyma10g02210	-3.36	0.0336
Oxidoreductase, zinc-binding dehydrogenase family protein		Glyma19g01120	-3.29	0.0089
Gigantea protein	GI	Glyma20g30980	-3.20	0.0121
Unknown		Glyma15g02110	-3.08	0.0080
Snf1-related protein kinase 2.4	SNRK2.4	Glyma08g13380	-2.98	0.0378
ABC transporter G family member	ABCG37	Glyma17g03860	-2.97	0.0362
Unknown		Glyma13g21410	-2.90	0.0121
SUPPRESSOR OF ACAULIS 51 bHLH transcription factor	SAC51	Glyma14g35810	-2.90	0.0050
ABC transporter family protein		Glyma19g01980	-2.88	0.0333
Unknown		Glyma20g11250	-2.88	0.0026
Disease resistance protein (TIR-NBS-LRR class) family		Glyma06g40740	-2.85	0.0050
Disease resistance protein (TIR-NBS-LRR class), putative		Glyma16g25080	-2.79	0.0064
Zinc finger (C3HC4-type RING finger) family protein		Glyma17g11370	-2.74	0.0292
BETA-VPE beta vacuolar processing enzyme		Glyma17g14680	-2.73	0.0136
NAD(P)-linked oxidoreductase superfamily protein		Glyma03g11580	-2.70	0.0023
Calcium/calmodulin-dependent protein kinase-related		Glyma11g04220	-2.69	0.0073
Calcium/calmodulin-dependent protein kinase-related		Glyma17g17840	-2.69	0.0073
Oxidoreductase, zinc-binding dehydrogenase family protein		Glyma19g01150	-2.66	0.0362
Homogentisate phytyltransferase 1	HPT1	Glyma10g44170	-2.65	0.0438
Unknown		Glyma15g19380	-2.63	0.0110
MYB-related		Glyma06g44120	-2.62	0.0015
MYB-related		Glyma16g16800	-2.62	0.0015
Serine acetyltransferase 3;2	SERAT3;2	Glyma02g04770	-2.59	0.0392
Conserved peptide upstream open reading frame 37	CPuORF37	Glyma14g35820	-2.59	0.0004
IQ calmodulin-binding motif family protein		Glyma10g07860	-2.58	0.0179
Leucine-rich repeat protein kinase family protein		Glyma18g05710	-2.58	0.0005
Unknown		Glyma06g07160	-2.57	0.0048
RNA-binding KH domain-containing protein		Glyma11g31530	-2.56	0.0002
Unknown		Glyma17g06560	-2.51	0.0174
ABC transporter of the mitochondrion 3	ATM3	Glyma14g38800	-2.48	0.0304
Calcium/calmodulin-dependent protein kinase-related		Glyma06g30920	-2.47	0.0081
NB-ARC domain-containing disease resistance protein		Glyma12g36510	-2.40	0.0089
MYB-related		Glyma14g27020	-2.40	0.0014
Unknown		Glyma01g28790	-2.39	0.0018
Unknown		Glyma09g29240	-2.39	0.0018
Unknown		Glyma10g14830	-2.39	0.0018
Unknown		Glyma15g33040	-2.39	0.0018
Unknown		Glyma16g32100	-2.39	0.0018
SKP2A F-box/RNI-like superfamily protein		Glyma06g12640	-2.39	0.0202
Senescence-associated protein-related		Glyma20g28500	-2.39	0.0113
Unknown		Glyma11g07270	-2.35	0.0138
Terpene synthase 04	TPS04	Glyma13g25270	-2.34	0.0392
Ribosomal protein S12/S23 family protein		Glyma03g22190	-2.32	0.0167
Gigantea protein	GI	Glyma09g07240	-2.30	0.0055
Elicitor-activated gene 3-2	ELI3-2	Glyma18g38670	-2.29	0.0082
Pyridoxal phosphate (PLP)-dependent transferases superfamily		Glyma06g17810	-2.28	0.0049
Unknown		Glyma08g27810	-2.26	0.0067
Unknown		Glyma11g12510	-2.26	0.0061
Senescence-associated protein-related		Glyma10g39210	-2.26	0.0003
Cold, circadian rhythm, and rna binding 2	CCR2	Glyma06g01470	-2.24	0.0000
Glucuronokinase G	GLCAK	Glyma05g09130	-2.24	0.0415
Unknown		Glyma08g26570	-2.23	0.0427
Disease resistance family protein / LRR family protein		Glyma16g28570	-2.22	0.0365
NAC (No Apical Meristem) domain transcriptional regulator	ATAF1	Glyma05g32850	-2.21	0.0467
Disease resistance protein (TIR-NBS-LRR class), putative		Glyma16g24940	-2.20	0.0112
Developmental regulator, ULTRAPETALA	ULT1	Glyma06g45180	-2.19	0.0221
Unknown		Glyma12g05080	-2.17	0.0008
Unknown		Glyma18g14100	-2.16	0.0147
Beta-1,4-N-acetylglucosaminyltransferase family protein		Glyma08g14070	-2.16	0.0299
Glycerol-3-phosphate acyltransferase 6	GPAT6	Glyma18g42580	-2.14	0.0307
2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase super		Glyma12g05890	-2.13	0.0252
Pseudo-response regulator 7	PRR7	Glyma12g07860	-2.13	0.0103
Unknown		Glyma03g31410	-2.12	0.0031
Calcium-binding EF-hand family protein		Glyma03g28650	-2.12	0.0024
Serine-rich protein-related		Glyma04g06860	-2.12	0.0049
Carbohydrate transmembrane transporter	MSS1	Glyma20g28230	-2.11	0.0212
Tetratricopeptide repeat (TPR)-containing protein		Glyma04g11230	-2.11	0.0022
Cold regulated gene 27	COR27	Glyma17g36060	-2.11	0.0070
Nicotinamidase 1	NIC1	Glyma08g07210	-2.10	0.0416
NADH-dependent glutamate synthase 1	GLT1	Glyma06g13280	-2.10	0.0451
Uricase, nodulin 35, putative		Glyma20g17440	-2.09	0.0123
NF-YA3 "nuclear factor Y, subunit A3		Glyma15g18970	-2.08	0.0355
Putative endonuclease or glycosyl hydrolase		Glyma10g28450	-2.07	0.0109
Vacuolar sorting receptor homolog 1	VSR1	Glyma18g44580	-2.06	0.0013
Unknown		Glyma09g33530	-2.06	0.0014
Leucine-rich receptor-like protein kinase family protein		Glyma18g44600	-2.06	0.0142
Unknown		Glyma05g15860	-2.05	0.0027
Unknown		Glyma13g16070	-2.04	0.0365
NB-ARC domain-containing disease resistance protein	RPS2	Glyma07g08500	-2.04	0.0153
Unknown		Glyma13g32240	-2.03	0.0023
LRP1 Lateral root primordium (LRP) protein-related		Glyma13g26730	-2.02	0.0081
GAMMA-VPE gamma vacuolar processing enzyme		Glyma17g34900	-2.01	0.0361
NF-YA3 "nuclear factor Y, subunit A3"		Glyma07g04050	-2.01	0.0230
Acyl-CoA thioesterase family protein		Glyma18g46610	-2.01	0.0091
Unknown		Glyma14g35490	-2.01	0.0017
Peroxidase superfamily protein		Glyma02g15290	-2.00	0.0422
Translation elongation factor EFG/EF2 protein		Glyma01g14740	-1.99	0.0112
Unknown		Glyma10g27430	-1.99	0.0281
Major facilitator superfamily protein	(GmMFS1.5)	Glyma09g33680	-1.98	0.0070

Table 2.2. List of genes significantly downregulated (≥ -2 fold change, p-value ≤ 0.05) upon silencing of *GmSAT1* by RNAi.

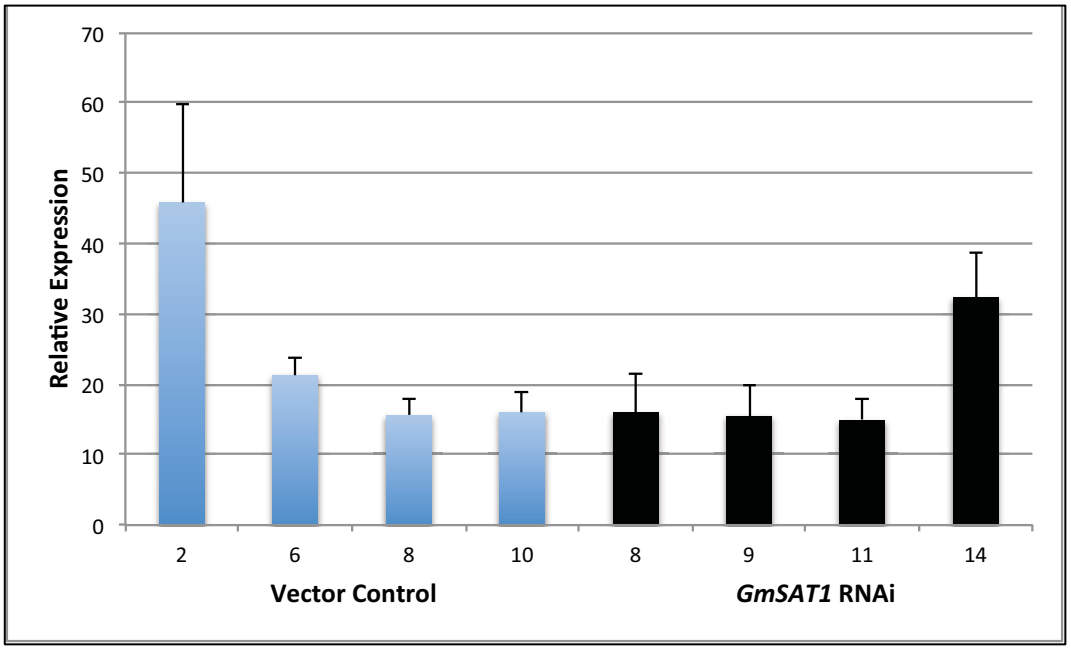
Gene Name	Fold Change qPCR	Fold Change Array
GmSAT1	-8.09	-3.76
GmSAT2	-4.34	-1.30
GI (Glyma20g30980)	-2.05	-3.12
GI (Glyma09g07240)	-3.40	-2.30
GmMFS1.5	-2.38	-1.98
SNRK2.4	-3.00	-2.98
Zip1	-6.42	-4.10

Table 2.3. Expression analysis of selected transcripts from *GmSAT1* nodule RNAi. qPCR was performed with primers shown in Table 2.5 on the four RNA replicates used for the microarray experiment. qPCR fold change values were generated using the calculation $2^{-\Delta\Delta CT}$ (Schmittgen and Livak, 2008) using *cons6* (Libault et al., 2008) for normalization. Fold changes are expressed as a ratio of RNAi to empty vector control. All samples chosen showed at least a 2-fold downregulation by qPCR, in agreement with the microarray results. Note that *GmSAT2*, which is expressed at a relatively low level, was found to be significantly downregulated by qPCR, in contrast to the microarray results.

Gene Description	*Arabidopsis* Gene Symbol	Reference Sequence	Fold-Change	p-value
Peroxidase superfamily protein		Glyma04g39860	4.78	0.0034
Bifunctional inhibitor/lipid-transfer protein/seed storage		Glyma17g14890	3.84	0.0048
Gamma tonoplast intrinsic protein		Glyma03g34310	3.83	0.0010
Plasma membrane intrinsic protein 1;4	PIP1;4	Glyma11g35030	3.62	0.0280
O-methyltransferase 1	OMT1	Glyma20g31700	3.53	0.0138
Cytochrome c oxidase subunit Vc family protein		Glyma10g01600	3.48	0.0219
Unknown		Glyma0973s00200	3.27	0.0002
Basic leucine-zipper 42	bZIP42	Glyma18g51250	3.16	0.0278
Peroxidase superfamily protein		Glyma06g15030	2.99	0.0049
GRAM domain-containing protein / ABA-responsive protein		Glyma20g23520	2.89	0.0361
Unknown		Glyma02g47770	2.86	0.0162
Eukaryotic aspartyl protease family protein		Glyma08g23600	2.77	0.0046
CAMV movement protein interacting protein 7	MPI7	Glyma10g32370	2.75	0.0387
Unknown		Glyma19g43010	2.71	0.0178
BANQUO 3	BNQ3	Glyma04g33920	2.62	0.0262
SAUR-like auxin-responsive protein family		Glyma06g16640	2.62	0.0473
Mitochondrial outer membrane translocase complex, subunit		Glyma08g03150	2.61	0.0216
Cellulose synthase-like A02	CSLA02	Glyma12g08990	2.59	0.0183
Unknown		Glyma13g21290	2.52	0.0303
Unknown		Glyma02g43620	2.49	0.0205
basic helix-loop-helix (bHLH) DNA-binding superfamily		Glyma20g39220	2.47	0.0494
Pectin lyase-like superfamily protein		Glyma19g00230	2.47	0.0220
Xyloglucan endotransglucosylase/hydrolase 5	XTH5	Glyma16g26630	2.45	0.0350
Peroxidase superfamily protein		Glyma18g06230	2.44	0.0149
Unknown		Glyma08g06250	2.37	0.0020
Bifunctional inhibitor/lipid-transfer protein/seed storag		Glyma05g04390	2.36	0.0386
Unknown		Glyma11g06600	2.34	0.0146
Uridine kinase-like 4	UKL4	Glyma18g08180	2.32	0.0304
Gamma tonoplast intrinsic protein		Glyma19g37000	2.31	0.0258
FH interacting protein 1	FIP1	Glyma06g09980	2.30	0.0278
Carbonic anhydrase 2	CA2	Glyma06g19400	2.29	0.0192
Ribose 5-phosphate isomerase, type A protein		Glyma19g43310	2.28	0.0037
TCP family transcription factor		Glyma07g08710	2.28	0.0178
Unknown		Glyma13g30350	2.28	0.0177
O-methyltransferase family protein		Glyma14g38110	2.27	0.0429
NADH-ubiquinone oxidoreductase 20 kDa subunit		Glyma06g05410	2.26	0.0004
S-adenosyl-L-homocysteine hydrolase	MEE58	Glyma11g36620	2.25	0.0224
EIN3-binding F box protein 1	EBF1	Glyma04g20330	2.23	0.0342
Unknown		Glyma15g38440	2.23	0.0415
Unknown		Glyma16g02970	2.23	0.0191
Unknown		Glyma09g35650	2.22	0.0124
Coatomer epsilon subunit		Glyma15g14450	2.21	0.0266
Cytochrome C oxidase 6B	COX6B	Glyma13g01590	2.21	0.0144
Senescence associated gene 20	SAG20	Glyma18g02610	2.20	0.0008
Unknown		Glyma12g32250	2.17	0.0019
S-adenosylmethionine synthetase family protein	MTO3	Glyma17g04330	2.16	0.0205
Carbohydrate-binding X8 domain superfamily protein		Glyma10g28470	2.12	0.0472
Ubiquinol-cytochrome C reductase UQCRX/QCR9-like family		Glyma13g22090	2.10	0.0273
ATPase, V0/A0 complex, subunit C/D		Glyma18g49110	2.09	0.0055
zinc finger protein 6	ZFP6	Glyma08g48230	2.05	0.0473
Subtilase family protein		Glyma03g42440	2.05	0.0061
Phloem protein 2-A13	PP2-A13	Glyma09g26940	2.05	0.0213
Unknown		Glyma03g28430	2.03	0.0220
Phospholipases;galactolipases		Glyma07g37200	2.03	0.0422
O-methyltransferase 1	OMT1	Glyma20g31610	2.02	0.0139
Plant invertase/pectin methylesterase inhibitor superfami		Glyma15g20060	2.02	0.0062
Protein kinase superfamily protein		Glyma02g02840	2.01	0.0078
Protein of unknown function (DUF1442)		Glyma05g03920	1.99	0.0112
Unknown		Glyma01g03280	1.99	0.0220
Ubiquinol-cytochrome C reductase iron-sulfur subunit		Glyma02g07340	1.98	0.0001
Disease resistance-responsive (dirigent-like protein) fam		Glyma11g21010	1.98	0.0003
Expansin A4	EXPA4	Glyma02g41590	1.98	0.0196
Unknown		Glyma06g04110	1.98	0.0366
SELT-like protein precursor		Glyma03g07550	1.95	0.0116
Ribose-5-phosphate isomerase 2	RPI2	Glyma20g26110	1.95	0.0002
Protein of unknown function (DUF1664)		Glyma18g52020	1.94	0.0250
Gibberellin-regulated family protein		Glyma14g09620	1.94	0.0183
Unknown			1.93	0.0134
Leucine-rich repeat (LRR) family protein		Glyma10g43450	1.93	0.0038
Homeobox-leucine zipper protein family	HAT22	Glyma0041s00350	1.92	0.0164
Farnesylated protein 6	FP6	Glyma04g00500	1.90	0.0055
Unknown		Glyma04g03760	1.90	0.0239
Ethylene response factor 1	ERF1	Glyma10g04210	1.90	0.0204
UDP-glucose pyrophosphorylase 2	UGP2	Glyma14g39140	1.89	0.0305
Wound-responsive family protein		Glyma13g37720	1.88	0.0468
EIN3-binding F box protein 1	EBF1	Glyma14g14410	1.88	0.0310
		Glyma18g07870	1.88	0.0011
Phospholipase D alpha 1	PLDALPHA1	Glyma08g22600	1.87	0.0401
VPS54	VPS54	Glyma15g22220	1.86	0.0381
PfkB-like carbohydrate kinase family protein		Glyma10g03920	1.86	0.0270
Cox19-like CHCH family protein		Glyma13g32930	1.85	0.0071
Myb family transcription factor		Glyma05g01640	1.85	0.0449
Ribonuclease III family protein	NFD2	Glyma17g09900	1.84	0.0294
Protein of unknown function (DUF544)		Glyma11g02970	1.84	0.0430
Ethylene responsive element binding factor 1	ERF-1	Glyma17g15480	1.83	0.0433
Cystathionine beta-synthase (CBS) family protein	LEJ2	Glyma07g33870	1.83	0.0153
EMB30032-oxoacid dehydrogenases acyltransferase family		Glyma20g31630	1.83	0.0281
Pyridoxal phosphate phosphatase-related protein		Glyma08g03350	1.82	0.0479
GRAS family transcription factor		Glyma12g02060	1.82	0.0252
Plasma membrane intrinsic protein 2	PIP2B	Glyma10g35520	1.80	0.0108

Table 2.4. List of genes significantly upregulated (≥ 2 fold change, p-value ≤ 0.05) upon silencing of *GmSAT1* by RNAi.

A



B

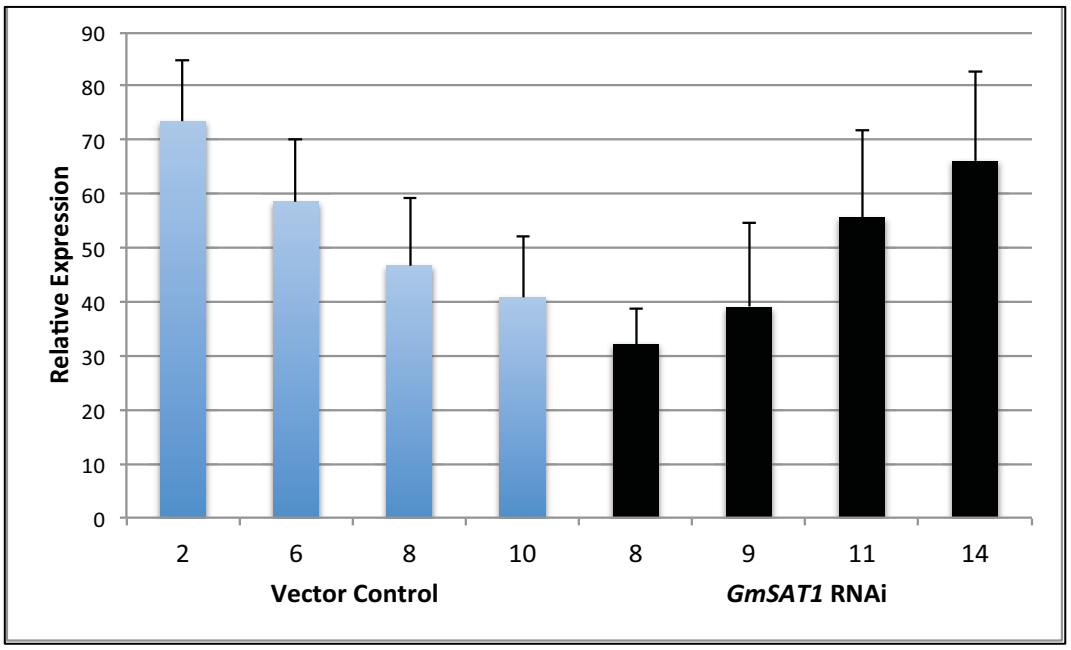


Figure 2.26. qPCR expression analysis of *B. japonicum* fixation marker genes. qPCR analysis of **A**, *FixU*, and **B**, *NifH* transcripts from bacteroid mRNA. cDNA was generated from vector control and *GmSAT1* RNAi RNA using random decamer primers. Expression values were normalized to the average expression of the housekeeping genes *SigA* and *GapA*. Data values represent the means of three independent technical replicates ±SD, as calculated by the ΔCT method (Livak and Schmittgen, 2001).

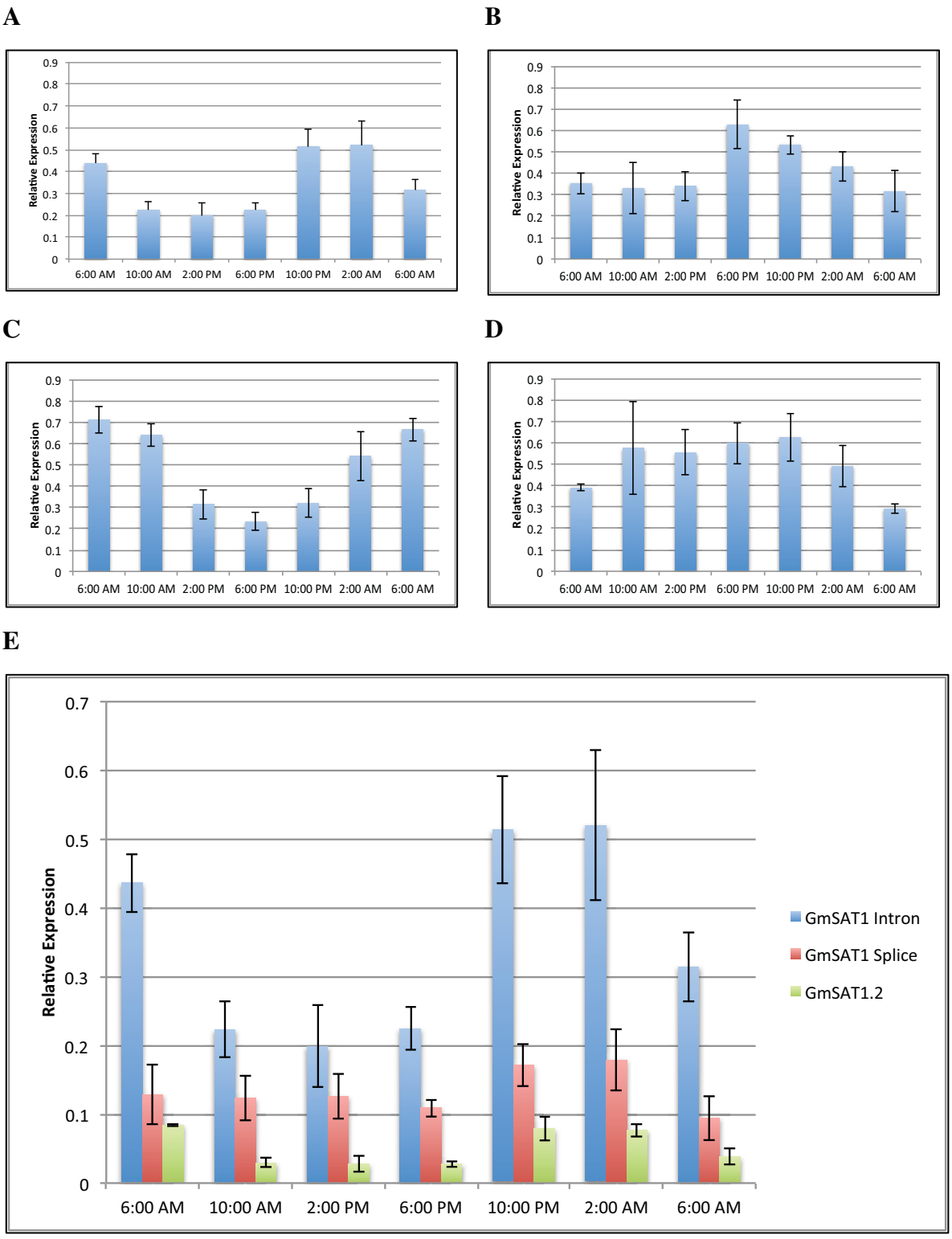


Figure 2.27. Diurnal expression of selected soybean transcripts.
24 day-old soybean nodules were harvested every 4 hours over a 24 hour period, starting at 6:00 am (sunrise/sunset, 5:57am/8:07pm). Tissue from six plants was pooled for RNA extraction and analyzed by qPCR. **A**, *GmSAT1*. **B**, *GI* (Glyma09g07240). **C**, *GmLHY* (Glyma16g01980). **D**, *GmMFS1.5* (Glyma09g33680). **E**, *GmSAT1* splice variants. Data values represent the means of three independent technical replicates ±SD relative to *cons6* (Libault et al., 2008), as calculated by the ΔCT method (Livak and Schmittgen, 2001). The expression value of *cons6* was stable over the 24-hour period.

Primer Name	Sequence
GmSAT1 pMAL F	CATATGCCTAAAATTGACAACAATGCTCTTG
GmSAT1 pMAL R	GATCCTTATCTTGCTTCAATTTCAGGTAGTG
pMAL Forward Seq	GGTCGTCAGACTGTCGATGAAGCC
pMAL Reverse Seq	TGTCCTACTCAGGAGAGCGTTCAC
YOR378W Promoter 3F	GGGTTGTCGCTACGAATGTT
YOR378W Promoter 1R	ATTGATTTTATGTTCTTTATGTTCCAG
PHO84 Promoter 2F	TATTACGCACGTTGGTGCTG
PHO84 Promoter 1R	TTGGATTGTATTCGTGGAGTTTT
GmSAT1 CDS F	ATGAGGAGTTCTCATATGGAGA
GmSAT1 CDS R	TCACACGAAATATGAAAAAGC
GmSAT1 T1 F	ATGGCTAACTTCCTCCATCAGTGG
GmSAT1 T2 F	ATGTCATTCTCTGATAACTCAAACTTC
GmSAT1 T3 F	ATGGCAAAACAGCTCGGTGATAAC
GmSAT1 T4 F	ATGTCAAATCTTCTTTCATTTGTTAATACG
GmSAT1 N1 F	ATGGAGATTTCATCAATCAGAGG
GmSAT1 N2 F	ATGGGGCTACCTGAACTGGGG
GmSAT1 N3 F	ATGGGGATAATAGAGGATCCTAACTT
GmSAT1.TruncCterm.R	TCATTCAGGTAGTGCCTCGACAAA
GmSAT1PromoterF	GATTTAACCTAAGAAAACCAATTCC
GmSAT1PromoterR	ATACTCAAACTACAACATCCCATG
GmSAT2 CDS F	ATGAGGACTTCTCATATGGAGATTTC
GmSAT2 CDS R	TCACACGAAATATGAAAAGGC
GmSAT2PromoterF	GAGAATTAGACACAAAAATCATATAAC
GmSAT2PromoterR	ATCTCAACTGAAGTGAGAACAATTC
GmSAT2PromoterR2	GAACAATTCTTTGGAACTTGCAT
GFP pYES3 F	ATGGTAGATCTGACTAGTAAAGGAG
GFP pYES3 R	TCAAGCTTTGTATAGTTCATCCATG
GmSAT1 qPCR F	TTATTGCTCAGATGGATATGGAA
GmSAT1 qPCR R	GACGACCGAAAAATGCATAAC
GmSAT1 Splice F	CCATGAATTTGGTGATATTAGCTTG
GmSAT1 5'UTR F	TGTACATGGGATGTTGTAGTTTG
GmSAT1 Intron F	ATGGCCATTCAGCAAACATTAC
GmSAT1.2.UTR.F	TCAGCTTCTGTGGTGACCAAT
Gm.SAT1.qPCR2.R	CTATTATCCCCAGTTCAGGTAG
GmSAT2 qPCR F	TTATTGCTCAGATGGATATGGAA
GmSAT2 qPCR R	TCGGAATGCATAACGAGTTTC
GmSAT2.1.5UTR.F	GCAAGTTCCAAAGAATTGTTCTC
GmSAT2.2.5UTR.F	TCAGCTTCTGTGGTGACCAAT
GmSAT2.pPCR2.R	TCTATTATCCCCATTTCAGGCAA
GmSAT2 5 Intron F	ATGTGGCTTTCCTTCTTCCAT
GmSAT2 5 Splice F	TGGGATGTTGTACTTTTTGATATAT
Brady SigA qPCR F	ACACCGGCTCGGAGCTCGAT
Brady SigA qPCR R	CGCAGCGAGCTGATGCACCT
Brady GapA qPCR F	GCGCCACCGACGTGAAGGAA
Brady GapA qPCR R	CGATCGAGACGTTCGGCGCA

Brady NifH qPCR F	TGCGTCGCATGACGGTGCTT
Brady NifH qPCR R	GGATGATGCCCTTGCCGCCA
Brady FixU qPCR F	ATGATCCGCCGCTCTCCGGA
Brady FixU qPCR R	CCAGCCGCCCCATAGGGTCT
SNF1 Kinase qPCR F	TGAGAAGTGAGTCAAGGAAGCA
SNF1 Kinase qPCR R	CAGGCAAAACATTCGCCTTA
GmLHY qPCR F	AAAGGCAACCTTTGGTTTTTG
GmLHY qPCR R	ACGACAAGTTTCGTGAGCTG
GmGI20g qPCR2 F	CCCACAACCCCTCTTATCAATAC
GmGI20g qPCR2 R	GGCAGCGGTATCAAGAAAGT
GmGI09g qPCR F	GAGTCCTAAGCCACTGCAAAAG
GmGI09g qPCR R	GCTTCCCATGTCAAGCAGTT
GmZIP1 qPCR2 F	TCTATTACAATAATGGGATTGTTCTTC
GmZIP1 qPCR2 R	AGGGCAGTTGGACTGTTTTC
cons6 qPCRF	AGATAGGGAAATGGTGCAGGT
cons6 qPCRR	CTAATGGCAATTGCAGCTCTC
StrepHA GmSAT1 F	ATGTGGAGCCACCCGCAGTTCGAAAAAGCTAGCCCA TACCCATACGATGTTCCAGATTACGCTAGCGCTAGG AGTTCTCATATGGAGATTTCA
GmSAT1 Signal Peptide F	ATGTGGCTTTCTTTCTTCCATGAATTTGGTGATATTA GCTTGATAGCTTCAGCTGTTTTTCATTGCATGGTAGA TCTGACTAGTAAAGGAG
GmSAT2 Signal peptide F	ATGATTGATTTTAATGCATGTACATGGGATGTTGTAC TTTTTGATATATTAATTAGGTTGATAGCTTCAGCTGT TTTTCATTTCATGGTAGATCTGACTAGTAAAGGAG
MtN25 Signal Peptide F	ATGGTTTATTCAAATTCATATATGTTCCTTGGTCTTG GTGTTTTTGTTCTTCTCTCTTCTCATGTTTTGGCTTAT AATATGGTAGATCTGACTAGTAAAGGAG
GmSAT1 L17M F	ATGAGGAGTTCTCATATGGAGATTTCATCAATCAGA GGGCTACCTGAA**A**TGGGGATAATAGAGG
GmSAT1 G260D F (added silent AgeI site)	CTTCTTCAGAGAC**C**GGTG**A**TACATTTGTCGAGGC
GmSAT1 G260D R	GCCTCGACAAATGTATCACCGGTCTCTGAAGAAG
GmSAT1 V263D F (added silent AgeI site)	CGAAGACTCTTCTTCAGAGAC**C**GGTGGTACATTTG**A** **T**GAGGCACTACCTGAAATTGAAGC
GmSAT1 V263D R	GCTTCAATTTCAGGTAGTGCCTCATCAAATGTACCA CCGGTCTCTGAAGAAGAGTCTTCG
GmSAT1 W272Y F (added silent EcoRI site)	CCTGAAATTGAAGCAAGATTTT**AT**GAAAGAAATGTC CTCATAAGAAT**T**CATTGTGAGAAG
GmSAT1 W272Y R	CTTCTCACAATGAATTCTTATGAGGACATTTCTTTCA TAAAATCTTGCTTCAATTTCAGG

Table 2.5. List of primers used in this chapter.

2.3 Materials and Methods

2.3.1 Expression and purification of GmSAT1

A truncated version of GmSAT1 (amino acids 128-270) was amplified (GmSAT1 pMAL F and GmSAT1 pMAL R primers) with Phusion DNA polymerase (Finnzymes) and inserted into CloneJET (Fermentas). The fragment was then excised with *Nde*I and *Bam*HI and ligated into pMAL-c5X (New England Biolabs) to create a N-terminal fusion to the maltose-binding protein (MBP). The fusion was sequenced using primers pMAL Forward Seq and pMAL Reverse Seq. The maltose binding protein was expressed by using the empty pMAL-c5X plasmid. The plasmids were transformed into the *E. coli* strain NEB Express. MBP-GmSAT1$_{128\text{-}270}$ and MBP alone were expressed and purified according to the manufacturers instructions using amylose resin (New England Biolabs).

2.3.2 EMSA

Fragments of the *YOR378W* (-284 to 0, called Y3) and *PHO84* (-594 to 0, called P3) promoters were amplified (YOR378W Promoter 3F + 1R and PHO84 Promoter 2F + 1R primers) and ligated into pGEM-T Easy (Promega). The promoter fragment was then excised with *Not*I and end-filled with ^{32}P dCTP (3000 ci/mmol, Easytide 5' Triphosphate γ-^{32}P, PerkinElmer) using the Klenow fragment (3'->5' exo-, New England Biolabs) in a 50 µl reaction (1 µl ^{32}P-dCTP, 1 µl 1mM dGTP, 1 µl Klenow, 400 ng promoter fragment). Labeled probe was desalted with S-200 HR Microspin Columns (GE Healthcare). 1 µg of MBP-GmSAT1$_{128\text{-}270}$ or MBP alone was incubated with the labeled probe in a 20 µl reaction at room temperature for 25 mins. The reaction contained 15 % (v/v) glycerol, 20 mM Tris pH 8.0, 175 mM NaCl, 5 mM EDTA, 50 mM KCl, 0.5 mM DTT, 0.5 µl Poly [d(I-C)] (1µg/ul) and 4 µl of probe. The reaction was then loaded on a 6% (w/v) native tris-glycine polyacrylamide gel containing 10% (v/v) glycerol. After separation, the gel was vacuum dried and exposed to film for 24 hours at -20°C.

2.3.3 Generation of clones for Yeast Complementation

GmSAT1 (primers GmSAT1 CDS F and GmSAT1 CDS R), T1-/T2-/T3-/T4-*GmSAT1* (primers T1-/T2-/T3-/T4-GmSAT1 F and GmSAT1 CDS R), N1-/N2-/N3-*GmSAT1* (primers N1-/N2-/N3-GmSAT1 F and GmSAT1 CDS R), Strep-HA tag (StrepHA GmSAT1 F and GmSAT1 CDS R) were amplified by PCR using platinum taq high fidelity (Invitrogen) and inserted into pCR8-GW-TOPO (Invitrogen). *GmSAT2* was cloned by nested PCR (Round 1: primers GmSAT2.2 5UTR F and GmSAT2 qPCR R, Round 2: primers GmSAT2 CDS F and GmSAT2 CDS R). Clones were then sequenced with GW1 and GW2 primers and recombined into pYES3-DEST using LR Clonase II (Invitrogen).

To create the *Bam*HI swap, *GmSAT1* and *GmSAT2* pYES3-DEST were digested with *Bam*HI. For the *Hin*dIII and *Spe*I swaps, *GmSAT1* and *GmSAT2* pCR8 was digested with either *Hin*dIII/*Xba*I or *Spe*I/*Xho*I. The fragments were then purified from agarose gels. The vector backbones were dephosphorylated, ligated with the desired insert using T4 DNA ligase (New England Biolabs), transformed into DH5α, and the resulting clones were sequenced.

For mutagenesis, substitutions were made by PCR with Phusion DNA polymerase, using *GmSAT1* pCR8 as a template. GmSAT1 S3T and L17M were generated with the primers GmSAT2 CDS F and GmSAT1 L17M F respectively, in combination with GmSAT1 CDS R. The three other mutations were made via site-directed mutagenesis similar to the strategy for QuikChange (Stratagene). GmSAT1 G260D (primers GmSAT1 G260D F+R) and GmSAT1 W272Y (primers GmSAT1 W272Y F+R) were generated by PCR (initial denaturation of 98° for 30 sec, then 18 cycles of 98° for 10 sec, 60° for 1 min, 72°C for 4 min) using 10 ng of template in a 50 µl reaction. GmSAT1 V263D (primers GmSAT1 V263D F+R) was generated in a similar manner, except that a two-step PCR cycle was used (20 cycles of 98°C for 20 sec, then 72°C for 4.5 min). 1 µl of DpnI was then added to the PCR reaction and incubated at 37°C for 1 hour. 5 µl of the DpnI treated PCR reaction was used to transform *E. coli* strain DH5α. Eight colonies were picked for each transformation and subsequently grown for plasmid preparation. Isolated plasmids were then test digested for an added restriction enzyme site (G260D and V263D contained an

added *Age*I site, while a *Eco*RI site was added to W272Y). The new clones were then recombined into pYES3-DEST using LR clonase.

All constructs were transformed by lithium acetate/PEG (Gietz and Schiestl, 2007) into the yeast strain 26972c. Positive transformants were selected on YNB/Proline/Glucose (1.7 g L^{-1} yeast nitrogen base without amino acids and ammonium sulfate, 0.1% (w/v) proline, 2% (w/v) glucose) and restreaked onto the same media. A single colony was then picked and grown in YNB/Proline/Glucose for two days, washed, and diluted to an OD600 of 0.5 in water. Three serial 10-fold dilutions were then made and 5 µl of each dilution was spotted onto Grenson's media (with 2% (w/v) galactose) plates containing either 0.1% (w/v) proline, 0.1% (w/v) proline plus 100 mM methylamine, or 1 mM NH_4Cl. Plates were incubated for 4 days at 28°C, then photographed with a GelDoc imager (Bio-Rad).

2.3.4 Western Blots

For western blots, 26972c containing either GFP:GmSAT1, GmSAT1-pYES3 or empty pYES3 were grown in YNB with 2% (w/v) glucose. Cells were then pelleted, washed in water and diluted to an OD_{600} of 0.4 in YNB containing 2% (w/v) galactose. Cells were grown for 24 hours, pelleted and total protein was extracted according to (Kushnirov, 2000). 20 µg of total protein was separated by 12% SDS-PAGE and transferred to nitrocellulose for blotting with a SNAPid system (Millipore). The blot was blocked with 1% (w/v) BSA in TBST (50 mM Tris, 150 mM NaCl, 0.1% (v/v) Tween 20, ph 7.5) and all washing steps were conducted with TBST. Rabbit anti-GFP antibody (Cell Signaling) was used at a dilution of 1:3000, followed by 1:6000 secondary anti-rabbit conjugated to peroxidase (Sigma) in TBST with 1% (w/v) BSA. Signals were detected by chemiluminescence according to Haan and Behrmann (2007).

2.3.5 Isolation of *GmSAT1* and *GmSAT2* Splice Variants

RT-PCR was carried out on cDNA initially using GmSAT1 5'UTR F and GmSAT1 CDS R primers to identify possible slice variants. The two bands were excised and ligated into pGEM-T-easy for DNA sequencing. Using the sequencing results and available ESTs in Genbank, three primer sets (GmSAT1.2, GmSAT1 Splice, GmSAT1 Intron) were designed

to amplify each variant in conjunction with the GmSAT1 qPCR2 R primer. All amplicons were subsequently cloned and verified by DNA sequencing. The exact same methodology was used to identify three variants of the GmSAT2 transcript (primers GmSAT2.2 5UTR F, GmSAT2 5 Intron F, GmSAT2 5 Splice F, and GmSAT2 qPCR2 R). qPCR was then carried out on 36-day old soybean tissue with SYBR Green (Bio-Rad) in a Light Cycler (Bio-Rad) qPCR machine. After initial denaturation, the reaction was run for 36 cycles (96°C for 10sec, 55°C for 5 sec, 72°C for 20 sec). The mean C_T value of three replicates was used to generate a ΔC_T value using *cons6* as an internal control (Libault et al., 2008). The relative expression was then calculated by using the formula $2^{-\Delta CT}$ (Livak and Schmittgen, 2001).

2.3.6 Cloning and Expression of *GmSAT1* and *GmSAT2* Promoters

Using the available soybean genomic sequence from Phytozome (http://www.phytozome.net), primers were designed to clone a DNA fragment upstream of the start codons of *GmSAT1* and *GmSAT2*. The *GmSAT1* promoter (1926 bp) was cloned with GmSAT1 Promoter F + R primers, and the *GmSAT2* (1989 bp) promoter was cloned with GmSAT2 Promoter F + R2 using platinum taq high-fidelity (Invitrogen). The fragments were ligated into pCR8-GW-TOPO (Invitrogen) and recombined into the destination vector pKGWFS7. The plasmids were then transformed into *Agrobacterium rhizogenes* K599 (selection on spectinomycin 150 μg/ml due to partial resistance by K599) for hairy root production. Hairy roots were generated according to Kereszt et al. (Kereszt et al., 2007). Hairy roots were fixed in 90% acetone on ice for 15 minutes, partially sectioned with a razor blade, then incubated in GUS staining buffer (0.1 M sodium phosphate buffer pH 7, 3% (w/v) sucrose, 0.5 mM EDTA, 0.5 mM potassium ferrocyanide, 0.5 mM potassium ferricyanide, 0.05% (w/v) X-Gluc). Nodules were then fixed in 5% (v/v) glutaraldehyde and embedded in Technovit 7100 (Heraeus Kulzer). 8 μm sections were cut on a Leica RM2265 microtome, dried onto glass slides, stained with 0.05% (w/v) ruthenium red for 30 min, and embedded by adding DPX (Sigma) onto the slide with a coverslip. Images were obtained with a Leica ASLMD laser-assisted micro-dissection microscope.

2.3.7 Onion Transformations

GFP:GmSAT1 (primers GFP pYES3 F and GmSAT1 CDS R), GmSAT1:GFP (primers GmSAT1 CDS F and GFP-pYES3 R), S1-GFP (primers GmSAT1 Signal Peptide F and GmSAT1 CDS R), S2-GFP (GmSAT2 Signal Peptide F and GmSAT1 CDS R), N25-GFP (Nodulin-25 Signal Peptide F and GmSAT1 CDS R), and GFP-Alone (primers GFP pYES3 F and GFP-pYES3 R) were amplified from a GFP:GmSAT1-pYES3 template. T1-GmSAT1 was created by amplifying the coding sequence of GmSAT1 without the N-terminal 21 amino acids (primer GmSAT1 T1 F) from a GmSAT1-pCR8 template. All fragments were inserted into pCR8-GW-TOPO (Invitrogen). GFP:GmSAT1, GmSAT1:GFP, S1-GFP, S2-GFP, N25-GFP, and GFP-Alone were recombined into the destination vector pK7WG2D and T1-GmSAT1 was recombined into pP7WGF2 using LR Clonase II (Invitrogen). All plasmids were then digested with *Sac*I and *Xba*I and the resulting fragment (including a 35S promoter and terminator) was ligated into pBluescript.

5 μg of purified plasmid DNA (in 10 μl) was combined with a 50 μl suspension (1.5 mg of 0.6 μm gold macrocarriers (Bio-Rad) in 50% (v/v) glycerol). Onion epidermal peels were bombarded with a PDS-1000/He particle delivery system (Bio-Rad) using a 1100 psi Rupture Disc. Epidermal peels were maintained on Murashige and Skoog basal medium (with vitamins, Austratec) supplemented with 9 g L^{-1} TC grade agar (Austratec), 120 g L^{-1} sucrose and 500 mg L^{-1} tryptone. After bombardment, the peels were incubated in the dark for 24 hours at room temperature before viewing. Images were obtained using a Zeiss LSM 5 Pascal confocal microscope. GFP fluorescence was monitored by excitation at 488 nm with an argon laser combined with a 505-520 nm bandpass filter.

2.3.8 Soybean Microarray

Nodule tissue from *GmSAT1* RNAi and empty vector control hairy roots (Kereszt et al., 2007) were collected 24 days post inoculation with *Bradyrhizobium japonicum* USDA110. The construct used for silencing contained a 359 bp portion of the *GmSAT1* 3'UTR (Loughlin, 2007) inserted into the plasmid pK7GWIWG2D(II) (Karimi et al., 2002). Plant were grown in Waikerie sand and watered with nitrogen-free Herridge's media (Herridge, 1982). Total RNA was extracted from individual plant nodules with a Spectrum Plant

Total RNA kit (Sigma-Aldrich). RNA was treated with Turbo DNase (Ambion), ethanol precipitated, and resuspended in water. 1 μg of total RNA was then used for cDNA synthesis with Superscript III (Invitrogen). Samples showing similar *GmSAT1* RNA levels by qPCR (normalised to *cons6*) were then chosen for further analysis. RNA was checked by a bioanalyzer before proceeding with labeling. 100 ng of total RNA from four replicates was then biotin-labeled with an Applause WT-Amp ST System (NuGEN) and hybridized to the Soybean 1.0 ST array (Whole Transcript, Affymetrix) according to the manufacturers protocol at the Ramiciotti Centre (The University of New South Wales, Sydney). Raw .CEL files were then analyzed using the Partek Genomics Suite using default import settings. Subsequently, expression of a selection of genes from the array was analyzed by qPCR (as outlined in 2.3.5) with primers from Table 2.5 to verify differential regulation. Fold changes were calculated by the $2^{-\Delta\Delta CT}$ method (Schmittgen and Livak, 2008).

For qPCR of bacterial transcripts, RNA from the same pool used for the microarrays was reverse-transcribed with Superscript III (Invitrogen) as above using random decamer primers. *FixU* and *NifH* transcripts were normalized to the average expression of the control genes, *GapA* and *SigA* (primers included in Table 2.5), as outlined in 2.3.5.

2.3.9 Soybean Diurnal Expression

Soybean plants were grown in river sand and inoculated with *Bradyrhizobium japonicum* USDA110 upon planting. Twenty-four days after sowing, tissues (nodule, root, leaf) were collected at 6:00 am and every four hours until 6:00 am the following day. The plants were harvested from November 24-25, 2011 (sunrise/sunset, 5:57am/8:07pm). Total RNA was extracted from pooled plant nodules (from a single pot) with a Spectrum Plant Total RNA kit (Sigma-Aldrich). RNA was treated with Turbo DNase (Ambion) and reverse transcribed with Superscript III (Invitrogen) using and oligo-dt20 primer. *GmSAT1*, *GI*, *GmMFS1.5*, and *LHY* expression was determined by qPCR, relative to *cons6* (as described in section 2.3.5).

2.4 Discussion

2.4.1 GmSAT1 is a *bona fide* transcription factor

Work by Loughlin (2007) had established that GmSAT1 was likely a transcription factor. GmSAT1 was observed in the nucleus of yeast and mutations to the DNA-binding domain of GmSAT1 disrupted its activity. However, direct evidence for DNA binding was lacking. By purifying a soluble truncated version of GmSAT1 fused to MBP, this study demonstrates that GmSAT1 is able to bind DNA *in vitro*. As the probes used for the binding assay were from the promoters of yeast genes upregulated in the presence of GmSAT1 (Mazukiwicz and Loughlin, unpublished results), it is likely that GmSAT1 binds to these promoters *in vivo*. To identify all the GmSAT1-occupied genomic sites, chromatin immunoprecipitation (ChIP) experiments would need to be carried out. ChIP experiments require an antibody to the transcription factor of interest, which should identify a single band by western blot. Attempts where made here to identify a suitable tag, since the anti-GmSAT1 serum was not deemed suitable for ChIP experiments. Due to cleavage and interference with GmSAT1 activity, N-terminal tags are not recommended. Since C-terminal tags also interfere with GmSAT1 activity (Loughlin, 2007), generating a new polyclonal antibody to a GmSAT1 unique peptide would be ideal.

2.4.2 Residues contributing to GmSAT1 activity in yeast

It was observed that GmSAT2 was able to complement the ammonium phenotype of 26972c, but not the MA uptake. Therefore the residues responsible for the MA activity were investigated by swapping domains of GmSAT2 into GmSAT1. It was discovered that a few key residues were contributing to the difference in activity. Specifically, the combination of the S3T and L17M mutations (*GmSAT2-BamHI-GmSAT1* construct) led to a loss of MA uptake. Further, the combination of Q245R, G260D, V263D, and W272Y mutations (*GmSAT1-HindIII-GmSAT2* construct) also led to loss in activity. Since Loughlin (2007) identified a putative site-1 protease cleavage site at position 274-277, these nearby mutations (Q245R, G260D, V263D, and W272Y) may influence processing at this site. The effect of the S3T and L17M mutations remains to be determined. Attempts were made to assign a role to the N-terminus of GmSAT1 by truncations and

mutagenesis, however the results were ambiguous. Since the GFP-T1-GmSAT1 construct was able to enter the nucleus of onions intact and processing of GFP-GmSAT1 was observed in yeast, it is intriguing to speculate that a cleavage event occurs in the N-terminus of GmSAT1. This cleavage may be required for nuclear entry, before or after the C-terminal transmembrane domain is removed. Fusing the GmSAT2-BamHI-GmSAT1 and GmSAT1-HindIII-GmSAT2 proteins to GFP for localization in onion epidermal cells would be recommended in the future.

2.4.3 *GmSAT2* is a partially redundant version of *GmSAT1*

Having located and cloned *GmSAT2*, and noting the high degree of similarity to *GmSAT1*, experiments were conducted to determine if it might perform a redundant role *in planta*. Promoter expression analysis of *GmSAT1* and *GmSAT2* revealed essentially identical localization. Both promoters were active in the root stele, as well in the inner cortex and uninfected cells of the nodule. *GmSAT1*, in contrast to *GmSAT2*, was also expressed in infected cells. qPCR did reveal some differences, though, between *GmSAT1* and *GmSAT2*. *GmSAT1* was nodule-enriched relative to roots, as well as being 5-fold higher in nodules relative to *GmSAT2*. Additionally, *GmSAT2* was expressed at higher level in lateral roots, relative to nodules. Therefore, it would seem that GmSAT1 could be the dominant protein in nodules. The RNAi experiment successfully downregulated both genes, and therefore we cannot yet say with certainty what the role of GmSAT2 is *in planta*. Based on the yeast transcriptional assay for MA uptake, it could be that *GmSAT2* is a relic of the soybean genome duplication and has since become inactive as a transcription factor. Selectively knocking out each gene would be necessary to determine their individual activities (via EMS or insertional mutagenesis).

2.4.4 Localisation of GmSAT1 *in planta*

Previously, GmSAT1 was shown to localize to the soybean SM and nucleus using GmSAT1 antiserum (Kaiser et al., 1998; Loughlin, 2007). Additionally, GFP-GmSAT1 was observed in vesicles and the nucleus in yeast (Loughlin, 2007). Further investigations in this study found that GFP-GmSAT1 localized to ER vesicles in onion epidermal cells,

but not nuclei. Interestingly, GmSAT1-GFP (C-terminal) was also observed in vesicles, which appeared to be of a slightly different size and shape (although the exact origin was not determined). Having found that the N-terminus of GmSAT1 may be cleaved in yeast, a truncated version of GmSAT1 (T1) was fused to GFP. Since this fusion was observed in nuclei, it appears that the N-terminus of GmSAT1 is also cleaved *in planta*. The data point to a cleavage event that seems to be necessary for nuclear entry, as N-terminal mutations in yeast to this region disrupted GmSAT1 transcriptional activity in yeast. Therefore, it could be hypothesized that some of the vesicular GFP signal belongs to free GFP that has remained after N-terminal cleavage of GmSAT1, since cytoplasmic GFP was not observed. If this were true, then the N-terminus of GmSAT1 would be associated with vesicle membranes, and potentially the C-terminus as well (since C-terminal GFP fusion also revealed GFP associated with vesicles). Therefore, in order to be released to the nucleus, GmSAT1 may need to be cleaved sequentially, similar to SREBP and ATF6 (Brown and Goldstein, 1999; Ye et al., 2000). In these two cases, the membrane-bound transcription factors are first cleaved by a site-1 protease, then a site-2 protease. Previously, Loughlin (2007) identified a site-1 protease site (RXXL) in the C-terminus of GmSAT1. Therefore, a yet unidentified site-2 protease site could be present in the N-terminus of GmSAT1.

2.4.5 *GmSAT1* is linked to the circadian clock

The plant circadian clock is a biological cycle that is entrained by day-night cycles and has a period of approximately 24 hours. The clock is maintained by morning, core, and evening interlocking feedback loops. The core loop so far includes the MYB transcription factors CCA1 (Wang and Tobin, 1998), and LHY (Schaffer et al., 1998), which negatively regulate the expression of the pseudo-response regulator TOC1. TOC1, which is expressed at dusk, in turn negatively regulates CCA1 and LHY by directly binding to their promoters (Gendron et al., 2012). The morning loop is activated by CCA1/LHY, which promote the expression of *PRR9* and *PRR7* (Locke et al., 2006; Zeilinger et al., 2006). PRR9 and PRR7, in turn, repress the expression of *CCA1* and *LHY*. The evening loop involves GI and ZTL, which inhibit the accumulation of TOC1. GI stabilizes ZTL (Kim et al., 2007), which then targets proteasome-dependent degradation of TOC1 and PRR5 (Somers et al., 2000; Kim et al., 2003). Removing any of these clock components does not induce oscillatory arrest, however three other proteins (ELF3, ELF4, and LUX) are

essential for sustained circadian function (Herrero and Davis, 2012). The phenotypes of the loss-of-function alleles are similar, and it now believed that these three proteins form a repressor complex, controlling the expression of key clock genes (Nusinow et al., 2011).

Of the ninety-five genes identified as being downregulated in the *GmSAT1* RNAi nodules, twelve are associated with the timing of the circadian clock. Included are such genes as *GIGANTEA (GI), PSEUDO-RESPONSE REGULATORs (PRR5 and PRR7), COLD, CIRCADIAN RHYTHM, AND RNA BINDING 2 (CCR2), NUCLEAR FACTOR Y SUBUNIT A3 (NF-YA3), COLD REGULATED GENE 27 (COR27),* and *JUMONJI C DOMAIN-CONTAINING PROTEIN 5 (JMJD5)* (Park et al., 1999; Chen et al., 2007; Covington et al., 2008; Streitner et al., 2008; Mikkelsen and Thomashow, 2009; Lu et al., 2011).

The misexpression of members of the morning (*PRR5* and *PRR7*) and evening (*GI*) components of the circadian clock could explain the observed phenotype of the *GmSAT1* RNAi nodules. Loss of *GI* activity has pleiotropic effects, altering such process as flowering time (Park et al., 1999), response to sucrose (Dalchau et al., 2011), and hypocotyl growth (Huq et al., 2000). *GI* has been shown to be functionally equivalent in soybean, where a mutation in *GI* (*GmGIa,* Glyma10g36600) leads to early flowering (Watanabe et al., 2011). Interestingly, the soybean genome contains three GI genes, but *GmGIa* was not significantly downregulated in *GmSAT1* RNAi (-1.2 fold). On the other hand, *GmGIb* (Glyma20g30980, from the same study) and Glyma09g07240 (this study), which are 79% and 86% identical to *GmGIa*, showed a reduction upon silencing of *GmSAT1*. According to the SoySeq database (http://www.soybase.org/soyseq/#), all three genes show a similar expression pattern, being present in all soybean tissues (not shown).

Although the vast majority of studies have focused on aerial organs, James et al. (2008) focused on the circadian clock in roots. The authors found that most of the components (including *GI, PRR5,* and *PRR7*) oscillated in roots like shoots. Once transferred to constant light, members of the morning-phased loop (*CCA1*, *LHY*, *PRR7*, and *PRR9*) remained rhythmic, however the core and evening loops became uncoupled. This study concluded that synchronizing the clock between the shoot and root depends on photosynthesis, perhaps via the supply of sucrose.

The transcripts for *GmSAT1, GI, and LHY* were assayed over a twenty-four hour period in soybean nodules. *GmSAT1*showed a maximum (2.5 fold increase) at 10 pm and 2 am, *GI* had a maximum at 6pm (2 fold increase), and *LHY* showed increased expression at 6 am (3 fold increase). Notably, the transcriptional patterns of *GI* and *LHY* are greatly dampened in nodules relative to known patterns in *Arabidopsis* leaves (up to 20 fold changes in transcript levels), similar to the findings in *Arabidopsis* roots (James et al., 2008). Sullivan et al. (2004) previously monitored the expression of *GmPPCK4* (PEPc Kinase 4) in soybean and found that is was expressed rhythmically in leaves, but not in roots. They used *LHY* as a positive control (same as this study), and also found a relative peak in expression in the early morning. Thus, it would appear that the circadian clock is operating in nodules, but at a dampened level. No studies to date have monitored the circadian clock in nitrogen fixing nodules.

Yazdanbakhsh and co-authors (2011) investigated the circadian control of root elongation and carbon partitioning. They found that in the *cca1/lhy* mutant there was a premature depletion of sugars, leading to a decline in root elongation. Therefore, in the *GmSAT1* RNAi nodules, there could be a disruption of the sucrose-sensing network. Downregulation of *GI, PRR5,* and *PRR7* (and the remaining nine genes) could disrupt the circadian clock, thus affecting proper timing of carbon utilization, and therefore nodule growth. Monitoring the diurnal expression of key genes from the circadian clock in *GmSAT1* RNAi nodules versus wild type would likely shed light on this situation. Fukushima et al. (2009) demonstrated that in the triple mutant *prr9/7/5* there was a significant decrease in the metabolites required for the TCA cycle. Thus, assaying key carbon input metabolites in *GmSAT1* RNAi nodules would also be desirable.

2.4.6 Is nitrogen export disrupted in *GmSAT1* RNAi nodules?

The microarray also identified *Uricase* II (Nodulin-35, Glyma20g17440), NADH-dependent glutamine synthetase (*GLT1*) and a newly identified major facilitator protein (*GmMFS1.5*, chapter 4 this study) as being downregulated after silencing of *GmSAT1*. As *GmSAT1* RNAi leaves do not appear to receive adequate nitrogen (Loughlin, 2007), combined with the fact that *B. japonicum* nitrogen fixation gene expression appears

unaffected, the phenotype could be explained by an inability to either incorporate the fixed nitrogen or export it from the nodule. Further, *GmSAT1* promoter activity is localized in uninfected cells and the inner cortex, both sites of ureide synthesis (Newcomb et al., 1989).

A recent study investigated the function of two nodule ureide permeases in soybean (Collier and Tegeder, 2012). The transporters (GmUPS1-1 and GmUPS 1-2) were localized to the plasma membrane and were expressed in the nodule inner cortex and vascular bundles. The authors then silenced the transporters by RNAi and observed a very similar phenotype to *GmSAT1* RNAi. The silenced plants contained similar nodule numbers relative to controls, however the nodules remained small and infected cells were not properly developed. Further, the level of ureides was significantly increased in the silenced nodules, with reduced partitioning to the leaves. Interestingly, the transcript for *uricase* (called *UR9*) was significantly downregulated. Uricase is localized to uninfected cell peroxisomes, and is responsible for the conversion of uric acid to allantoin (Bergmann et al., 1983; Nguyen et al., 1985). The results of Collier and Tegeder (2012) indicate that excessive accumulation of ureides feeback-inhibit nitrogen fixation and nodule development.

In the *GmSAT1* RNAi nodules, downregulation of *uricase* may indicate the beginning of a feedback loop, where ureide (or a precurser, such as glutamine) levels are accumulating. As the control of many metabolic enzymes occurs post-transcriptionally, there could also be changes, such as phosphorylation, that are missed by transcriptomics. The downregulation of *GmMFS1.5*, which is closely related to *GmMFS1.3*, may hint at the function of this protein. *GmMFS1.3* (chapter 4) is expressed in the nodule inner cortex and is able to transport ammonium and methylammonium in yeast. However, the level of transport would indicate that these compounds are not the preferred substrates. Based on the location of the MFS transporters, it is possible that they are involved in transporting ureides or related compounds from the infected region into vascular tissues to be loaded into the xylem. To determine if there is an imbalance of end products in *GmSAT1* RNAi nodules, metabalomics should be carried out to monitor the levels of compounds such as ureides and ureide precursors.

3 Identification and Cloning of *MtSAT1* and *MtSAT2*

3.1 Introduction

3.1.1 *Medicago truncatula* as a model Species

Medicago truncatula belongs to the Hologalegina subclade of Papilionoids, which split from the Milletioid/Phaseoloid subclade (soybeans, Figure 3.1) approximately 54 million years ago (Lavin et al., 2005). Since *Medicago* is dipoloid, self-fertilizing, and has a relatively quick regeneration time, it was identified as a potential model legume (Barker et al., 1990). Additionally, *Medicago* can be readily transformed by both *Agrobacterium rhizogenes* and *Argobacterium tumefaciens* (Chabaud et al., 1996; Boisson-Dernier et al., 2001). The nodules of *Medicago* are indeterminate and contain distinct zones (meristem, infection, fixation, senescence) from the apex to the base (Vasse et al., 1990). Therefore, the distinct zonation can aid in identifying the effects of mutation (or knock-down) relative to a determinate nodule.

In the past decade, a large number of resources have been directed towards *Medicago* to construct insertional mutant lines, generate a transcriptomic atlas, and sequence the genome (Benedito et al., 2008; Tadege et al., 2008; Young et al., 2011). The *Medicago* genome contains eight chromosomes with 62,388 assigned gene loci, however pseudomolecules and unassigned BACs are still being mapped by next-generation sequencing (Young et al., 2011). Because of the relatively high rate of local gene duplications, *Medicago* contains nearly as many genes as soybean (65,781), which underwent a whole genome duplication event 13 million years ago (Schmutz et al., 2010). Even with the local duplications, mutations in single genes have yielded many nodulation phenotypes (Popp and Ott, 2011). Due to the increasing number of resources available, it was decided to clone and characterize orthologues of *GmSAT1* in *M. truncatula*.

3.2 Results

3.2.1 Sequences of *MtSAT1* and *MtSAT2*

As the *M. truncatula* genome was not fully annotated when this research was initiated, many gene models were incomplete. This was the case for two sequences found to be most similar to *GmSAT1*, called *MtSAT1* (Medtr2g010480) and *MtSAT2* (Medtr4g121940). Therefore, 5' and 3' RACE (rapid amplification of cDNA ends) was conducted to determine the sequences of the *MtSAT1* and *MtSAT2* mRNAs. *MtSAT1* (Figure 3.2) and *MtSAT2* (Figure 3.3) were determined to have open reading frames encoding for 353 and 310 amino acids, respectively. Compared to GmSAT1, the proteins are of a similar size and contain a bHLH region and a potential transmembrane domain (Figure 3.4A). Relative to each other, MtSAT1 and MtSAT2 share 53.0% identity (Figure 3.4B). MtSAT1 is 69.8% identical to GmSAT1, while MtSAT2 is 60.4% identical. However, the high degree of homology dramatically decreases when these proteins are compared to non-legume SAT proteins. Relative to MtSAT1, the *Arabidopsis* SATs are 28-33% identical. The proteins are most similar from the bHLH region to the C-terminus, with the N-terminal region being most variable. Of note, the four *Arabidopsis* proteins all contain a potentially helix-breaking/bending proline residue(s) in the transmembrane domain (Figure 3.4C). The phylogenetic tree shown in Figure 3.5 provides an overview of these proteins.

3.2.2 Expression of *MtSAT1* and *MtSAT2*

To further investigate the two genes, qPCR analysis was carried out on *Medicago* tissues harvested over a period of three weeks. The plants were sown in perlite and starved of external nitrogen for five days after transfer from agar plates. At day zero, the plants were inoculated with *S. meliloti* 2011. Figure 3.6 shows the results of the qPCR tissue time course. At day zero and day four, *MtSAT1* showed a four-fold higher level of expression in roots relative to *MtSAT2*. Between days four and seven, the expression level of *MtSAT2* increased by five-fold in roots, while *MtSAT1* levels remained stable. By day fourteen, both genes showed a decrease in root expression, a pattern that remained until day twenty-one. The expression of *MtSAT2* was enhanced in nodules, relative to roots, while *MtSAT1* showed a similar level in both tissues.

3.2.3 Tissue-specific expression of *MtSAT1* and *MtSAT2*

The promoters of both *MtSAT1* and *MtSAT2* were both cloned and inserted in the plasmid p243-RedRoot-GUS (Limpens et al., 2005) for localization studies. The *MtSAT1* (Figure 3.7) and *MtSAT2* (Figure 3.8) promoters were active in the stele of non-inoculated roots. Notably, the expression pattern of *MtSAT2* showed a change in spatial expression in response to rhizobia (Figure 3.8B). Both promoters showed activity in early nodule development, eventually becoming restricted to the inner cortical cells. In a fourteen day-old nodule, both promoters were localized to the inner cortex, with *MtSAT1* also being expressed in infected cells of the fixation zone (Figure 3.7D). At twenty-one days after inoculation, *MtSAT1* was expressed in the inner cortex, meristem, and infected cells, while *MtSAT2* seemed to be restricted to the inner and outer cortex, as well as the meristematic region.

3.2.4 Subcellular localization of MtSAT1 and MtSAT2

To monitor subcellular localization, GFP was fused to the N-terminus of both MtSAT1 and MtSAT2 under the control of the *ubiquitin3* promoter (Figure 3.9A). *The ubiquitin3* promoter is generally active in all cell types at a high level (Limpens et al., 2009). Looking at root tips, GFP:MtSAT1 was observed in nuclei and small vesicles (Figure 3.10A and B). Similarly, GFP:MtSAT2 was also found associated with small vesicles throughout the cell and the nucleus (Figure 3.10D).

3.2.5 Effect of silencing *MtSAT1* and *MtSAT2*

As silencing of *GmSAT1* is known to impair nodule development, *MtSAT1* and *MtSAT2* were targeted for knock-down experiments. A 253 bp portion of the *MtSAT1* C-terminus was cloned and inserted into p277-RedRoot-RNAi (Figure 3.9B). This region was chosen because it is unique to *MtSAT1* and *MtSAT2*, and shares 80.7% identity at the DNA level (70.2% amino acid identity) to *MtSAT2*. After selecting positive transgenic roots, nodules were chosen and embedded for microscopy. Overall, the number of nodules was similar

for the *MtSAT1/2* RNAi compared to empty vector controls. However, there seemed to be issues related to nodule development (Figure 3.11). Of the silenced nodules analyzed, the majority were impaired in structure. At early stages, a subset was aborted while others were disorganized and showed an increased number of cells undergoing senescence. The other older affected nodules showed a rounded base similar to a fourteen-day old nodule, but contained what appeared to be new meristematic growth on top. qPCR was carried out on *MtSAT1/2* RNAi nodules and found *MtSAT1* and *MtSAT2* to be downregulated -2.9 and -1.5 fold, respectively. The degree of knockdown likely reflects the chosen RNAi fragment, as it is smaller than conventional constructs and not 100% homologous to *MtSAT2*.

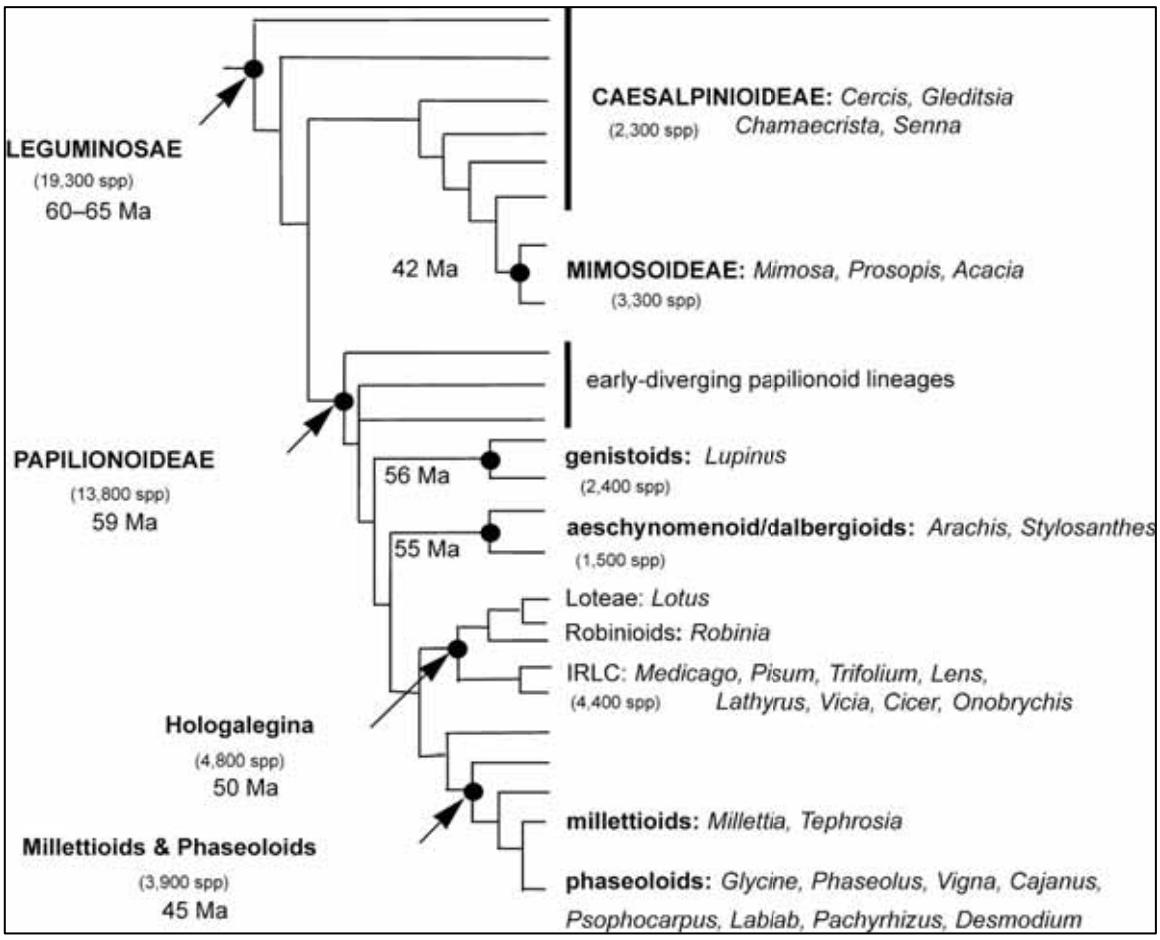


Figure 3.1. Schematic tree of legume evolutionary history.
Overview of the legume lineage showing the three subfamilies: Caesalpinoideae, Mimosoideae, and Papilionoideae. The Papilionoideae are further subdivided into four clades: Genistoid, Aeschynomenoid, Hologalegina, and Phaseoloid/Millettioid. The Hologalegina (*Medicago*), and Millettioid/Phaseoloid (soybean) are estimated to have diverged ~54 million years ago. Figure taken from (Gepts et al., 2005).

Figure 3.2. mRNA sequence and translation of *MtSAT1*.
The mRNA sequence of *MtSAT1* was determined by 5' and 3' RACE analysis combined with cloning and sequencing of the full-length product.

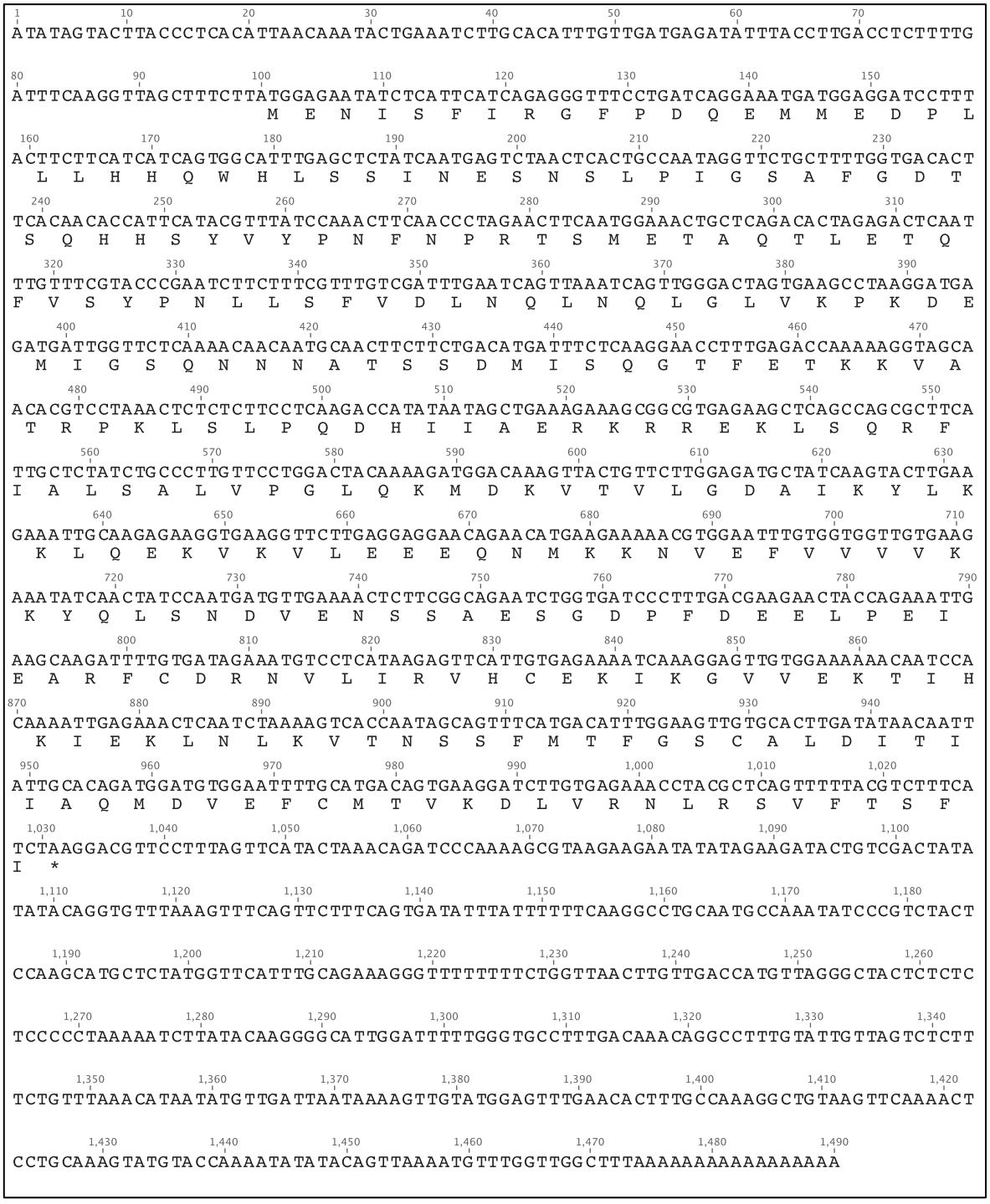
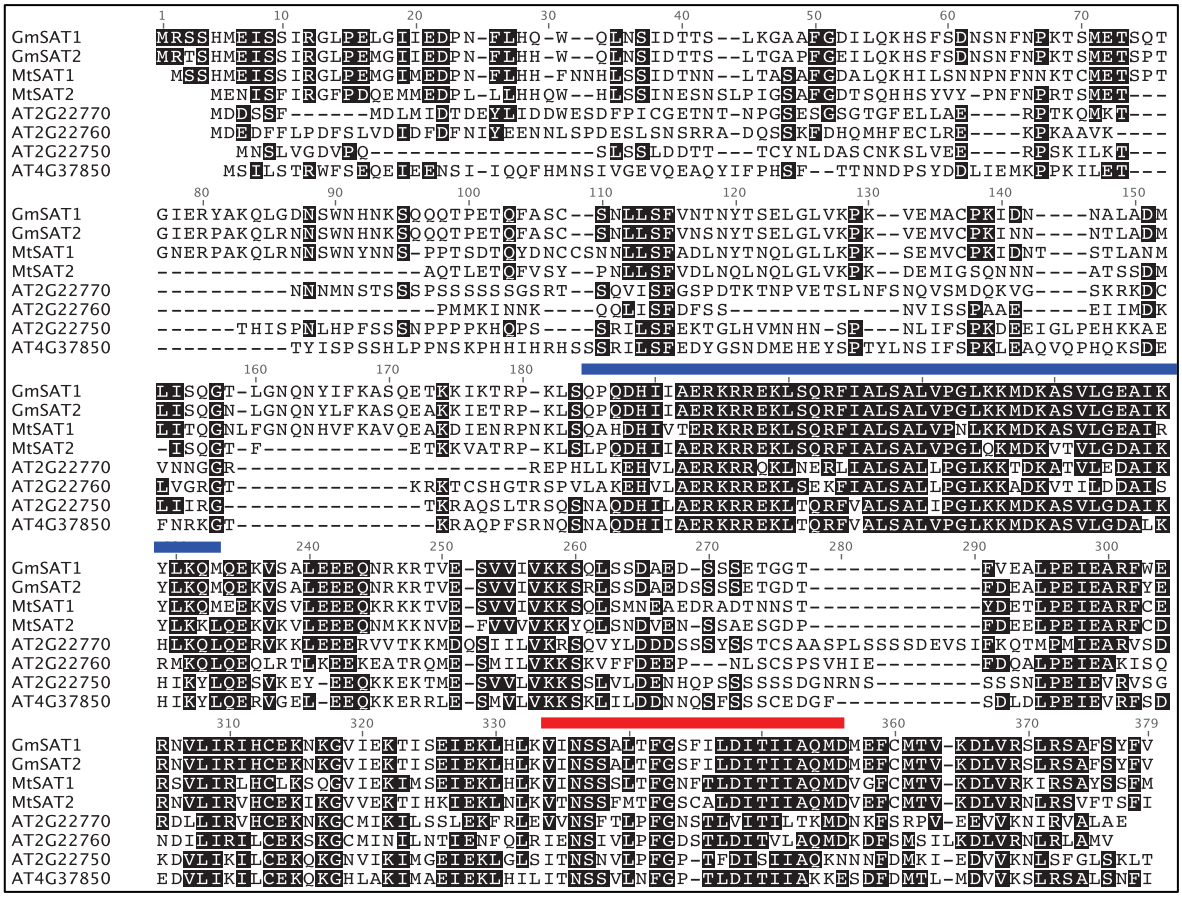


Figure 3.3. mRNA sequence and translation of *MtSAT2*.
The mRNA sequence of *MtSAT2* was determined by 5' and 3' RACE analysis combined with cloning and sequencing of the full-length product.

A



B

	GmSAT1	GmSAT2	MtSAT1	MtSAT2	AT2G22770	AT2G22760	AT2G22750	AT4G37850
GmSAT1		93.4%	69.8%	60.4%	28.8%	29.3%	34.9%	36.2%
GmSAT2	93.4%		72.0%	60.2%	29.4%	29.5%	34.7%	36.2%
MtSAT1	69.8%	72.0%		53.0%	28.0%	27.8%	32.8%	33.3%
MtSAT2	60.4%	60.2%	53.0%		30.9%	32.4%	32.9%	37.3%
AT2G22770	28.8%	29.4%	28.0%	30.9%		36.4%	32.6%	31.2%
AT2G22760	29.3%	29.5%	27.8%	32.4%	36.4%		34.0%	31.5%
AT2G22750	34.9%	34.7%	32.8%	32.9%	32.6%	34.0%		50.4%
AT4G37850	36.2%	36.2%	33.3%	37.3%	31.2%	31.5%	50.4%	

C

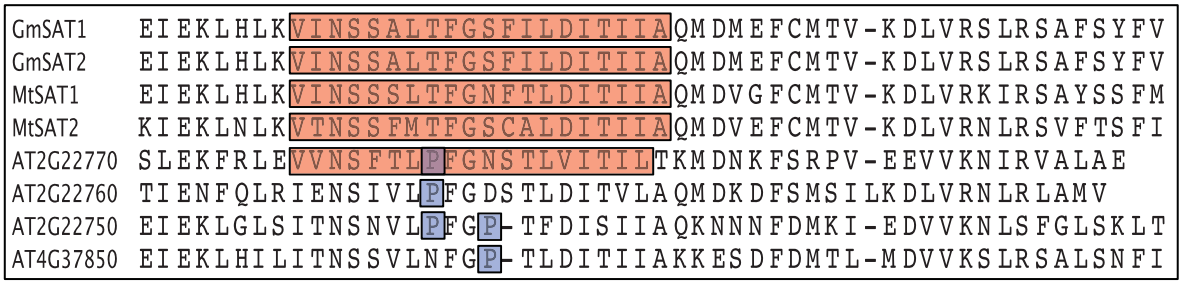


Figure 3.4. Protein sequence alignment of Soybean, *Medicago*, and *Arabidopsis* SATs. **A**, MUSCLE alignment of the protein sequences of soybean, *Medicago*, and *Arabidopsis* SATs using Geneious software. The bHLH domain is indicated in blue, the putative transmembrane domain in red. **B**, Table showing the percentage of identical residues shared between the aligned proteins. **C**, C-termini of soybean, *Medicago*, and *Arabidopsis* SATs showing TMpred (http://www.ch.embnet.org/software/TMPRED_form.html) prediction of transmembrane domains (red) along with proline residues (blue).

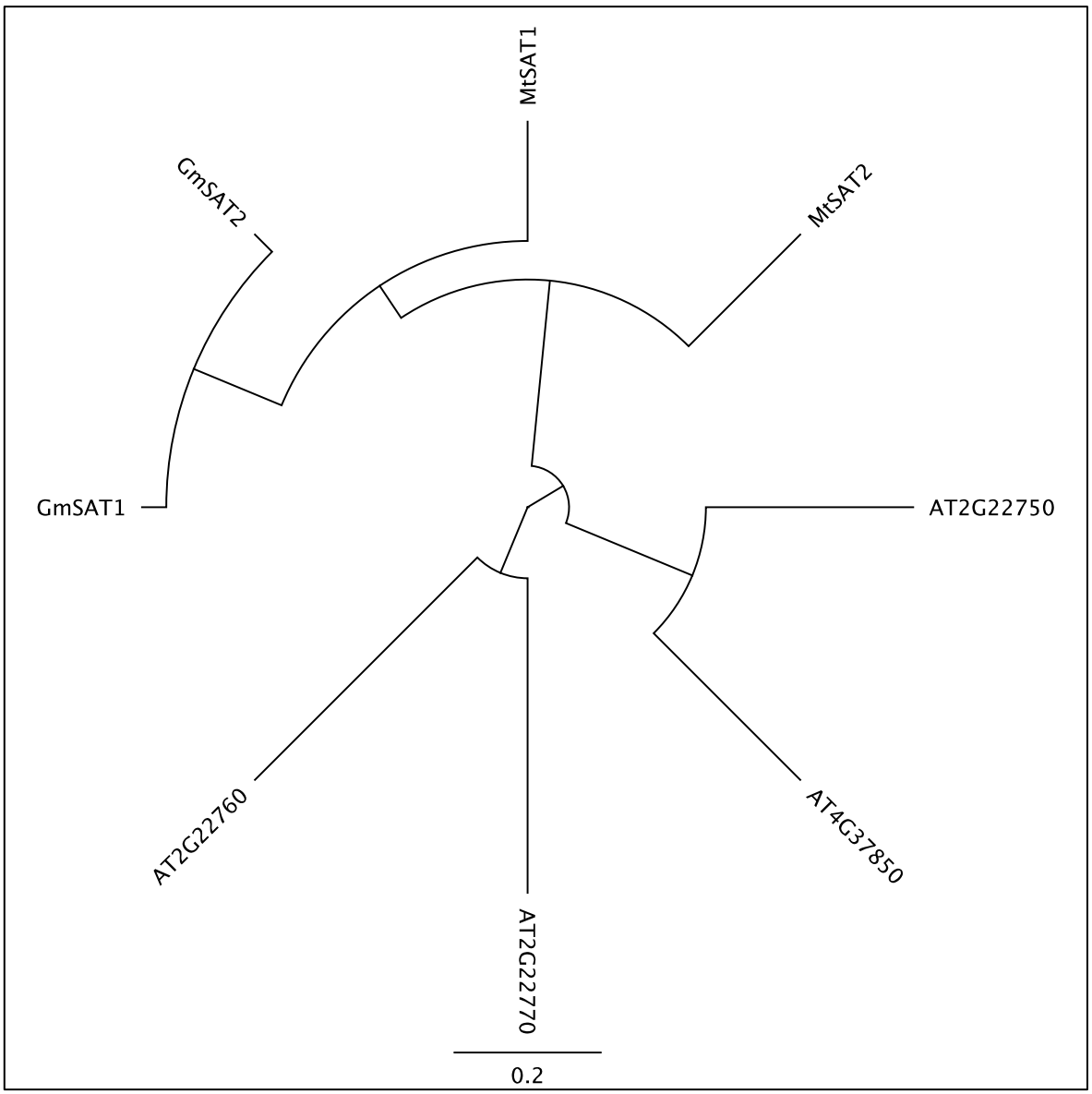


Figure 3.5. Circular phylogenetic tree of soybean, *Medicago*, and *Arabidopsis* SATs. The neighbor-joining tree was generated with Geneious software, using the Jukes-Cantor distance model. Scale represents the number of amino acid substitutions per site.

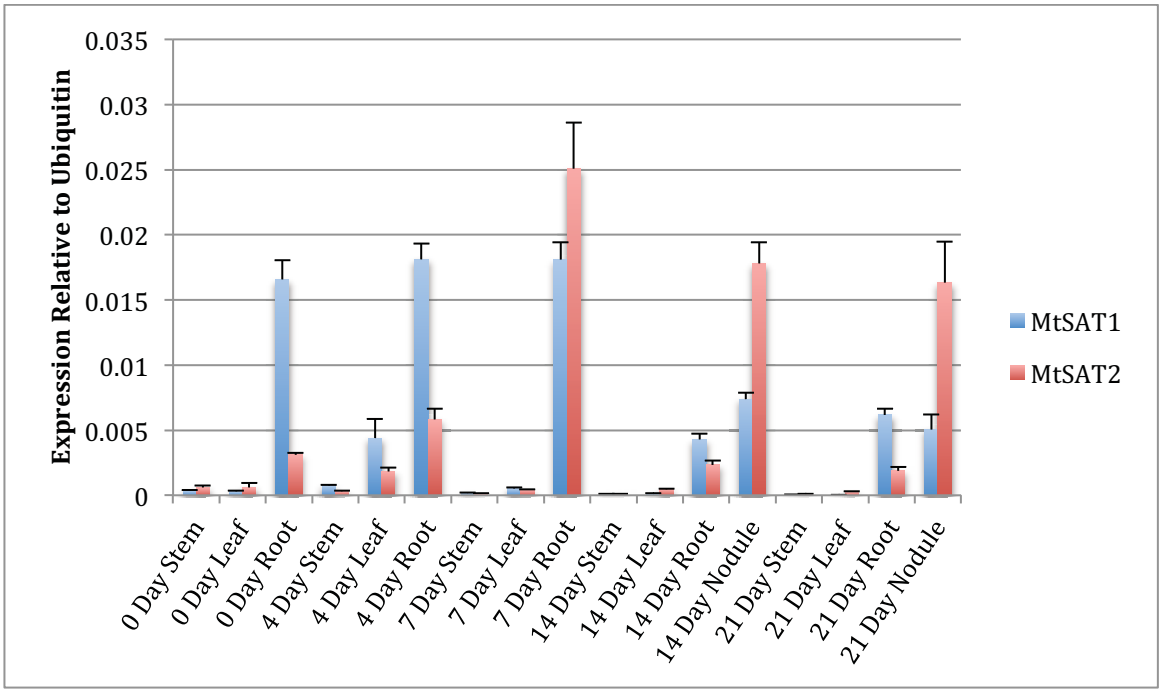


Figure 3.6. qPCR expression analysis of *MtSAT1* and *MtSAT2* in *Medicago*.
Medicago stems, leaves, roots, and nodules were harvested at 0, 4, 7, 14, and 21 days after inoculation with *S. meliltoti* 2011. RNA was extracted from pooled plants grown under identical conditions. qPCR was carried out with MtSAT1 qPCR F + R and MtSAT2 qPCR F + R primers, respectively (Table 3.1). Data values represent the means of three independent technical replicates ±SD relative to *MtUBQ10* (Ivanov et al., 2012), as calculated by the ΔCT method (Livak and Schmittgen, 2001).

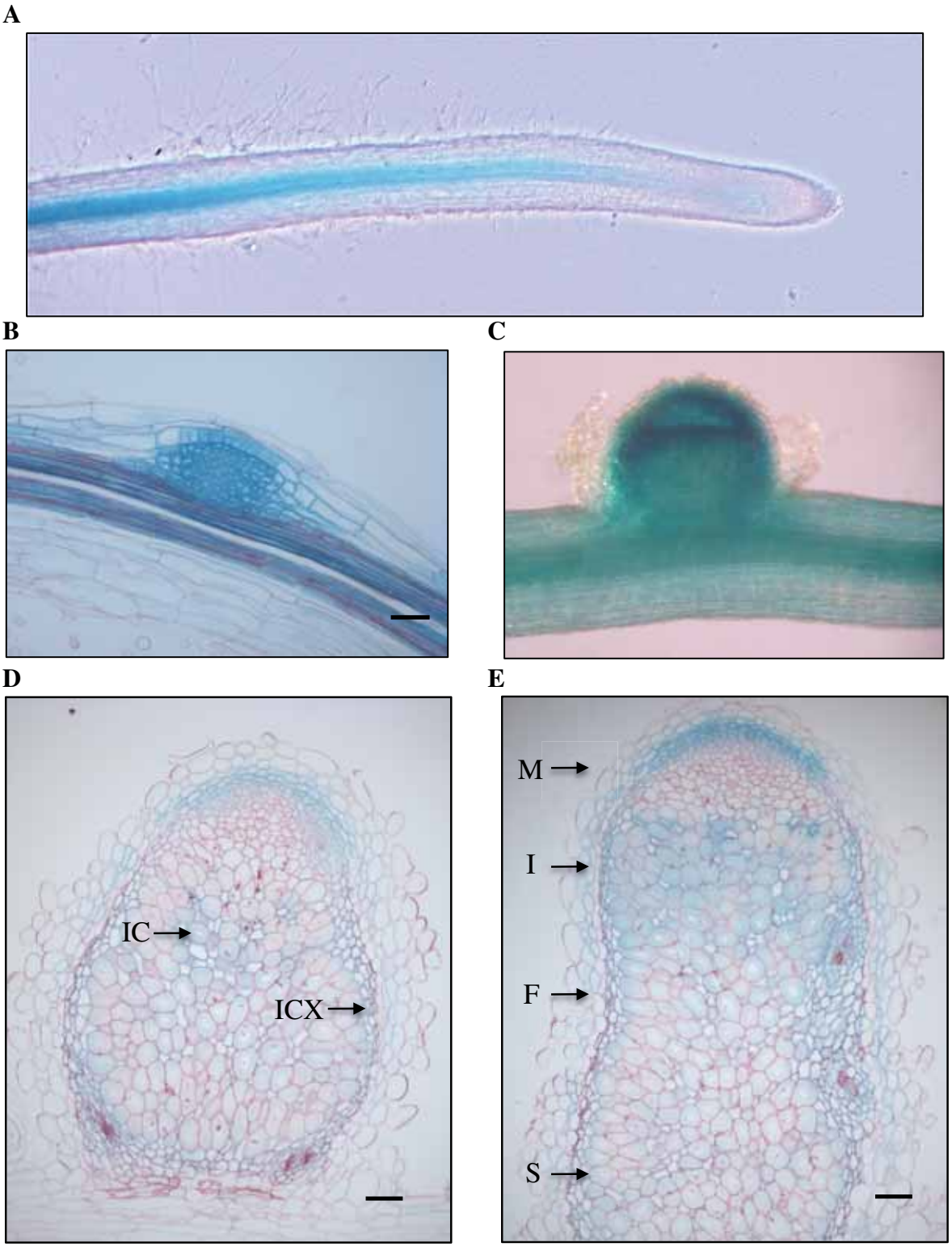


Figure 3.7. Promoter analysis of *MtSAT1* in *Medicago* roots and nodules.
Transgenic roots were generated by *A. rhizogenes* MSU440 carrying the plasmid p243-RedRoot-GUS (Figure 4.14) containing a copy of the *MtSAT1* promoter and inoculated with *S. meliloti* 2011. DS-Red-positive hairy roots were stained for GUS activity. The time since inoculation was: (**A**) 0 days, (**B**) 6 days, (**C**) 9 days, (**D**) 14 days, (**E**) 21 days. Infected cell (IC), inner cortex (ICX), meristem (M), infection zone (I), fixation zone (F), and senescence zone (S) are indicated. Bar = 80 μm (B), 200 μM (D and E).

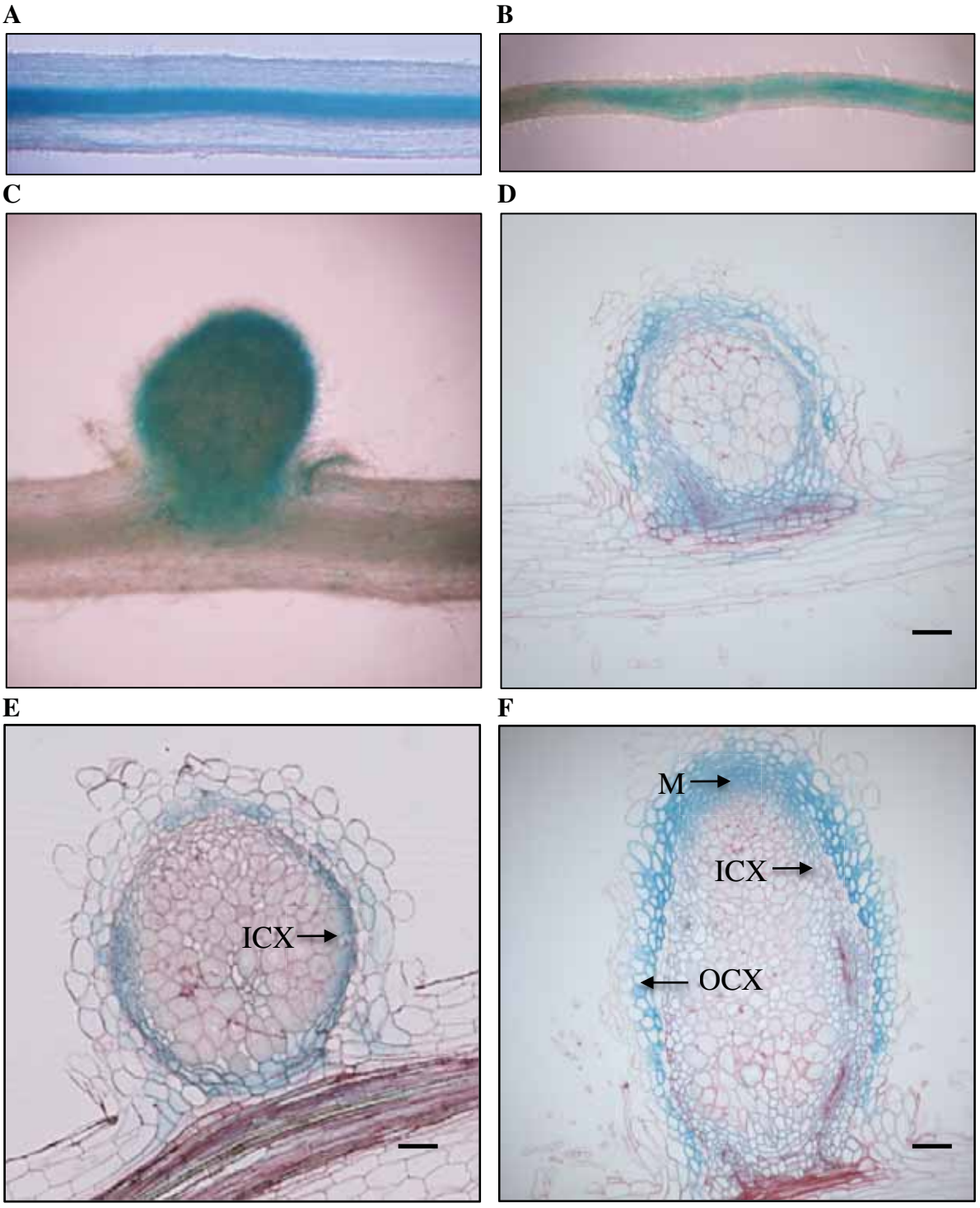
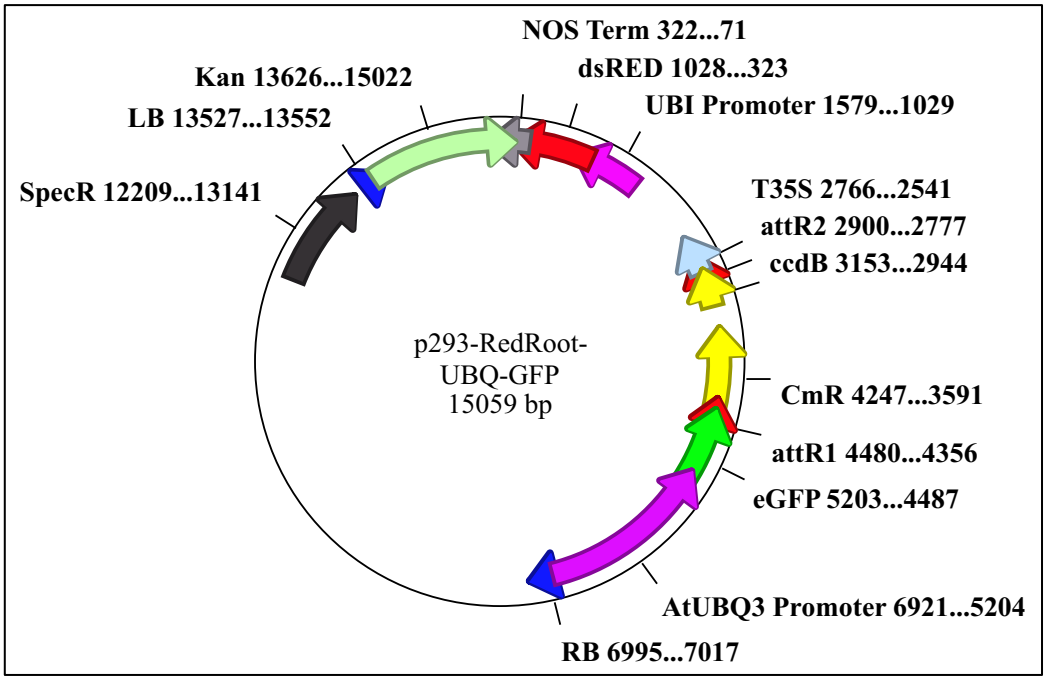


Figure 3.8. Promoter analysis of *MtSAT2* in *Medicago*.
Transgenic roots were generated by *A. rhizogenes* MSU440 carrying the plasmid p243-RedRoot-GUS (Figure 4.14) containing a copy of the *MtSAT2* promoter and inoculated with *S. meliloti* 2011. DS-Red-positive hairy roots were stained for GUS activity. The time since inoculation was: (**A**) 0 days, (**B**) 3 days, (**C**, **D**) 9 days, (**E**) 14 days, (**F**) 21 days. Inner cortex (ICX), outer cortex (OCX), and meristem (M) are indicated. Bar = 80 μm (D and E), 200 μm (F).

A



B

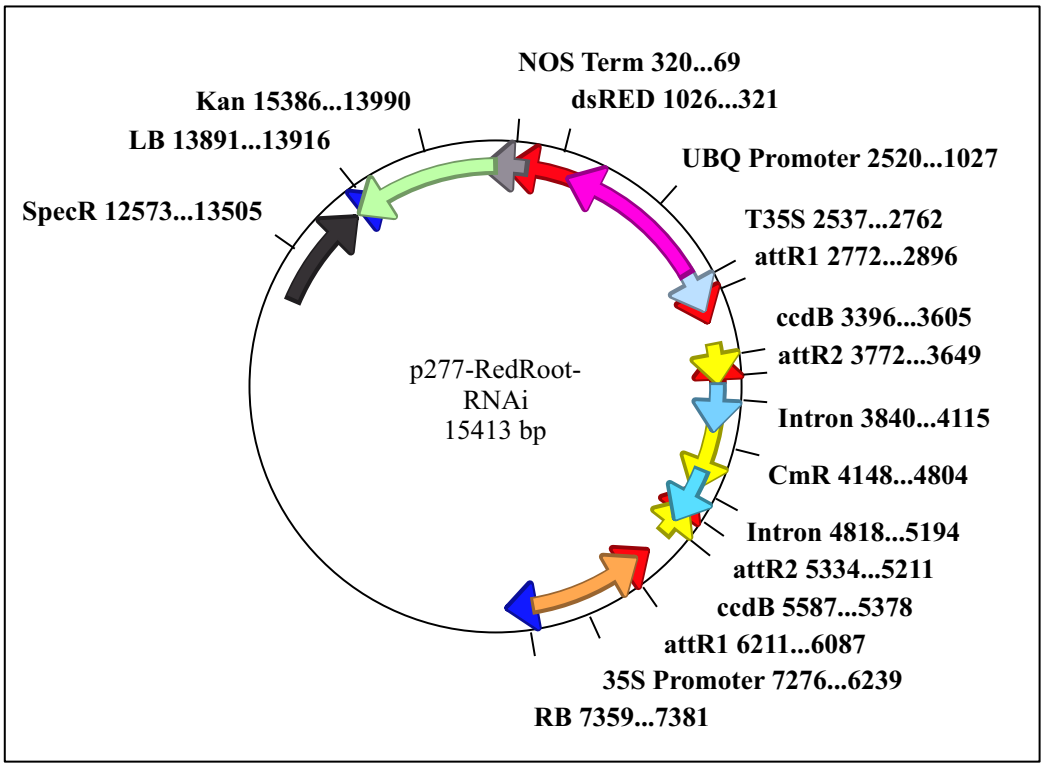


Figure 3.9. Plasmids used for GFP-fusion and RNAi experiments.
A, p293-RedRoot-UBQ-GFP plasmid used for N-terminal GFP fusions, driven by the *ubiquitin3* promoter (Limpens et al., 2009). **B**, p277-RedRoot-RNAi plasmid used for RNA interference (Limpens et al., 2005). Both plasmids carry a dsRED fluorescent reporter under control of the UBQ10 promoter.

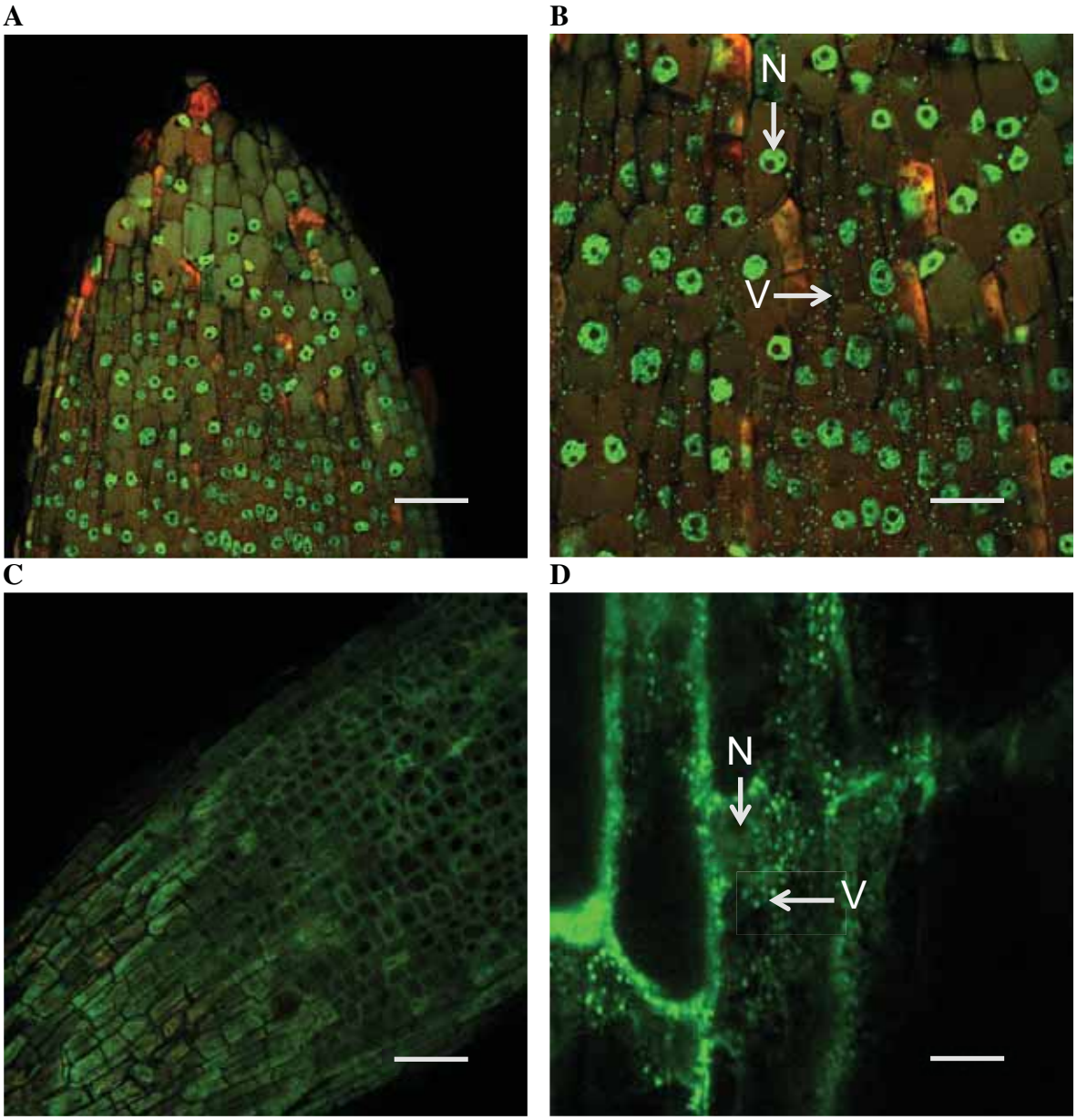


Figure 3.10. Localisation of MtSAT1 and MtSAT2 *in planta* by GFP fusion.
Transgenic roots were generated by *A. rhizogenes* MSU440 carrying the plasmid p293-RedRoot-UBQ-GFP containing a copy of either *MtSAT1* or *MtSAT2*. **A**, GFP:MtSAT1 in a root tip. **B**, Close-up of GFP:MtSAT1 showing nuclei (N) and vesicle (V) localization. **C**, GFP:MtSAT2 in root tip. **D**, Close-up of GFP:MtSAT2 in roots, also showing vesicle (V) and nuclear (N) localization. Images were obtained by confocal microscopy. Bar = 400μm (A, C), 200 μm (B), and 40 μm (D).

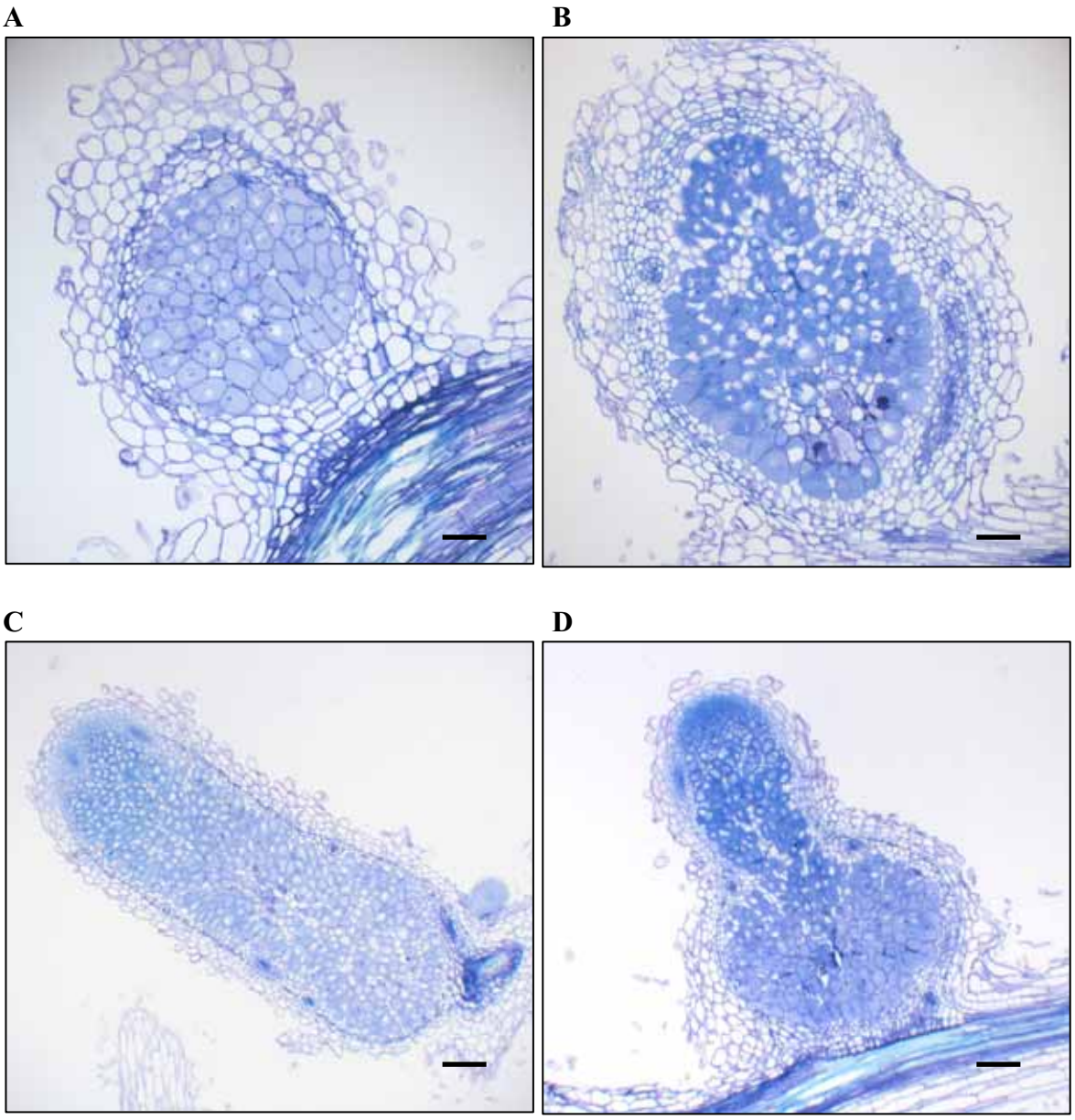


Figure 3.11. Effect of silencing *MtSAT1* and *MtSAT2* in *Medicago*.
Transgenic roots were generated by *A. rhizogenes* MSU440 carrying the plasmid p277-RedRoot-RNAi containing a copy of the *MtSAT1 RNAi* fragment (targeting both *MtSAT1* and *MtSAT2*) and inoculated with *S. meliloti* 2011. Hairy roots positive for DS-Red fluorescence were selected were embedded in Technovit 7100. Nodules were sectioned to 8μm and stained with toluidine blue **A**, 9 day-old empty vector control nodule. **B**, 9 day-old *MtSAT1/2* RNAi. **C**, 24 day-old empty vector control. **D**, 24 day-old *MtSAT1/2* RNAi. Bar = 200 μM (C and D). Bar = 80 μm (A and B), 200 μm (C and D).

Primer Name	Primer Sequence
MtSAT1 CDS F	ATGAGTTCTCATATGGAGATTTCA
MtSAT1 CDS R	TCACATGAAAGATGAATATGCTGAA
MtSAT1 5'RACE R	CATCTCCAAAAGCAGAAGCA
MtSAT1 5509 F	GAGAATCGTCCTAATAAACTTTCTC
MtSAT1 7538 F	CTCTTGCAAACATGTTGATCACTC
MtSAT1 Promoter F	GGGAACCGCCAAGAGAGATT
MtSAT1 Promoter R	ATGAGAACTCATCCTGAAAAAGAAATTA
MtSAT2 CDS F	ATGGAGAATATCTCATTCATCAGAG
MtSAT2 CDS R	TTAGATGAAAGACGTAAAAACTGAG
MtSAT2 5'RACE R	GAGCTCAAATGCCACTGATGA
MtSAT2 Promoter F	GATCCTCCAGGTTCCATCTTC
MtSAT2 Promoter R	AAGAAAGCTAACCTTGAAATCAAAAG
MtSAT1qPCRF	CGCCTACCGGAAATGAAAGA
MtSAT1qPCRR	CAAATGAAAGAAGATTATTAGAGCA
MtSAT1 pPCR2 F	GTGAAGGATCTTGTGAGGAAGA
MtSAT1 qPCR2 R	TGCTTTATGGTTTTATTTATAGTGAAC
MtSAT2qPCRF	AGGTGAAGGTTCTTGAGGAGG
MtSAT2qPCRR	CGTCAAAGGGATCACCAGAT
MtSAT2 qPCR2 F	ACAGATCCCAAAAGCGTAAGAAG
MtSAT2 qPCR2 R	TGGCATTGCAGGCCTTGAAAAA
MtSAT1 RNAi F	TACCTGAAATTGAAGCAAGATTTTGTG
MtSAT1 RNAi R	TCACATGAAAGATGAATATGCTGAA
MtSAT2 RNAi F	ACAGATCCCAAAAGCGTAAGAAG
MtSAT2 RNAi R	AAAGCCAACCAAACATTTTAACTG
p277 RNAi 1F	GTTCATTTCATTTGGAGAGGACTG
p277 RNAi 1R	TGAAGACACAGAAAGCCGTAAGA
p277 RNAi 2F	ACTCGGTTATTTACAGCTTATTCATA
p277 RNAi 2R	CCCTTATCTGGGAACTACTCACA
MtUBI10 qPCR F	CCCTTCATCTTGTCCTTCGTCTG
MtUBI10 qPCR R	CACCTCCAATGTAATGGTCTTTCC

Table 3.1. List of primers used in this chapter.

3.3 Materials and Methods

3.3.1 Cloning *MtSAT1* and *MtSAT2* and RACE

RNA was extracted from mixed roots and nodules with a RNeasy kit (Qiagen). 5' RACE was carried out with a FirstChoice RLM RACE kit (Ambion). Instead of using the supplied MLV reverse transcriptase and random decamer primers, Superscript III (Invitrogen) was used in conjunction with an oligo dT primer to make full length cDNA for 5 'RACE. For *MtSAT1*, the first round of 5' RACE was carried out with the MtSAT1 qPCR R primer (with the 5' outer primer of the kit), and the second round with the MtSAT1 5' RACE R primer (with the supplied 5' inner primer). For *MtSAT2*, the same strategy was employed using the primers MtSAT1 qPCR R (round one) and MtSAT1 5'RACE R (round two). 3' RACE was carried out using an ExactSTART Eukaryotic mRNA 5'- & 3'-RACE Kit (Epicentre Biotechnologies). For *MtSAT1* 3' RACE, the primers MtSAT1 7538 F and PCR2 (supplied with kit) were used for round one, and MtSAT1 5509 F and PCR2 inner (custom nested primer) for round two. For *MtSAT2*, just one round was required using the primers MtSAT2 qPCR F and PCR2. In all cases, the amplified fragments were gel purified, ligated in pGEM-Teasy (Promega), and sequenced (results shown in Figure 3.2and Figure 3.3). Having the RACE sequencing results enabled the design of primers to clone the full-length open reading frames. The *MtSAT1* coding sequence was amplified with the MtSAT1 CDS F and R primers, and *MtSAT2* was amplified with the MtSAT2 CDS F and R primers. The fragments were then inserted into pCR8-TOPO and sequenced.

3.3.2 *MtSAT1* and *MtSAT2* gene expression

M. truncatula A17 seeds were scarified in concentrated H_2SO_4 for 10 min and rinsed with water six times. Next, the seeds were soaked in 3% (w/v) (active chlorine) calcium hypchlorite (made fresh) for 8 min, then washed seven times with sterile water. The seeds were then spread on sterile filter paper, which was resting on Färhaeus medium (Limpens et al., 2004) in a petri dish. The plates were sealed with parafilm, wrapped in aluminum foil, and placed upside down at 4°C overnight. The plates were then incubated in the dark, upside down at room temperature for 24 hours, and then exposed to light on the bench for

8 hours. The germinated seedlings were then planted on Färhaeus medium plates containing a sterile half round piece of filter paper (five per plate). The plates were sealed and placed vertically in a growth cabinet at 21°C with relatively low light from standard fluorescent bulbs (~50 PAR). The seedlings were grown for seven days, and then transferred to perlite soaked in Färhaeus medium (without nitrogen). The plants were misted with water daily for four days while the trays were wrapped with plastic to maintain humidity. At day seven, the first tissues were harvested (day zero), then inoculated with *S. meliloti* 2011 (2 ml of a OD_{600} = 0.1 suspension per plant). Tissues were harvested and frozen in liquid nitrogen. RNA was extracted with a Spectrum Total RNA kit (Sigma) and cDNA synthesized with Superscript III (Invitrogen). qPCR was carried out using SYBR Green (Bio-Rad) on a light cycler qPCR machine (Bio-Rad) similar to section 2.3.5. *MtSAT1* was monitored using the primers MtSAT1 qPCR2 F and R, and *MtSAT2* using the primers MtSAT2 pPCR2 F and R. Relative expression levels were determined by delta C_T method (Livak and Schmittgen, 2001) using *ubiquitin10* (Medtr4g124610) as the control gene (primers MtUBI10 qPCR F and R), as used by Ivanov et al. (2012).

3.3.3 Promoter analysis of *MtSAT1* and *MtSAT2*

The promoters of *MtSAT1* and *MtSAT2* were cloned using the primers MtSAT1 Promoter F and R and MtSAT2 Promoter F and R. The fragments were inserted into pCR8-TOPO and recombined into p243-RedRoot-GUS (Limpens et al., 2005), shown in Figure 4.14. The plasmids were then electroporated into *Agrobacterium rhizogenes* MSU440. To generate hairy roots, *Medicago* A17 seeds were germinated, and plated as above. After being placed on the Färhaeus medium (Limpens et al., 2004) plates with the half filter, the plants were incubated vertically for five days, after which the roots were removed with a scalpel (at the point where the stem turns green). The freshly cut hypocotol was then scraped over a plate containing a lawn of *A. rhizogenes* containing the construct and placed back on the half filter and resealed with parafilm. After five days, the plants were then transferred to a new plate containing emergence medium and sandwiched between two half filters. The plates were sealed and placed vertically in a chamber for two weeks. Plants containing newly emerged roots were transferred to perlite soaked in Färhaeus medium (without nitrogen), and inoculated seven days later with *S. meliloti* 2011. Positive hairy roots were screened for dsRED expression, then stained for GUS activity (same assay as outlined in

2.3.6). Nodules were then embedded in Technovit 7100 and 8 μm sections were cut with a glass knife. Sections were stained in 0.05% ruthenium red for 30 min and embedded by adding DPX (Sigma) onto the slide with a coverslip. Images were obtained with a Leica ASLMD laser-assisted micro-dissection microscope.

3.3.4 Analysis of GFP fusions to MtSAT1 and MtSAT2

To generate N-terminal GFP fusions, MtSAT1- and MtSAT2-pCR8-TOPO were recombined into the destination vector p293-pK7WGF2-RedRoot-UBQ-GFP (Figure 3.9) using LR clonase (Invitrogen). The plasmids were then transformed in *Agrobacterium rhizogenes* MSU440 by electroporation. Hairy roots were generated as described above. After fourteen days, roots staining positive were selected and viewed on a Zeiss PASCAL Confocal Laser Scanning microscope for GFP (588 argon laser excitation with GFP filter set).

3.3.5 Silencing of *MtSAT1* and *MtSAT2*

A 253 bp portion of the *MtSAT1* coding sequence was amplified with the primers MtSAT1 RNAi F and R. The fragment was inserted into pCR8-TOPO and recombined into p277-RedRoot-RNAi (Figure 3.9). The double insertion was checked by digestion and PCR with the primer combinations p277 RNAi 1F and 1R and p277 RNAi 2F and 2R. PCR with these primer sets is necessary since this double insertion event can cause unwanted rearrangements. The plasmid was then electroporated into *Agrobacterium rhizogenes* MSU440 and hairy roots were generated as described above. dsRED positive nodules were picked and embedded in Technovit 7100 for microscopy. Sections were stained with 0.1% Toludine Blue for 1 min, rinsed for 5 min, and embedded in DPX (Sigma). Nodules were also used for qPCR after extracting RNA with a Spectrum Total RNA kit (Sigma) and generating cDNA with Superscript III (Invitrogen), as described in 3.3.2.

3.4 Discussion

3.4.1 Comparison of *MtSAT1* and *MtSAT2* to *GmSAT1*

MtSAT1 and *MtSAT2* (*MtSATs*) are orthologous to *GmSAT1*, last sharing a common ancestor 54 mya. It would appear that the *MtSAT1* and *MtSAT2* arose during a duplication event (as evidenced by the synteny between these loci), possibly via the whole-genome duplication (WGD) event that occurred in the ancestor of legumes 58 mya (Pfeil et al., 2005). In contrast, *GmSAT2* appears to have arisen from the soybean WGD, 13 mya (Schlueter et al., 2007). Other *GmSAT1*-like sequences exist in soybean, but appear to have diverged greatly over time. The existence of *SAT*-like genes in *Arabidopsis*, which diverged from the legume lineage 115 to 93 mya (Wang et al., 2009) supports the notion that SATs are an ancient subfamily of bHLH transcription factors.

From the qPCR, GUS, and GFP localization studies, the *MtSATs* appear to overlap in distribution with *GmSAT1*. Both the *MtSAT*s and *GmSAT1* are confined to the stele of non-nodulated roots and the inner cortex of nodules. Additionally, *MtSAT1* and *GmSAT1* are observed at a low level in infected cells as well. *MtSAT1* and *MtSAT2* do differ in their expression level, where *MtSAT1* is present at a relatively higher level in non-nodulated roots, while *MtSAT2* appears to be stimulated during early nodulation events and is the more abundant transcript in nodules.

GFP fusions of MtSAT1, MtSAT2, and GmSAT1 *in planta* all show localization to vesicles and nuclei. Therefore, the function of SAT-like proteins in legumes may require membrane targeting. NAI is the only *Arabidopsis* SAT localized thus far, appearing to be exclusively nuclear in the only image presented (Tominaga-Wada et al., 2011). The *Arabidopsis* SATs contain a helix-breaking/bending proline residue (Barlow and Thornton, 1988) in the transmembrane domain. This residue is present in SAT-like proteins from corn, *Brachypodium*, and rice (monocots), but absent in SATs from grapevine and poplar. Testing the effect of the proline residue on transmembrane helix stability would be recommended. Based upon localization, *MtSAT1* and *MtSAT2* may be redundant, but could be functionally similar to *GmSAT1*. *MtSAT1* and *MtSAT2* were transformed into the yeast strain 26972c, but did not complement the ammonium phenotype or induce MA

uptake (not shown). Therefore, the residues that differ between these proteins may contribute to their differing functions in yeast. If a stable mutant line of *GmSAT1* was generated in soybean, then *MtSAT1* could be tested for its ability to complement *GmSAT1*. This approach has been used by studies to test the conservation of function of common symbiosis pathway genes (Banba et al., 2008).

3.4.2 Silencing of *MtSAT1* and *MtSAT2*

A portion of the C-terminus of *MtSAT1* was used for RNAi, which targeted both *MtSAT1* and *MtSAT2*. Positive hairy roots showed a spectrum of phenotypes, from nodules that were aborted at early stages to nodules that seemingly initiated a new meristem. Since poor nodule development is the hallmark of the *GmSAT1* RNAi phenotype, there could be some common functionality. The spectrum of phenotypes could reflect the fragment chosen for RNAi. Since *MtSAT1* was effectively targeted, but *MtSAT2* to a lesser degree, a new strategy is required to target both genes. Fusing a larger fragment from each gene by overlapping PCR would be advised, since this strategy was recently effective in targeting two unique genes (Ivanov et al., 2012). It would also be useful to test other promoters for driving the RNAi fragment. This study utilized the CaMV 35S promoter, but other promoters have been identified with broader tissue expression and higher levels of activity (Auriac and Timmers, 2007; Govindarajulu et al., 2008).

4 Identification of a Novel Family of Ammonium-Transporting Major Facilitator Proteins

4.1 Introduction

4.1.1 Ammonium transport in yeast

Yeast (*Saccharomyces cerevisiae*) contains three ammonium transporters, termed MEP1, MEP2, and MEP3 (Dubois and Grenson, 1979; Marini et al., 1994; Marini et al., 1997). MEP2 displays the highest affinity for NH_4^+ (K_m 1-2 μM), followed by MEP1 (K_m 5-10 μM), and MEP3 (K_m 1.4-2.1 mM). The MEP proteins are responsible for sequestering NH_4^+ from the medium as a nitrogen source, with MEP2 also serving as a sensor for pseudohyphal growth during NH_4^+ limitation (Lorenz and Heitman, 1998). Expression of any one MEP transporter is sufficient to allow growth on media containing low levels of NH_4^+, however they are not essential for growth on higher concentrations of NH_4^+ (>20 mM). MEP1 and MEP3 are closely related (79% identity), however MEP2 is more divergent (40% identity). The MEP protein sequences have been used to isolate novel NH_4^+ transporters from vertebrates. For example, the Rhesus-associated membrane protein RhAG was identified through homology to the MEPs, and was able to restore growth on low ammonium to a yeast triple mutant (strain 31019b) for the MEP transporters (Marini et al., 2000).

4.1.2 Ammonium transport in plants

Yeast has also been used to isolate NH_4^+ transporters from plants. At the same time that *MEP1* was first cloned and characterized, the yeast strain 26972c was used to isolate the first NH_4^+ transporter from plants, called AMT from *Arabidopsis* (Ninnemann et al., 1994). Since their discovery, AMT/Rh sequences have been identified in most sequenced organisms (von Wirén and Merrick, 2004). *Arabidopsis,* as a representative plant, contains six AMT genes (five AMT1 and one AMT2). Both AMT transporter families have a high affinity (saturated at ≤1 mM) and high selectivity for NH_4^+ (K_m in μM range), but only the

AMT1 family is able to transport the NH_4^+ analog methylammonium (Pantoja, 2012). Ammonium transporters have been shown to associate as homotrimers, with each monomer possessing a hydrophobic channel (Zheng et al., 2004; Andrade et al., 2005; Zheng et al., 2009). Loss of *AtAMT1;1* reduces the uptake of NH_4^+ by ~30%, and is reduced a further ~30% after the loss of *AtAMT1;3* (Kaiser et al., 2002; Loqué et al., 2006). Interestingly, the decrease in ammonium flux in these mutants had minor effects on normal plant growth. To observe an effect in the *AtAMT1;1* mutant, the plant must be grown in conditions where ammonium is the sole nitrogen source (along with the addition of sucrose). These and other observations have revealed a low affinity flux of NH_4^+ that is functioning in addition to the high affinity system (Wang et al., 1993; Rawat et al., 1999; Kaiser et al., 2002).

4.1.3 Major Facilitator Transporters

The major facilitator superfamily (MFS) is the largest and most diverse superfamily of secondary active transporters (Saier et al., 1999). Nearly half of all solute transports belong to either the MFS or ATP-binding cassette (ABC) superfamilies (Pao et al., 1998). These two superfamilies are found in all classifications of living organisms. MFS superfamily transporters are single-polypeptide secondary carriers, often with 12-14 transmembrane domains centered on a central pore. The high-resolution crystal structures of *E. coli* GlpT, LacY, and EMrD have revealed that despite sequence divergence, the 3-D structures are very similar (Law et al., 2008). The superfamily can be divided into three distinct classifications based on kinetics: 1) uniporters; 2) symporters; and 3) antiporters. The MFS superfamily contains 74 families (each transporting a related set of compounds) such as Organic Anion Transporters (OAT), Oligosaccharide:H^+ Symporters (OHS), Proton-dependent Oligopeptide Transporters (POT), and Organophosphate:P Antiporters (OPA). Studies have shown that MFS transporters can translocate such compounds as: amino acids, monosaccharides, oligosaccharides, nucleotides, peptides, vitamins, and organic and inorganic anions and cations (Reddy et al., 2012).

Using a Pfam ontology search of the soybean genome uncovers 978 transporters, with 278 being classified as major facilitator superfamily members. A similar search in *Arabidopsis* uncovers 137 major facilitator superfamily members, of which 25 have been assigned a

gene symbol. The characterized transporters belong to such diverse groups as: zinc-induced facilitator-like (ZIFL), tonoplast monosaccharide transporter (TMT), polyol/monosaccharide transporter (ATPMT), sucrose-proton symporter (SUC), sugar transporter (STP), organic cation/carnitine transporter (OCT), inositol transporter (INT), and equilibrative nucleoside transporter (ENT). The wide range of MFS substrates, coupled with very divergent primary sequences, makes a functional prediction challenging.

Recently, our lab has characterized a novel low-affinity ammonium MFS transporter from yeast, called YOR378W. YOR378W is a major facilitator family member and its sequence was used to identify a small MFS subfamily of the sugar transporter/spinster transmembrane proteins classification (KOG) from soybean. A representative member of this soybean subfamily (*GmMFS1.3*) was cloned and characterized. Interestingly, the expression of the soybean MFS transporter in yeast induces the uptake of ammonium and methyammonium (D. Mazurkiewicz, unpublished results). This subfamily of MFS transporters may be responsible for the previously observed low affinity NH_4^+ flux in plants.

4.2 Results

4.2.1 Identification of a novel family of major facilitator proteins

The protein encoded by the locus YOR378W contains 515 amino acids and is predicted to be 56.11 kDa in size with 10 transmembrane domains. Using the predicted sequence of the YOR378W locus from yeast, a blast search was conducted to find related proteins in plants. Here, an uncharacterized group of transporters belonging to the major facilitator superfamily was identified. Focusing on legumes, the top hit was a predicted protein encoded by the gene Glyma08g06880 (predicted to be 495 amino acids in length, with a MW of 53.67 kDa) from soybean (Figure 4.1). Further blast searches with Glyma08g06880 protein uncovered three more related genes in soybean. The soybean genes were named based upon individual alignments with YOR378W. The predicted genes are: *GmMFS1.1* (Glyma13g32670), *GmMFS1.2* (Glyma07g30370), *GmMFS1.3* (Glyma08g06880), *GmMFS1.4* (Glyma15g06660), ranging from 17.7% overall identity to 15.1% (Figure 4.2). A fifth sequence was identified at a later date, *GmMFS1.5*

(Glyma09g33680). This protein was initially looked over during BLAST searches because of an incorrect sequence prediction. However, after identifying this gene as being downregulated in the *GmSAT1* RNAi microarray, it was revisited.

Four proteins were also found in *Medicago* (and named according to chromosome location): Medtr2g010370 (*MtMFS1*), Medtr2g010390 (*MtMFS2*), Medtr4g121900 (*MtMFS3*), and Medtr5g030580 (*MtMFS4*). Similar to *GmMFS1.5*, *MtMFS4* was identified later because of similarity to *GmMFS1.5* and due to the fact that the *Medicago* genome was not fully annotated (a gene model was located by BLAST analysis of next-generation sequencing results from the Noble Foundation). In *Arabidopsis*, three proteins were found with high similarity (AT2G22730, AT5G64500, and AT5G65687). After constructing a phylogenetic tree based on amino acid differences, it was revealed that MtMFS1, MtMFS2, and GmMFS1.1 formed a cluster, with another containing GmMFS1.2, GmMFS1.3, and MtMFS3 (Figure 4.3). A third cluster contained GmMFS1.5 along with MtMFS4. AT5G64500 is the most similar to the legume proteins, with AT2G22730 and AT5G65687 forming a distinct branch.

4.2.2 Cloning of *GmMFS1.3* and *GmMFS1.5*

Initially, a primer set was designed to clone *GmMFS1.3* based on the Phytozome genome prediction however a product could not be amplified. Based on sequence alignments to the other MFS family members, it was found that there could be a missing exon in the genome prediction. Therefore, the downstream genomic sequence of *GmMFS1.3* was searched for potential exons. A candidate was found, and a new primer was created and used to successfully amplify a full-length product (Figure 4.4). The final sequence encodes for a protein of 537 amino acids with a predicted MW of 58,276 Da. Based on various programs, GmMFS1.3 is predicted to contain 12 transmembrane domains (Figure 4.1). *GmMFS1.5* was cloned after being identified in the *GmSAT1* RNAi microarray. Sequencing of the full-length product uncovered an exon that was not previously predicted (Figure 4.5). The final sequence of GmMFS1.5 is 496 amino acids in length (MW of 53,526 Da) and is predicted to contain 10 transmembrane domains. GmMFS1.3 and GmMFS1.5 are 67.1% identical at the amino acid level, with GmMFS1.3 containing an extra 25 amino acid extension at the N-terminus (Figure 4.2).

4.2.3 Genomic Synteny

Blast searches for major facilitator proteins were carried out initially to find sequences similar to YOR378W, which is a target of GmSAT1 when overexpressed in yeast. Surprisingly, the most highly related sequences in plants are physically located next to *GmSAT*-like genes in many sequenced dicot species (Figure 4.6). For example, *GmMFS1.4* (Glyma15g06660) sits very close to *GmSAT1* (Glyma15g06680) and *GmMFS1.1* (Glyma13g32670) is physically near *GmSAT2* (Glyma13g32650, Figure 2.5). In *Medicago*, MtSAT1 (Medtr2g010480) is linked with MtMFS1 (Medtr2g010370) and MtMFS2 (Medtr2g010390). The same is true for *AT2G22730* from Arabidopsis, which is located next to three copies of SAT-like sequences (*AT2G22750, AT2G22760, AT2G22770*). Next to *GmMFS1.2* (Glyma07g30370) is a *SAT*-like gene (Glyma07g30420) however next to *GmMFS1.3* (Glyma08g06880) there is no *SAT*-like gene model, only a short sequence with similarity to *GmSAT1* (52/64 bases). *GmMFS1.5* (Glyma09g33680), on the other hand, is not located near to a *GmSAT1*-like sequence, similar to *MtMFS4*.

4.2.4 Structure of GmMFS1.3 and mutations

A *de novo* structure prediction was made using the HMM-based Protein Structure Prediction, SAM-T08 (http://compbio.soe.ucsc.edu/SAM_T08/T08-query.html). GmMFS1.3 is predicted to be a typical major facilitator protein, with 12 transmembrane domains organized as a mirror image around a central pore (Figure 4.7). During the course of cloning *GmMFS1.3*, a number of mutations (D25G, M26T, W38R, T40I, V50A, I85M, S120G, I192V, and I251V) were isolated from *E. coli* grown at 37°C (conditions were later optimized). Using the predicted model, these mutations were found to map to a pore on the side of the protein (Figure 4.8).

4.2.5 Transport Properties of GmMFS1.3

Since YOR378W is known to transport methylammonium in yeast (D. Mazurkiewicz, unpublished results), *GmMFS1.3* was inserted into the yeast expression vector pYES3-DEST. Upon expression in 26972c, *GmMFS1.3*-pYES3 was able to induce a significant uptake of ^{14}C-methylammonium (MA) relative to control (Figure 4.9). Spheroplasts were then generated from the yeast cells and tested in a stopped-flow device coupled to a spectrophotometer. Here, upon mixing the cells with compatible buffers but differing in 0.5 mM MA, there was a significant change in light scattering at 475 nm for *GmMFS1.3*-pYES3 relative to controls.

4.2.6 Expression of MFS genes in Soybean

To test the expression of the identified soybean *GmMFS1.1-1.4* genes, qPCR primers were designed to amplify a unique portion of each transcript. In 36-day-old soybean tissues, all qPCR primer sets were able to amplify a product (Figure 4.10A) in most tissues by RT-PCR. The primer set for *GmMFS1.3* yielded the strongest product in nodules and was therefore chosen for qPCR. By qPCR, *GmMFS1.3* was expressed the highest in nodule tissue at both 15 and 36 days after inoculation, but present in all other organs. To complement the qPCR results, expression data was also obtained from the RNA seq database (http://www.soybase.org/soyseq/#). Similar to the results obtained by PCR, *GmMFS1.1-1.4* were expressed in all tissues at a relatively low level (Figure 4.11). Interestingly, *GmMFS1.5*, which was identified and cloned at a later date, is expressed at a higher level in all tissues, and shows enrichment in flowers, nodules, and roots.

4.2.7 Subcellular Localization of GmMFS1.3

In order to assess the localization of GmMFS1.3, it was sub-cloned into the expression plasmid pYFP-attr (Subramanian et al., 2006). pYFP-attr contains a double copy of the 35SCamV promoter, driving a N-terminal YFP fusion to a protein of interest (Figure 4.12). The plasmid was precipitated onto gold particles and bombarded into onion epidermal peels. After initial assessments of localization, the pYFP-GmMFS1.3 plasmid was co-precipitated with the pm-rk CD3-1007 construct. pm-rk CD3-1007 contains a C-terminal

mCherry fusion to the AtPIP2A aquaporin and is used as a plasma membrane marker. After viewing the co-bombarded tissues it was found that GmMFS1.3 and AtPIP2A co-localized to the plasma membrane (Figure 4.13).

4.2.8 Tissue-Specific Expression of *GmMFS1.3*

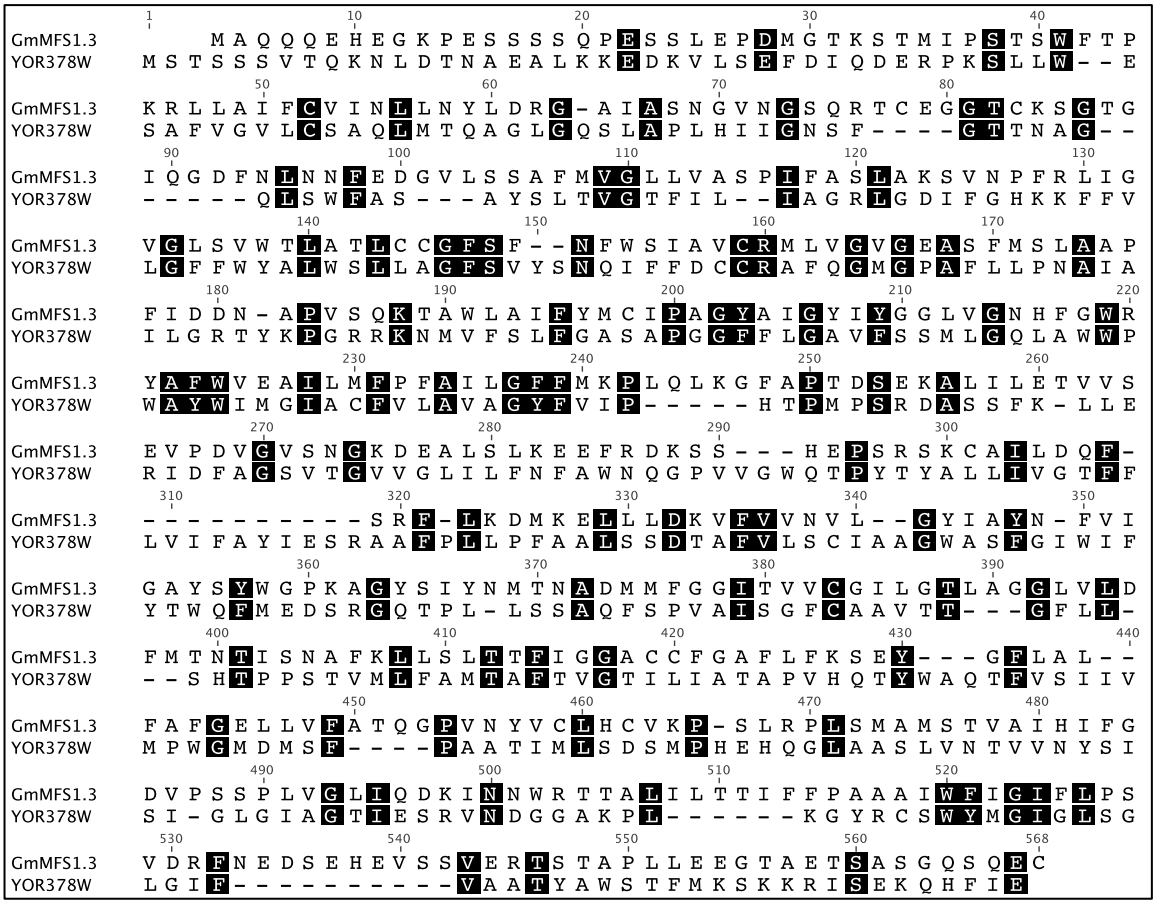
To determine the tissue-specific expression of *GmMFS1.3*, a 2 kb portion of the upstream DNA was cloned from soybean genomic DNA. The promoter was sub-cloned into the plasmid p243-RedRoot-GUS (Figure 4.14). p243-RedRoot-GUS contains a promoterless GUS/GFP fusion, as well as a ubiquitin-driven dsRED that is used for positive hairy root selection (Limpens et al., 2005). The construct was transformed into *Agrobacterium rhizogenes* K599 to generate hairy roots, similar to section 2.3.6. Positive roots were then stained for GUS activity and blue tissues were embedded in technovit 7100 and sectioned. In agreement to the qPCR results, the *GmMFS1.3* promoter was active in both roots and nodules. Nodule expression was confined to the inner cortex and the vasculature (Figure 4.15). In roots, the *GmMFS1.3* promoter was active in the cell layer found between the xylem and phloem in the stele.

4.2.9 Expression of *Medicago* MFS Transporters

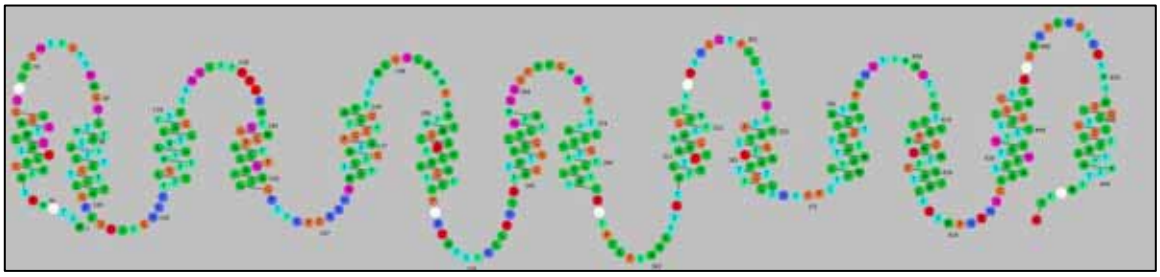
Four predicted genes were identified in *Medicago truncatula* with homology to *GmMFS1.3* (Figure 4.2). The expression of these genes was assessed using the MtGEA (http://mtgea.noble.org/v2/index.php) transcriptome database (Benedito et al., 2008). Both *MtMFS1* and *MtMFS2* are expressed exclusively in above ground organs (Figure 4.16). *MtMFS1* is enriched in leaves, stems, buds, flowers, and seed pods, while *MtMFS2* is expressed during seed development, but at a relatively low level (not visible with scale shown). *MtMFS3* is expressed in all organs at a similar expression level. *MtMFS4* shows the highest overall expression, being enriched in seeds, nitrogen-starved roots, and nodules. *MtMFS3*, being the only MFS transporter both linked with a SAT-like sequence and expressed in roots and nodules, was chosen for promoter-GUS analysis. Similar to *GmMFS1.3*, *MtMFS3* is expressed in the root stele (Figure 4.17). Upon inoculation with rhizobia, the *MtMFS3* promoter was active in developing nodules, as well as the root stele.

Cross-sections of GUS-positive nodules revealed localization to uninfected nodule cells, as well as the nodule inner cortex.

A



B



C

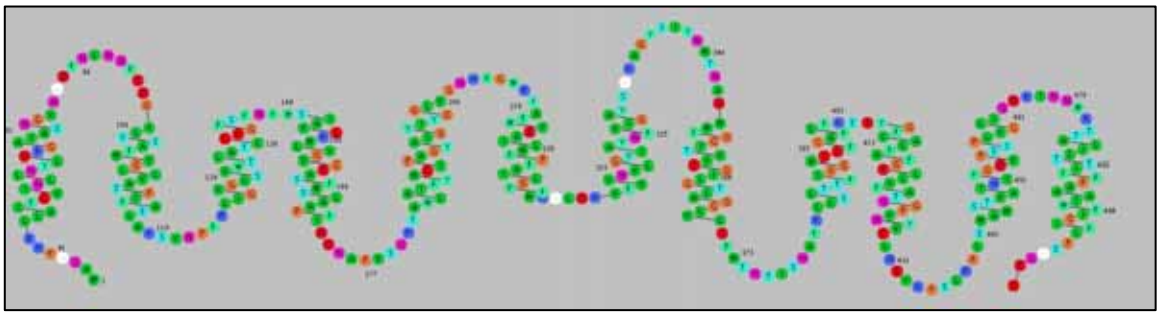


Figure 4.1. Comparison of sequence and topology of YOR378W with GmMFS1.3.
A, Pairwise MUSCLE alignment of YOR378W and GmMFS1.3 amino acid sequences generated with Geneious software. The RbDe (Residue-based Diagram editor) web application (http://icb.med.cornell.edu/crt/RbDe/) was used to generate a two-dimensional topology prediction of YOR378W (**B**) and GmMFS1.3 (**C**). RbDe uses HMMTOP to predict transmembrane helices (Konvicka et al., 2000; Tusnady and Simon, 2001).

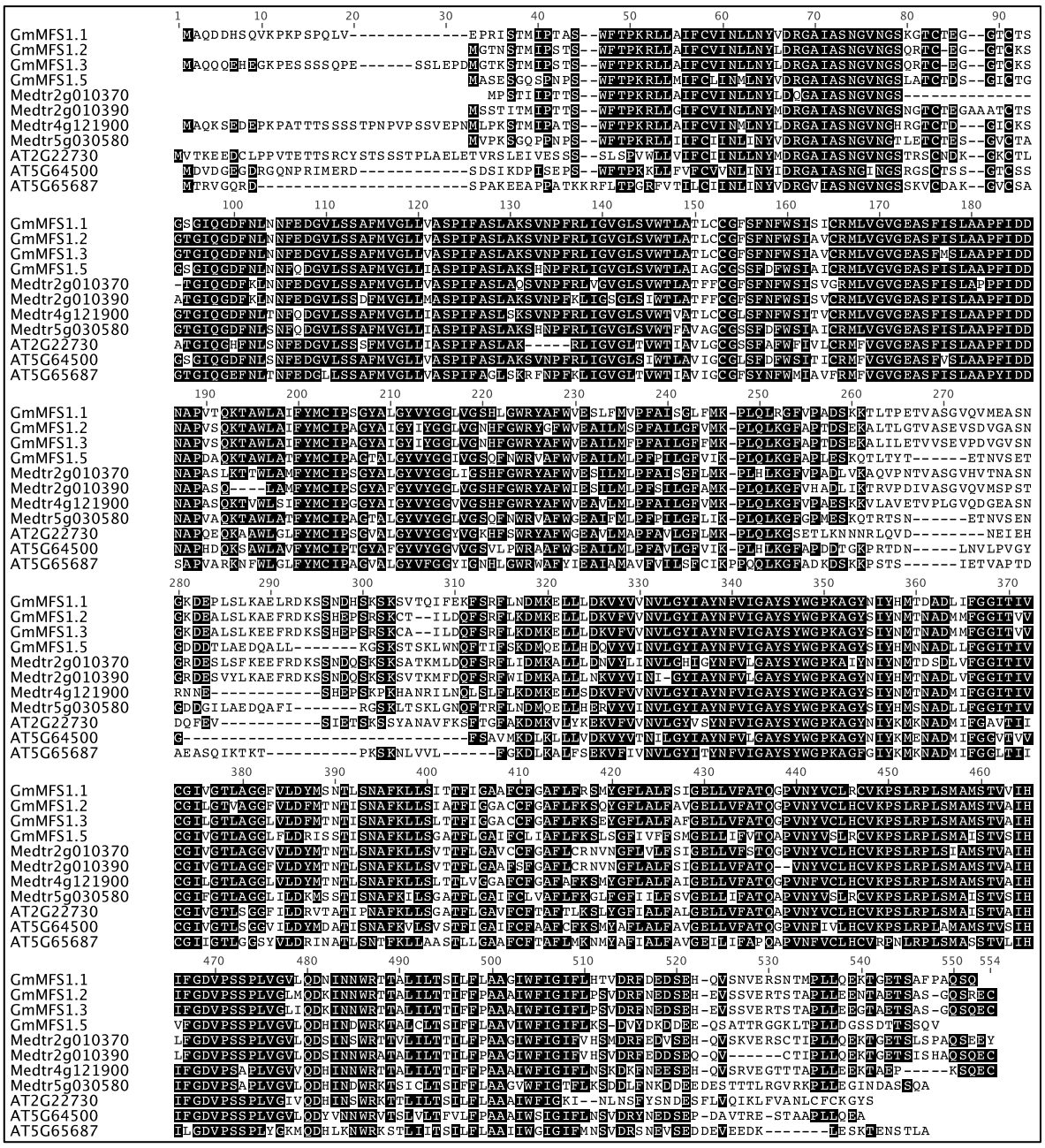


Figure 4.2. Alignment of the identified major facilitator subfamily members.
Multiple MUSCLE alignment of the MFS amino acid sequences as generated with Geneious software. GmMFS1.1 (Glyma13g32670), GmMFS1.2 (Glyma07g30370), GmMFS1.3 (Glyma08g06880), GmMFS1.5 (Glyma09g33680) are shown along with MFS proteins identified in *Medicago* and *Arabidopsis*. GmMFS1.4 was omitted due to an incomplete sequence prediction.

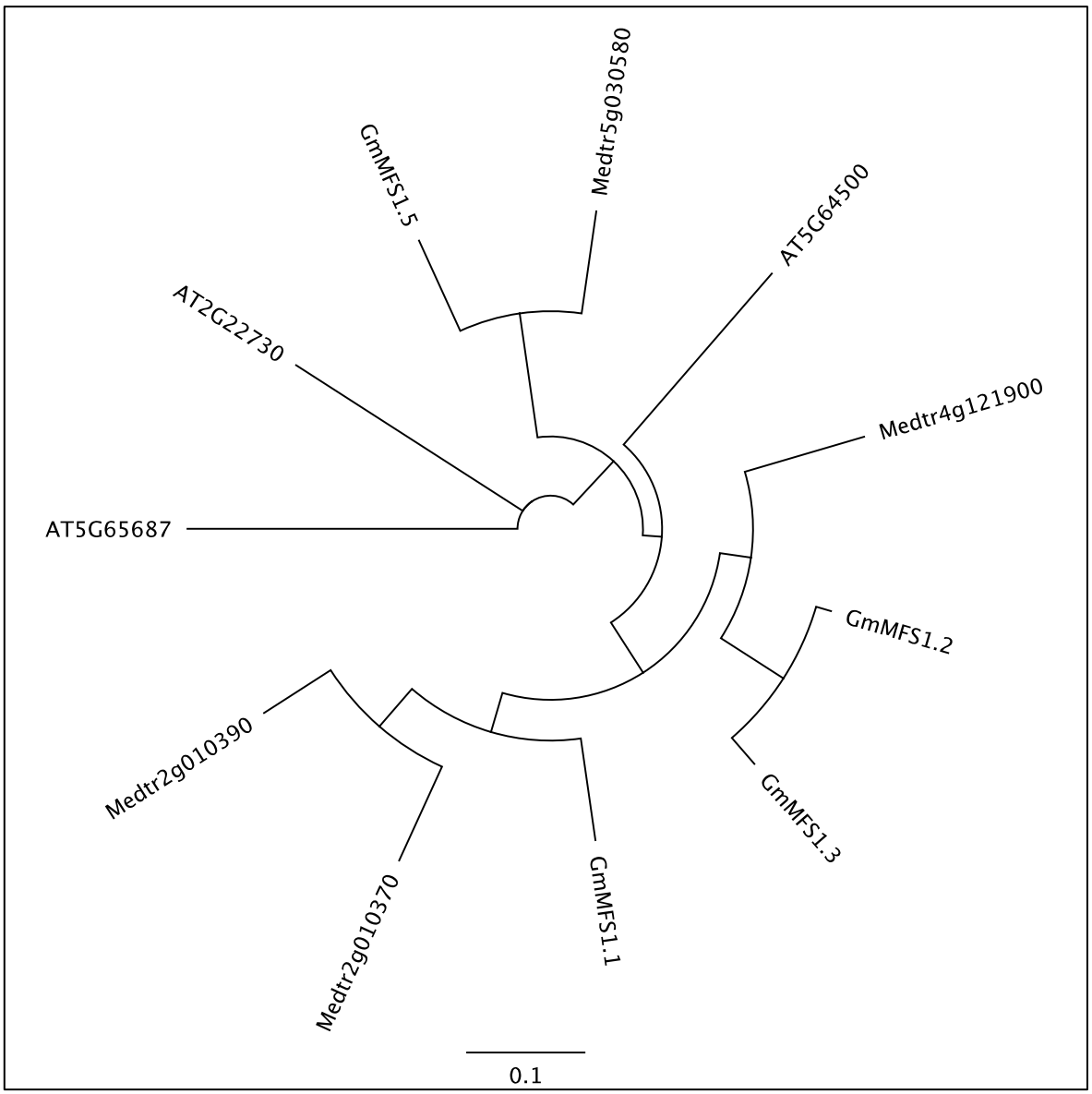


Figure 4.3. Circular phylogenetic tree of MFS transporters.
Analysis of soybean, *Medicago* and *Arabidopsis* MFS transporter proteins sequences. The neighbor-joining tree was generated using the program Geneious, using the Jukes-Cantor distance model. Scale represents the number of amino acid substitutions per site.

A

```
ATGGCACAACAACAAGAACATGAAGGAAAACCAGAGTCTTCTTCCTCTCAACCTGAGTCTTCTTTGGAACCAGACATGGGCACCAAGTCCACAAT
 M  A  Q  Q  Q  E  H  E  G  K  P  E  S  S  S  S  Q  P  E  S  S  L  E  P  D  M  G  T  K  S  T  M

GATTCCGTCAACTTCATGGTTTACACCAAAGAGGTTACTGGCTATATTTTGTGCGATCAACTTGTTAAACTATTTGGACCGAGGAGCGATTGCAA
  I  P  S  T  S  W  F  T  P  K  R  L  L  A  I  F  C  A  I  N  L  L  N  Y  L  D  R  G  A  I  A

GCAATGGGGTTAATGGAAGTCAACGAACTTGTGAAGGCGGTACCTGCAAATCTGGAACAGGAATACAGGGGGATTTTAACTTGAACAATTTTGAG
S  N  G  V  N  G  S  Q  R  T  C  E  G  G  T  C  K  S  G  T  G  I  Q  G  D  F  N  L  N  N  F  E

GATGGAGTTCTATCATCTGCTTTTATGGTTGGACTTCTTGTGGCTTCTCCAATATTTGCTTCTCTAGCAAAGAGCGTAAACCCATTTAGACTAAT
 D  G  V  L  S  S  A  F  M  V  G  L  L  V  A  S  P  I  F  A  S  L  A  K  S  V  N  P  F  R  L  I

TGGAGTGGGATTATCAGTTTGGACCCTTGCAACCTTATGTTGTGGTTTCTCTTTCAATTTCTGGTCCATTGCAGTCTGTCGCATGCTAGTTGGTG
  G  V  G  L  S  V  W  T  L  A  T  L  C  C  G  F  S  F  N  F  W  S  I  A  V  C  R  M  L  V  G

TTGGTGAGGCTTCATTTATGAGTCTTGCAGCACCTTTCATTGATGACAATGCCCCAGTTTCACAGAAAACGGCATGGCTTGCTATATTTTACATG
V  G  E  A  S  F  M  S  L  A  A  P  F  I  D  D  N  A  P  V  S  Q  K  T  A  W  L  A  I  F  Y  M

TGTATACCAGCAGGATATGCAATTGGCTACATCTATGGTGGTTTGGTTGGAAACCATTTTGGTTGGCGTTATGCGTTTTGGGTGGAAGCTATATT
 C  I  P  A  G  Y  A  I  G  Y  I  Y  G  G  L  V  G  N  H  F  G  W  R  Y  A  F  W  V  E  A  I  L

GATGTTTCCATTTGCTATTTTGGGATTTTTTATGAAGCCTTTGCAGTTAAAAGGTTTTGCCCCTACTGATTCAGAAAAGGCACTGATACTGGAGA
  M  F  P  F  A  I  L  G  F  F  M  K  P  L  Q  L  K  G  F  A  P  T  D  S  E  K  A  L  I  L  E

CAGTGGTATCAGAAGTTCCAGATGTCGGGGTTTCAAATGGAAAAGATGAAGCGTTGTCCTTGAAGGAAGAGTTCAGAGACAAAAGCTCACATGAA
T  V  V  S  E  V  P  D  V  G  V  S  N  G  K  D  E  A  L  S  L  K  E  E  F  R  D  K  S  S  H  E

CCTTCCAGGTCAAAATGTGCAATATTAGATCAGTTCTCAAGATTTCTGAAAGACATGAAAGAACTTTTGCTTGATAAGGTTTTTGTTGTCAATGT
 P  S  R  S  K  C  A  I  L  D  Q  F  S  R  F  L  K  D  M  K  E  L  L  L  D  K  V  F  V  V  N  V

TCTAGGTTACATAGCATACAACTTTGTAATAGGTGCTTACTCATATTGGGGCCCCAAAGCTGGTTATAGTATATATAACATGACTAATGCAGATA
  L  G  Y  I  A  Y  N  F  V  I  G  A  Y  S  Y  W  G  P  K  A  G  Y  S  I  Y  N  M  T  N  A  D

TGATGTTTGGAGGAATTACTGTTGTATGTGGGATATTGGGCACCTTAGCAGGAGGTCTTGTTCTTGATTTCATGACTAACACAATATCAAATGCA
M  M  F  G  G  I  T  V  V  C  G  I  L  G  T  L  A  G  G  L  V  L  D  F  M  T  N  T  I  S  N  A

TTTAAGCTTCTCTCATTAACAACATTTATTGGTGGTGCATGCTGTTTTGGTGCTTTCTTATTTAAAAGCGAGTATGGCTTCCTTGCCCTTTTTGC
 F  K  L  L  S  L  T  T  F  I  G  G  A  C  C  F  G  A  F  L  F  K  S  E  Y  G  F  L  A  L  F  A

TTTTGGTGAACTACTTGTATTTGCCACTCAGGGTCCTGTGAATTATGTATGTCTCCATTGTGTTAAACCAAGTTTGAGGCCGCTGTCTATGGCTA
  F  G  E  L  L  V  F  A  T  Q  G  P  V  N  Y  V  C  L  H  C  V  K  P  S  L  R  P  L  S  M  A

TGTCTACTGTTGCTATTCATATCTTTGGAGATGTGCCTTCCTCACCTCTCGTTGGACTTATCCAGGACAAAATTAATAACTGGAGAACGACTGCA
M  S  T  V  A  I  H  I  F  G  D  V  P  S  S  P  L  V  G  L  I  Q  D  K  I  N  N  W  R  T  T  A

TTGATTCTAACAACTATATTCTTTCCAGCAGCTGCAATATGGTTTATAGGAATATTTTTGCCCAGTGTGGATAGATTTAATGAAGATAGTGAGCA
 L  I  L  T  T  I  F  F  P  A  A  A  I  W  F  I  G  I  F  L  P  S  V  D  R  F  N  E  D  S  E  H

TGAAGTATCAAGCGTGGAAAGGACAAGCACAGCACCATTGCTTGAAGAGGGCACTGCTGAAACATCAGCATCTGGTCAATCCCAAGAATGCTGA
  E  V  S  S  V  E  R  T  S  T  A  P  L  L  E  E  G  T  A  E  T  S  A  S  G  Q  S  Q  E  C  *
```

B

```
                1                10                20                30                40                50
GmMFS1.3        MAQQQEHEGKPESSSSQPESSLEPDMGTKSTMIPSTSWFTPKRLLAIFCVINLLNY
GmMFS1.3 Prediction MAQQQEHEGKPESSSSQPESSLEPDMGTKSTMIPSTSWFTPKRLLAIFCVINLLNY
                60                70                80                90               100               110
GmMFS1.3        LDRGAIASNGVNGSQRTCEGGTCKSGTGIQGDFNLNNFEDGVLSSAFMVGLLVASP
GmMFS1.3 Prediction LDRGAIASNGVNGSQRTCEGGTCKSGTGIQGDFNLNNFEDGVLSSAFMVGLLVASP
               120               130               140               150               160
GmMFS1.3        IFASLAKSVNPFRLIGVGLSVWTLATLCCGFSFNFWSIAVCRMLVGVGEASFMSLA
GmMFS1.3 Prediction IFASLAK-----RLIGVGLSVWTLATLCCGFSFNFWSIAVCRMLVGVGEASFMSLA
               170               180               190               200               210               220
GmMFS1.3        APFIDDNAPVSQKTAWLAIFYMCIPAGYAIGYIYGGLVGNHFGWRYAFWVEAILMF
GmMFS1.3 Prediction APFIDDNAPVSQKTAWLAIFYMCIPAGYAIGYIYGGLVGNHFGWRYAFWVEAILMF
               230               240               250               260               270               280
GmMFS1.3        PFAILGFFMKPLQLKGFAPTDSEKALILETVVSEV--------PDVGVSNGKDEAL
GmMFS1.3 Prediction PFAILGFFMKPLQLKG---AHSSPALISCSSITSFPITKLLVHPDVGVSNGKDEAL
               290               300               310               320               330
GmMFS1.3        SLKEEFRDKSSHEPSRSKCAILDQFSRFLKDMKELLLDKVFVVNVLGYIAYNFVIG
GmMFS1.3 Prediction SLKEEFRDKSSHEPSRSKCAILDQFSRFLKDMKELLLDKVFVVNVLGYIAYNFVIG
               340               350               360               370               380               390
GmMFS1.3        AYSYWGPKAGYSIYNMTNADMMFGGITVVCGILGTLAGGLVLDFMTNTISNAFKLL
GmMFS1.3 Prediction AYSYWGPKAGYSIYNMTNADMMFGGITVVCGILGTLAGGLVLDFMTNTISNAFKLL
               400               410               420               430               440
GmMFS1.3        SLTTFIGGACCFGAFLFKSEYGFLALFAFGELLVFATQGPVNYVCLHCVKPSLRPL
GmMFS1.3 Prediction SLTTFIGGACCFGAFLFKSEYGFLALFAFGELLVFATQGPVNYVCLHCVKPSLRPL
               450               460               470               480               490               500
GmMFS1.3        SMAMSTVAIHIFGDVPSSPLVGLIQDKINNWRTTALILTTIFFPAAAIWFIGIFLP
GmMFS1.3 Prediction SMAMSTVAIHIFGDVPSSPLVGLIQDKINNWRTTALILTTIFFPAAAIWFIGKSL
               510               520               530               540   545
GmMFS1.3        SVDRFNEDSEHEVSSVERTSTAPLLEEGTAETSASGQSQEC
GmMFS1.3 Prediction
```

Figure 4.4. ***GmMFS1.3*** **sequence and translation.**
A, Coding sequence and translation of *GmMFS1.3* as determined after cloning. **B,** Pairwise alignment of the GmMFS1.3 amino acid sequence with the soybean genome prediction (Glyma08g06880, incomplete) as generated by Geneious software.

A

```
ATGGCATCGGAGTCAGGTCAAAGTCCCAATCCTTCTTGGTTTACGCCCAAAAGGCTTCTTATGATATTTTGTCTCATTAACATGCTAAACTATG
 M  A  S  E  S  G  Q  S  P  N  P  S  W  F  T  P  K  R  L  L  M  I  F  C  L  I  N  M  L  N  Y

TGGATCGAGGAGCTATAGCCAGTAACGGTGTCAATGGAAGTCTTGCAACTTGCACTGACTCTGGAATTTGTACAGGTGGCAGTGGAATTCAGGG
V  D  R  G  A  I  A  S  N  G  V  N  G  S  L  A  T  C  T  D  S  G  I  C  T  G  G  S  G  I  Q  G

GGATTTTAACTTGAACAATTTTCAAGATGGAGTTCTGTCATCTGCATTTATGGTTGGACTTCTAATTGCTTCCCCAATATTTGCTTCTCTGGCC
  D  F  N  L  N  N  F  Q  D  G  V  L  S  S  A  F  M  V  G  L  L  I  A  S  P  I  F  A  S  L  A

AAAAGTCACAATCCATTCCGGCTTATTGGTGTTGGATTGTCTGTTTGGACTTTAGCGATAGCTGGATGTGGTAGTTCGTTCGATTTTTGGTCTA
 K  S  H  N  P  F  R  L  I  G  V  G  L  S  V  W  T  L  A  I  A  G  C  G  S  S  F  D  F  W  S

TTGCAATATGCCGAATGCTAGTTGGTGTAGGTGAGGCTTCCTTCATAAGTCTTGCCGCACCATTTATAGATGACAATGCCCCTGATGCACAGAA
I  A  I  C  R  M  L  V  G  V  G  E  A  S  F  I  S  L  A  A  P  F  I  D  D  N  A  P  D  A  Q  K

AACAGCATGGCTTGCTACATTTTACATGTGTATACCAGCTGGGACTGCTCTGGGCTATGTTTATGGTGGAATTGTTGGAAGCCAATTTAACTGG
  T  A  W  L  A  T  F  Y  M  C  I  P  A  G  T  A  L  G  Y  V  Y  G  G  I  V  G  S  Q  F  N  W

CGTGTTGCATTTTGGGTTGAGGCAATTTTGATGCTTCCTTTTCCCATTTTGGGTTTTGTAATAAAGCCTTTGCAATTGAAAGGTTTTGCACCCT
 R  V  A  F  W  V  E  A  I  L  M  L  P  F  P  I  L  G  F  V  I  K  P  L  Q  L  K  G  F  A  P

TGGAATCCAAACAAACACTAACATATACTGAAACAAATGTTTCAGAAACTGGAGATGATGACACGCTTGCTGAAGATCAAGCCTTGCTCAAAGG
L  E  S  K  Q  T  L  T  Y  T  E  T  N  V  S  E  T  G  D  D  D  T  L  A  E  D  Q  A  L  L  K  G

ATCCAAGTCAACATCCAAATTGTGGAATCAGTTCACCATCTTCTCCAAAGATATGCAGGAGCTCTTGCATGACCAAGTGTATGTTATAAATGTT
  S  K  S  T  S  K  L  W  N  Q  F  T  I  F  S  K  D  M  Q  E  L  L  H  D  Q  V  Y  V  I  N  V

CTAGGGTACATAGCATACAATTTTGTCATTGGAGCTTACTCATACTGGGGCCCAAAGGCAGGATACAGCATTTACCATATGAATAATGCTGACT
 L  G  Y  I  A  Y  N  F  V  I  G  A  Y  S  Y  W  G  P  K  A  G  Y  S  I  Y  H  M  N  N  A  D

TGTTATTTGGAGGCATTACAATTGTTTGTGGTATTGTTGGGACATTAGCTGGAGGCTTGTTTCTTGACAGGATATCTTCAACCATCTCTAATGC
L  L  F  G  G  I  T  I  V  C  G  I  V  G  T  L  A  G  G  L  F  L  D  R  I  S  S  T  I  S  N  A

TTTCAAGCTTCTTTCTGGAGCAACATTTCTTGGTGCAATCTTCTGCTTGATCGCTTTCCTTTTCAAGAGCTTGTCTGGTTTCATAGTCTTCTTC
  F  K  L  L  S  G  A  T  F  L  G  A  I  F  C  L  I  A  F  L  F  K  S  L  S  G  F  I  V  F  F

TCTATGGGTGAACTACTGATTTTTGTCACTCAGGCTCCTGTAAATTATGTTTCTCTCCGTTGTGTAAAACCTAGTTTGAGGCCACTGTCTATGG
 S  M  G  E  L  L  I  F  V  T  Q  A  P  V  N  Y  V  S  L  R  C  V  K  P  S  L  R  P  L  S  M

CTATATCTACTGTTTCAATCCATGTCTTTGGTGATGTGCCTTCCTCACCACTTGTTGGGGTCCTGCAGGATCATATCAACGACTGGAGGAAAAC
A  I  S  T  V  S  I  H  V  F  G  D  V  P  S  S  P  L  V  G  V  L  Q  D  H  I  N  D  W  R  K  T

AGCACTTTGTTTAACATCCATTTTCTTTCTTGCTGCTGTAATATGGTTCATAGGAATCTTTTTGAAAAGTGATGTATATGACAAAGATGATGAG
  A  L  C  L  T  S  I  F  F  L  A  A  V  I  W  F  I  G  I  F  L  K  S  D  V  Y  D  K  D  D  E

GAGCAATCTGCCACAACTAGAGGAGGAAAACTGACGCCATTGCTTGACGGCAGCAGTGACACTACATCCAGTCAAGTTTAA
 E  Q  S  A  T  T  R  G  G  K  L  T  P  L  L  D  G  S  S  D  T  T  S  S  Q  V  *
```

B

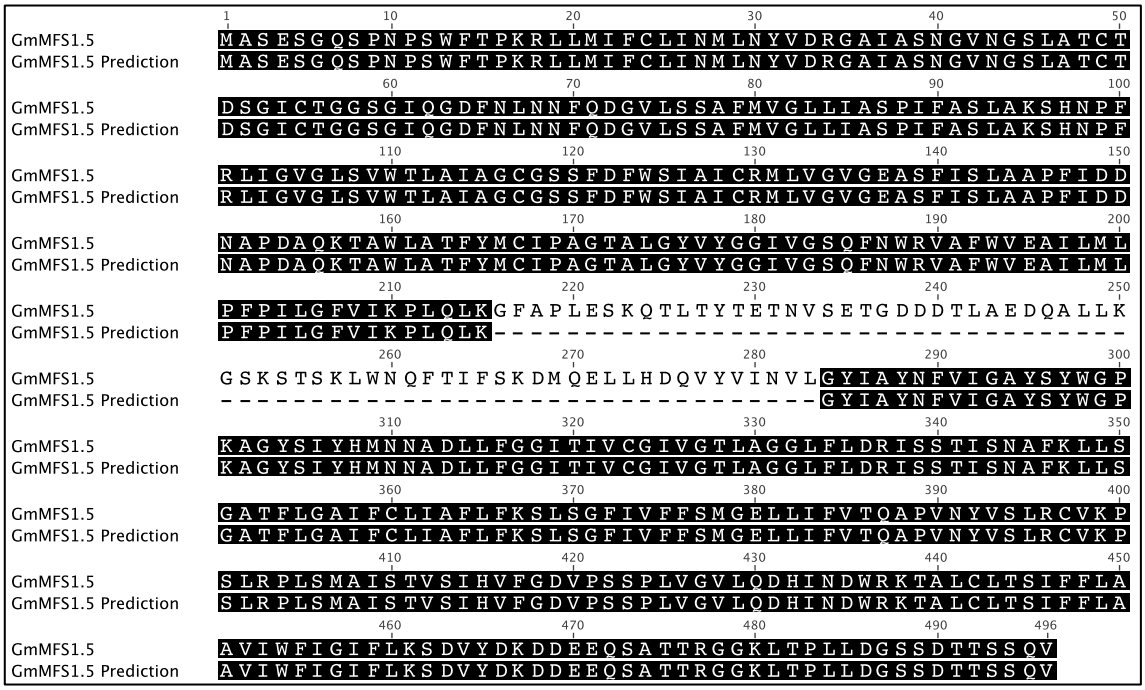


Figure 4.5. GmMFS1.5 sequence and translation.
A, Coding sequence and translation of *GmMFS1.5* as determined after cloning. **B,** Pairwise alignment of the GmMFS1.5 amino acid sequence with the soybean genome prediction (Glyma09g33680, missing an exon) as generated by Geneious software.

Figure 4.6. Conserved relative location of *MFS* and *SAT* loci in plant genomes.
Location of *GmSAT1* (15g06680, shown in red) and *GmMFS1.4* (15g06660, shown in blue) from soybean chromosome 15 relative to (**A)** *Medicago truncatula* chromosome 2 **(B**) *Arabidopsis thaliana* chromosome 2 and (**C**) *Solanum lycopersicum* (tomato) chromosome 2. Data was obtained from the Plant Genome Duplication Database (http://chibba.agtec.uga.edu/duplication/index/home) and redrawn.

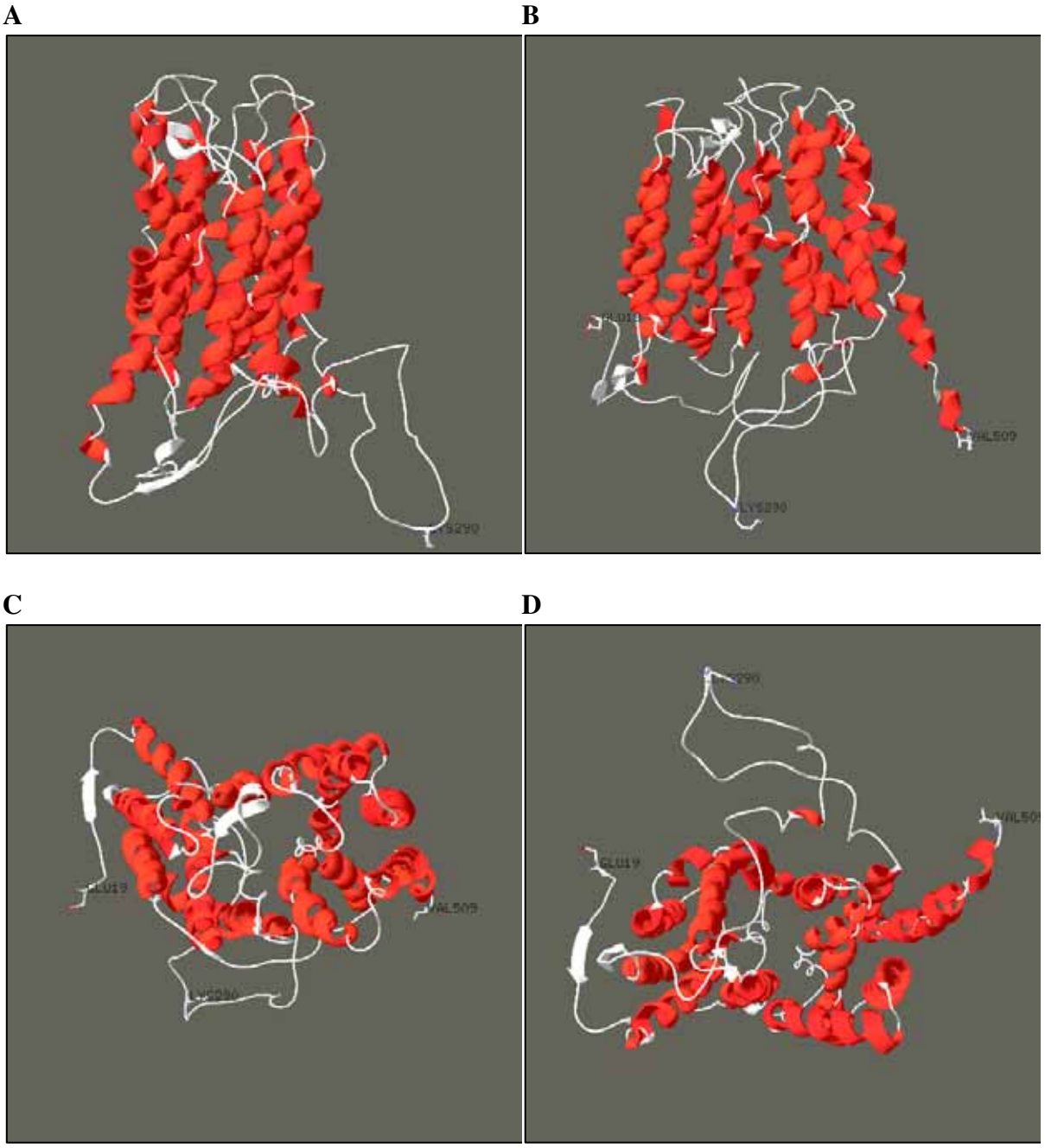


Figure 4.7. ***De novo*** **prediction of GmMFS1.3 structure.**
GmMFS1.3 predicted structure as viewed from the side (**A**, and **B**), overhead (**C**), and below (**D**). Residues (E19, K290, V509) are indicated for reference. The prediction was generated using the HMM-based Protein Structure Prediction, SAM-T08 (http://compbio.soe.ucsc.edu/SAM_T08/T08-query.html) and viewed with Swiss-PdbViewer (available from http://spdbv.vital-it.ch/).

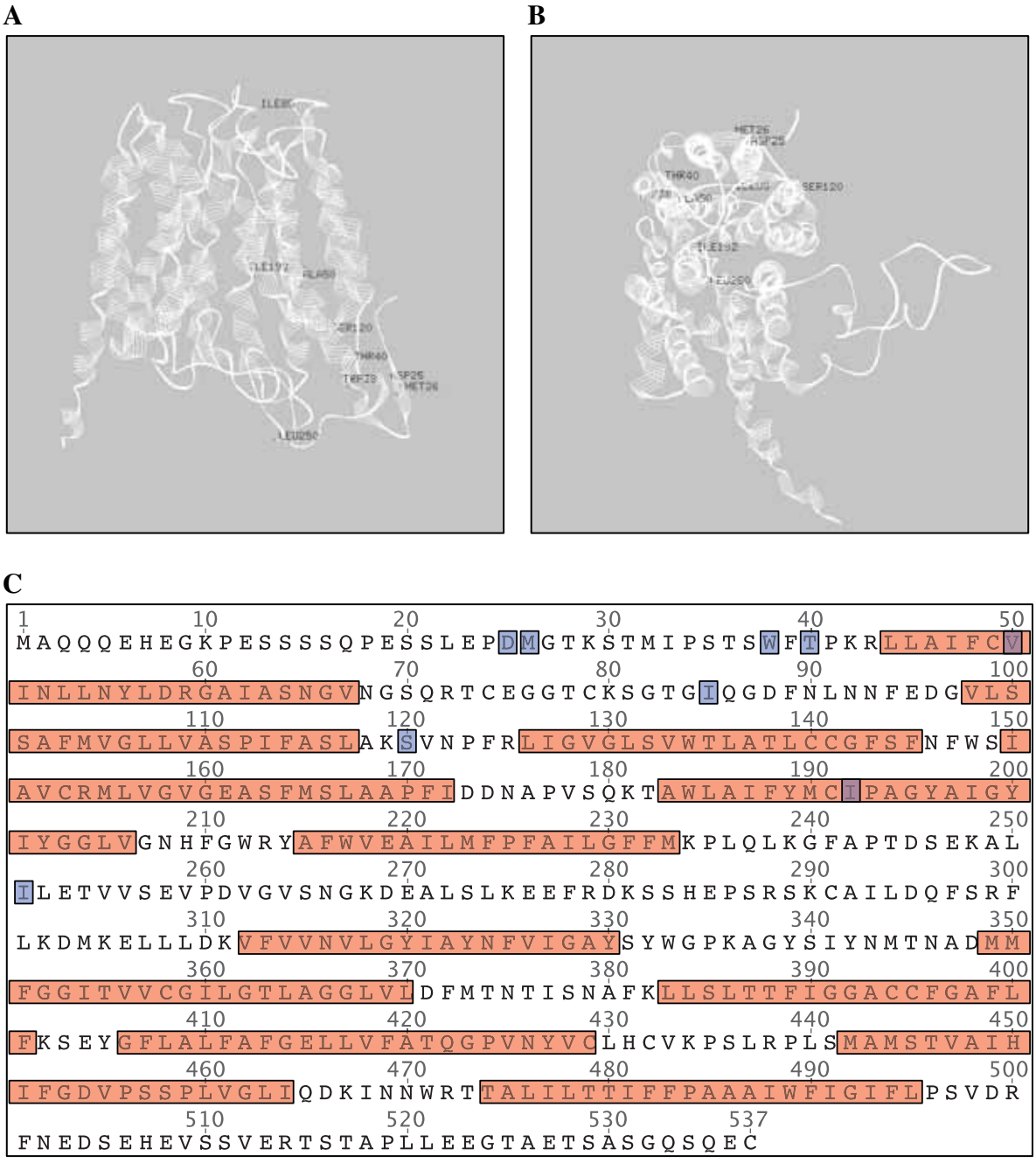


Figure 4.8. Location of isolated mutations in GmMFS1.3.
A, and **B**, separate views of GmMFS1.3 model with highlighted mutations isolated during cloning in *E. coli*. **C**, Location of mutations in primary amino acid sequence. The mutations are D25G, M26T, W38R, T40I, V50A, I85M, S120G, I192V, and I251V (shown in blue). Also indicated are the predicted transmembrane helices (red) as generated by HMMTOP (http://www.enzim.hu/hmmtop/).

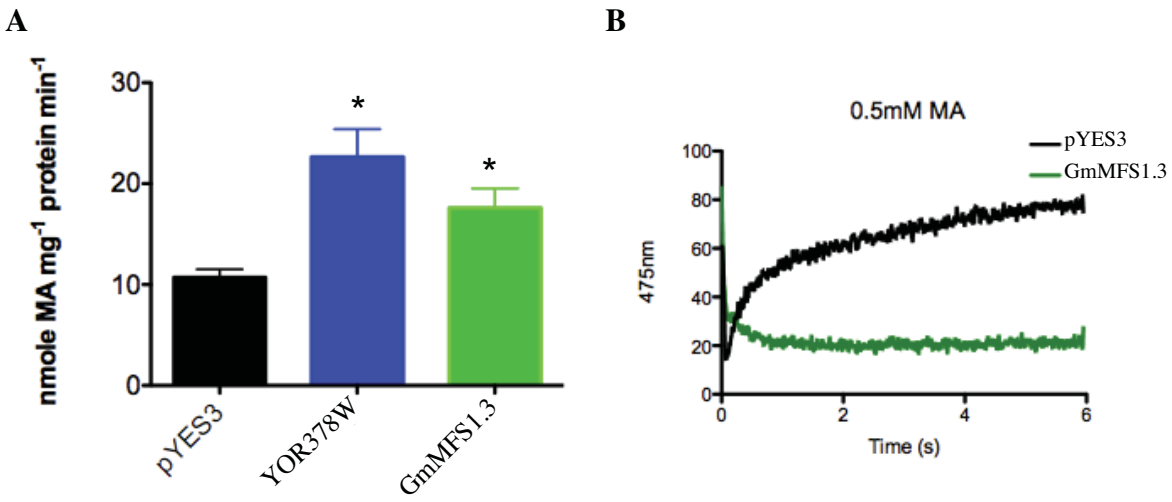
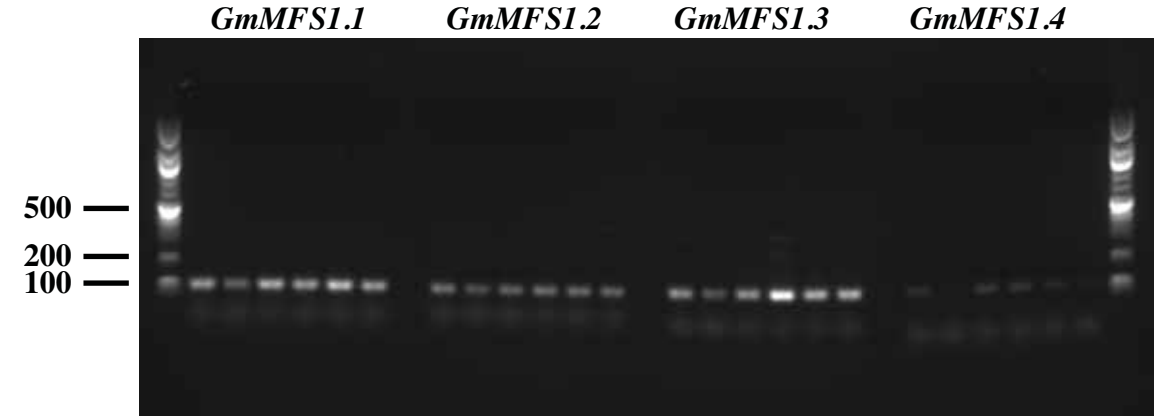


Figure 4.9. GmMFS1.3 transport of methylammonium in yeast.
A, Uptake of ^{14}C-methylammonium by the 26972c yeast strain expressing YOR378W-pYES3 or GmMFS1.3-pYES3 versus empty pYES3. Asterisks indicate significant uptake differences compared to control, as calculated by a student's t-test (p-value < 0.05). Yeast spheroplasts were generated from GmMFS1.3-pYES3 (**B**) and monitored for swelling in the presence of 0.5 mM methylammonium by stopped-flow spectroscopy. The uptake and swelling experiments were conducted by Danielle Mazurkiewicz.

A



B

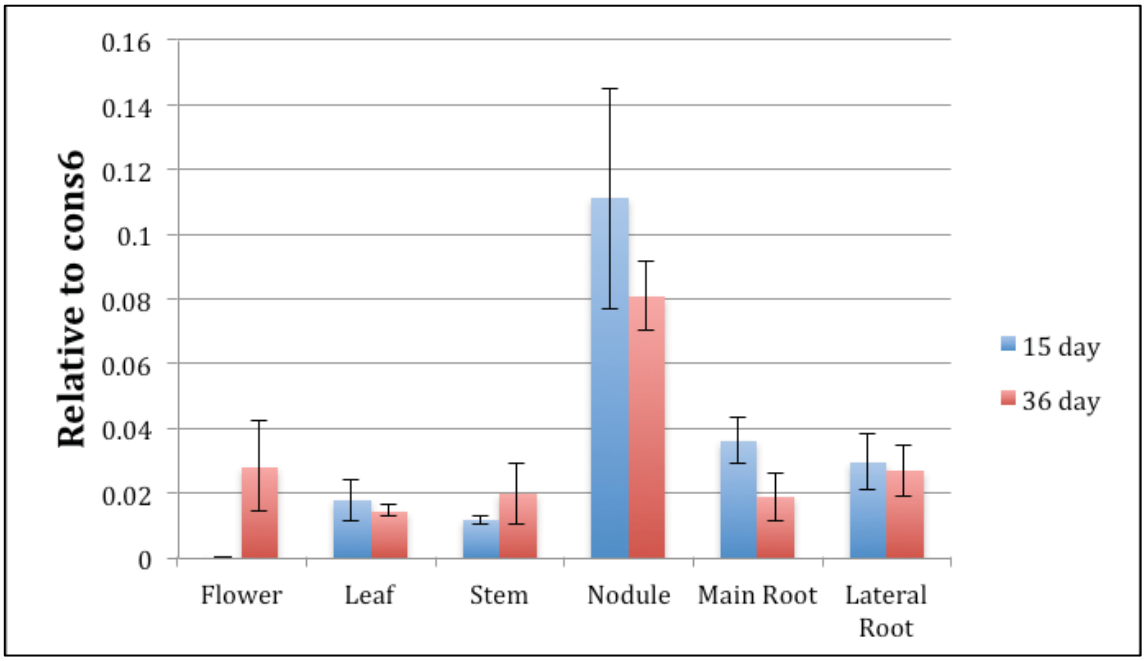


Figure 4.10. Expression of *GmMFS1.3* in soybean tissues.
A, RT-PCR analysis of *GmMFS1.1-1.4* in 36 day-old soybean tissues after 35 cycles using respective qPCR primers (Table 4.1). DNA ladder sizes are indicated on the left (bp). Tissues are, from left: flower, leaf, stem, nodule, main root, and lateral root. **B,** qPCR expression analysis of *GmMFS1.3* in 15 and 36 day-old soybean tissues inoculated with *Bradyrhizobium japonicum* USDA110 at planting. RNA was extracted from ten pooled plants grown under identical conditions. Data values represent the means of three independent technical replicates ±SD relative *cons6* (Libault et al., 2008), as calculated by the ΔCT method (Livak and Schmittgen, 2001).

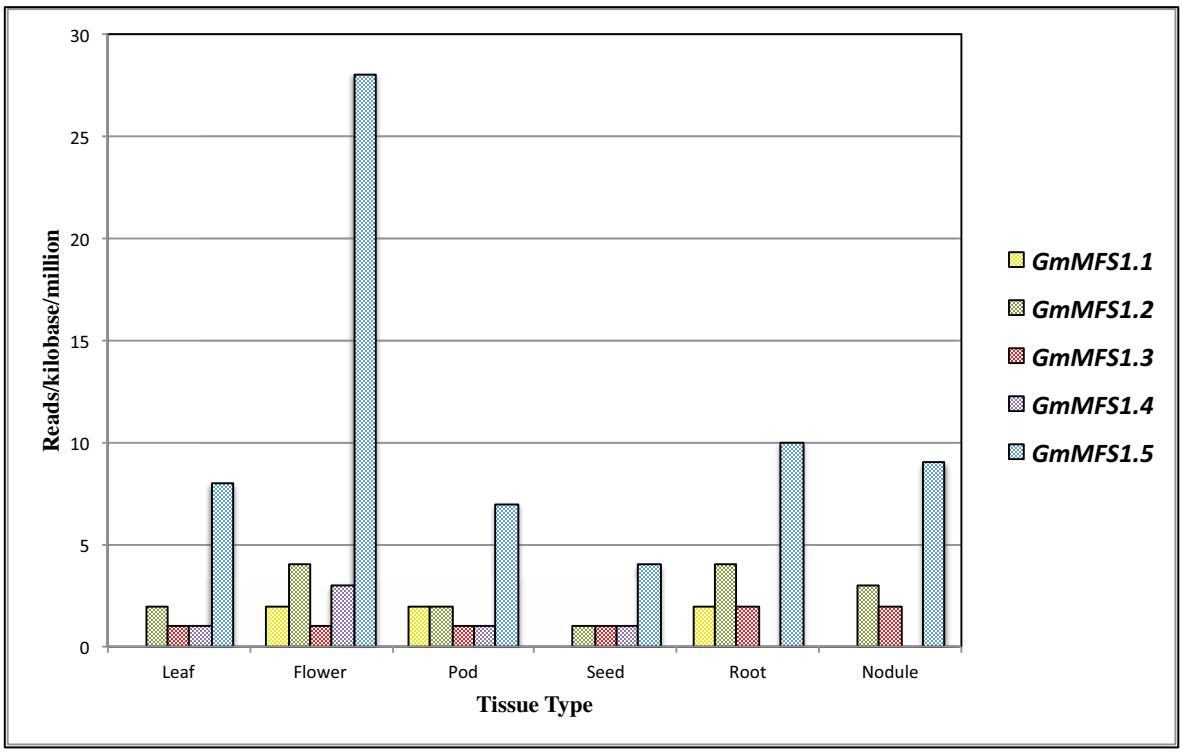
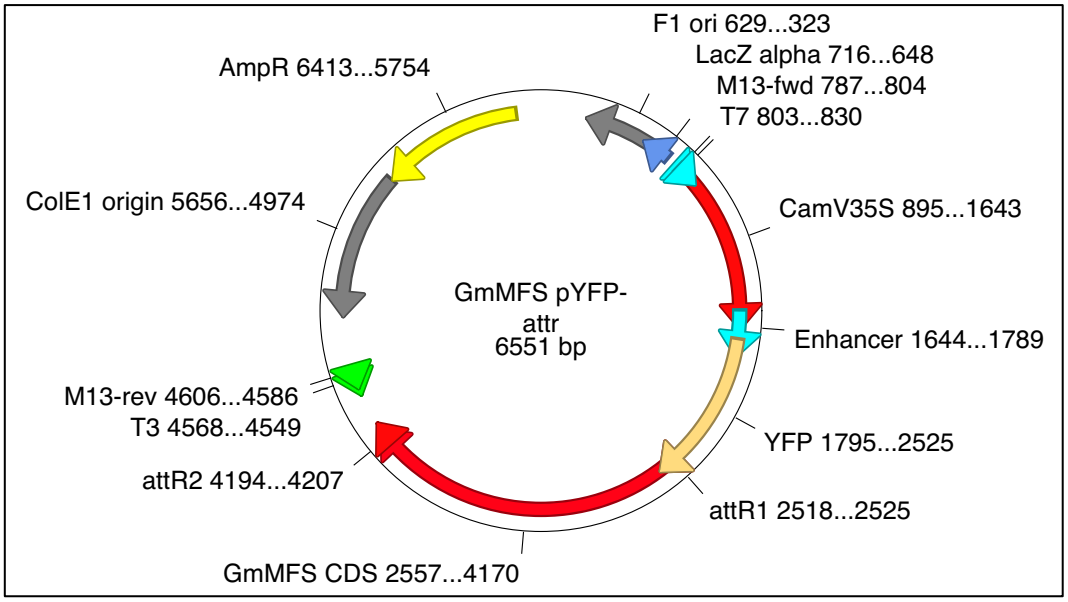


Figure 4.11. Expression of *GmMFS1.1-1.5* in soybean tissues by RNA-seq.
Expression data was obtained from the soybean RNA-seq transcriptome database (http://www.soybase.org/soyseq/#). Data represents the reads per kilobase of exon model per million mapped reads (RPKM), as generated by Severin et al. (2010).

A



B

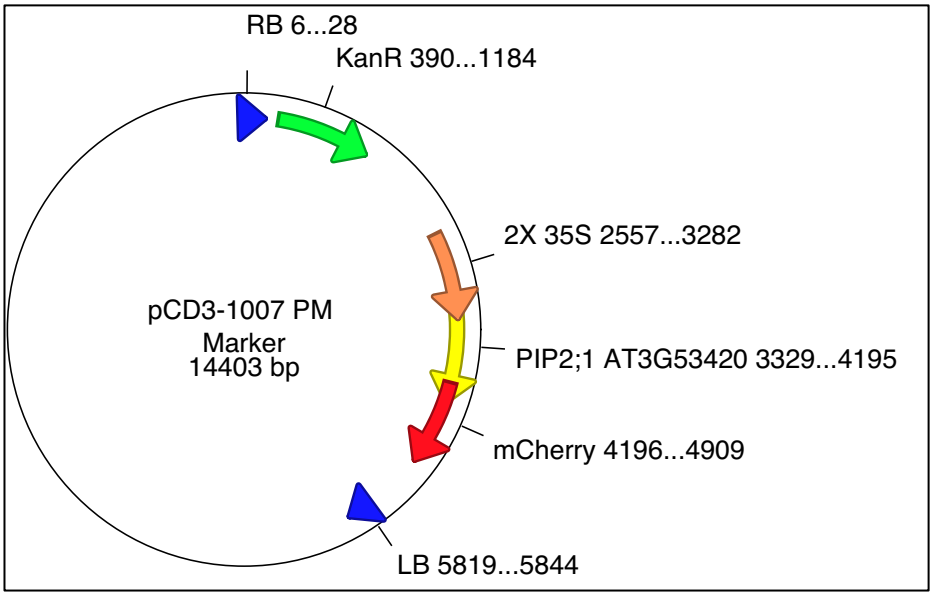
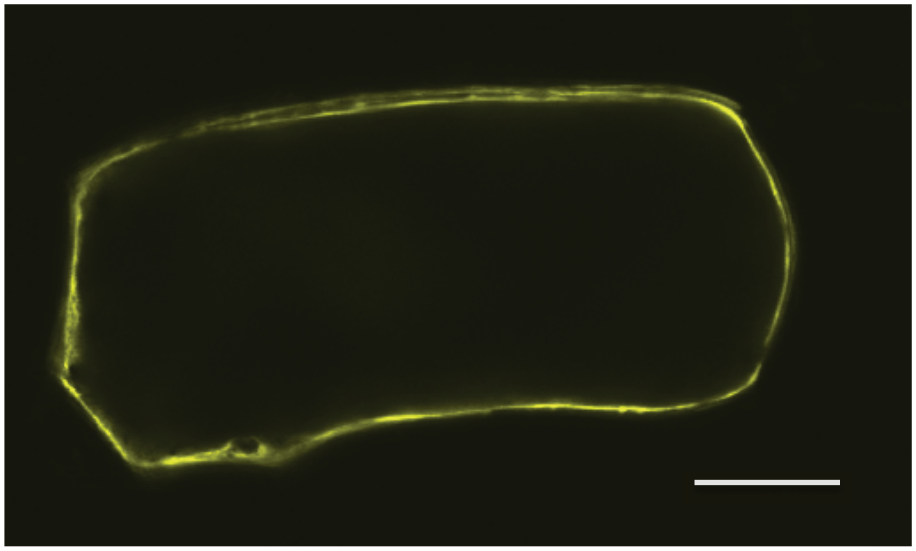


Figure 4.12. Expression plasmids used for onion localization.
A, pYFP-GmMFS1.3 plasmid used for N-terminal YFP fusion driven by a double 35S promoter. **B**, pCD3-1007 plasmid used to express the plasma membrane marker PIP2;1 fused C-terminal to mCherry.

A



B

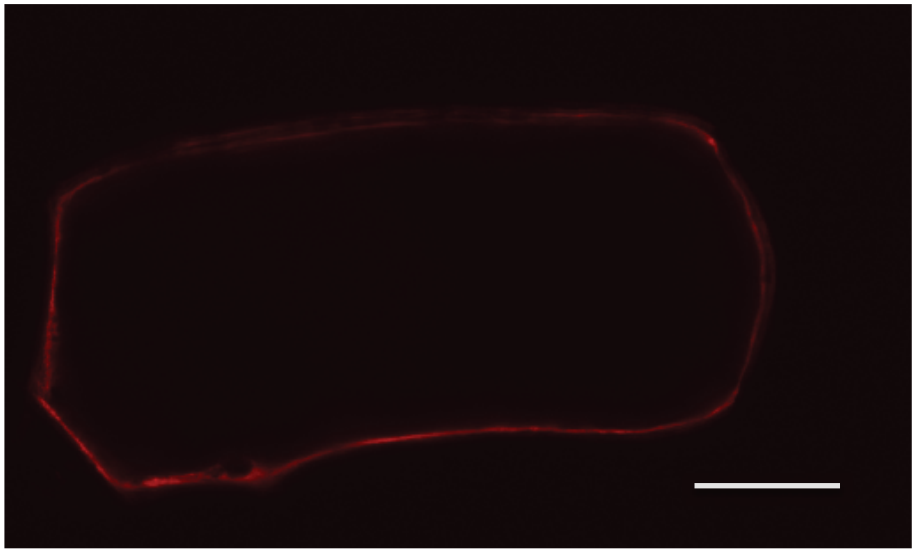


Figure 4.13. Localization of GmMFS1.3 in onion epidermal cells.
A, YFP-GmMFS1.3 and **B**, PIP2;1-mCherry expression in the same onion cell. Epidermal peels were co-bombarded with pYFP-GmMFS1.3 and pm-rk CD3-1007 plasmids, incubated for 24hr, then sequentially viewed for YFP (A) and mCherry (B) signals by confocal microscopy. Both fusions were found associated with the plasma membrane. Bar = 40 μm.

Figure 4.14. Plasmid used for promoter analysis of *GmMFS1.3* and *MtMFS3*.
The p243-RedRoot-GUS plasmid contains gateway destination sequences (R1 and R2) for recombination from an entry plasmid, as well as a dsRED marker driven by the *AtUBQ10* promoter for *in planta* selection (Limpens et al., 2005).

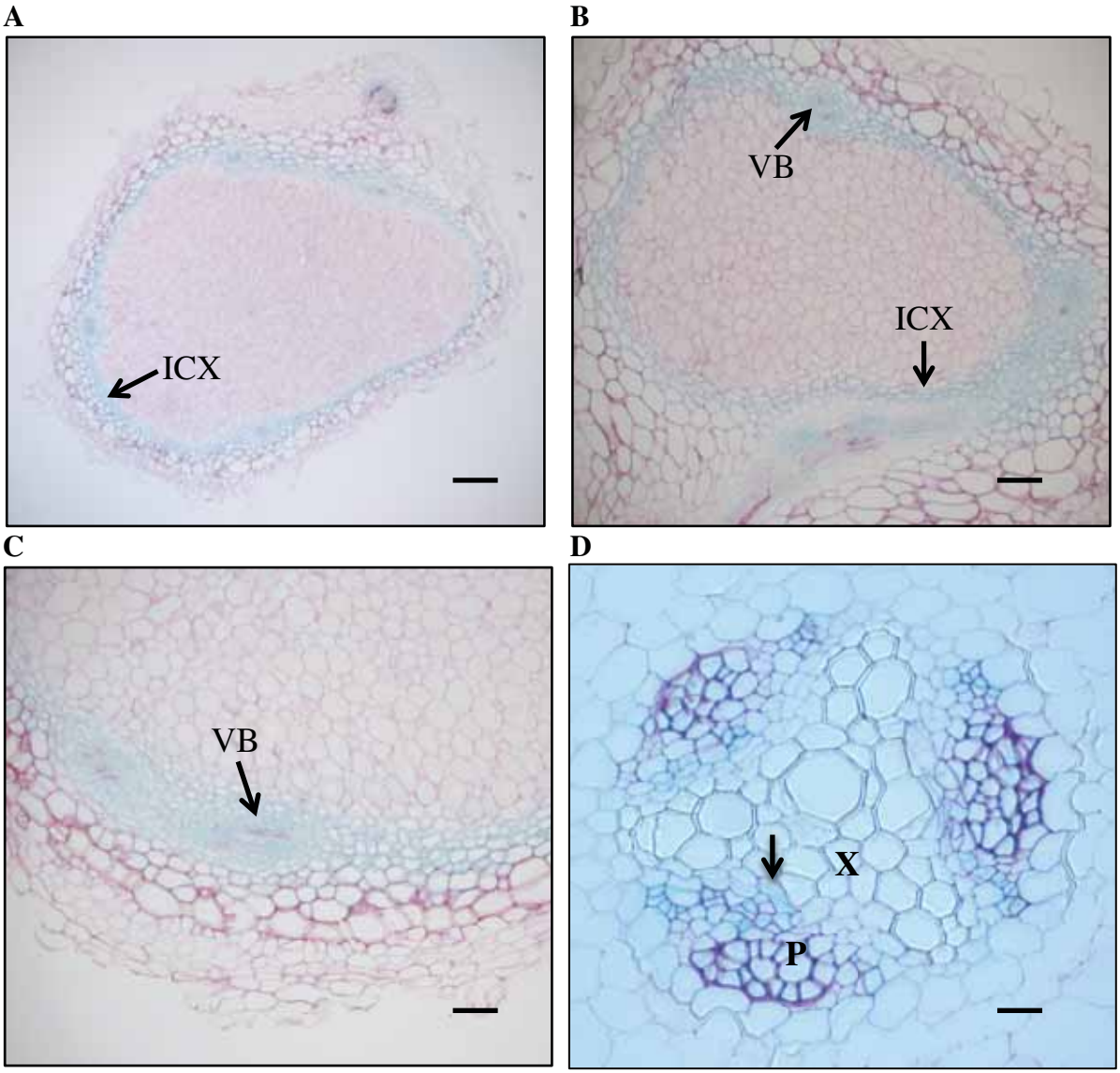


Figure 4.15. Activity of *GmMFS1.3* promoter in soybean nodules and roots.
Positive roots (identified by dsRED fluorescence) were stained for GUS activity and sectioned. Plants were analyzed 36 days (**A** and **C**) and 21 (**B** and **D**) days post-inoculation with rhizobia. **A,** and **B,** nodules showing expression in the inner cortex (ICX) and vascular bundles (VB). (**C**) Close-up of nodule showing expression in the inner cortex as well as surrounding vascular bundles (VB). (**D**) GUS activity detected in a root cross-section. Staining (arrow) was detected in the cell layer located between the xylem (X) and the phloem (P). Bar = 200 μm (A and B), 80 μm (C), 40 μm (D).

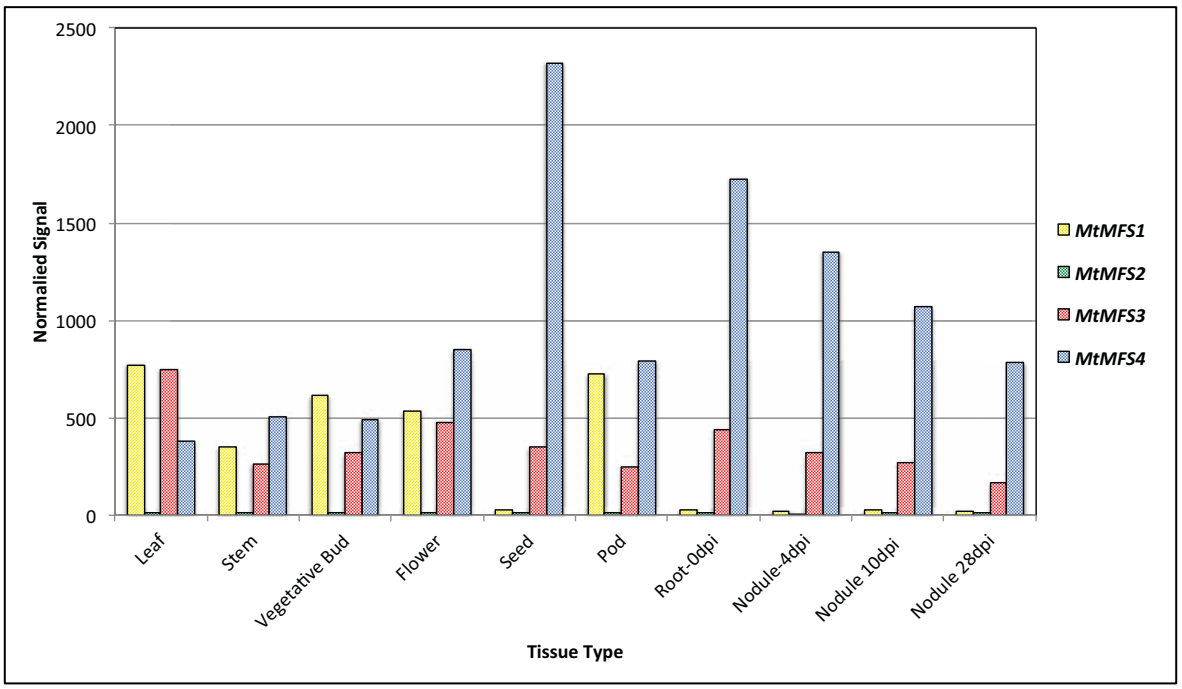


Figure 4.16. Expression of *Medicago MFS1-4* genes in various tissues.
Normalized expression of Medtr2g010370 (*MtMFS1*), Medtr2g010390 (*MtMFS2*), Medtr4g121900 (*MtMFS3*), and Medtr5g030580 (*MtMFS4*) as obtained from the *Medicago truncatula* gene atlas (MtGEA: http://mtgea.noble.org/v2/index.php). Data was acquired by DNA microarray analysis using the *Medicago* GeneChip (Ji et al., 2009).

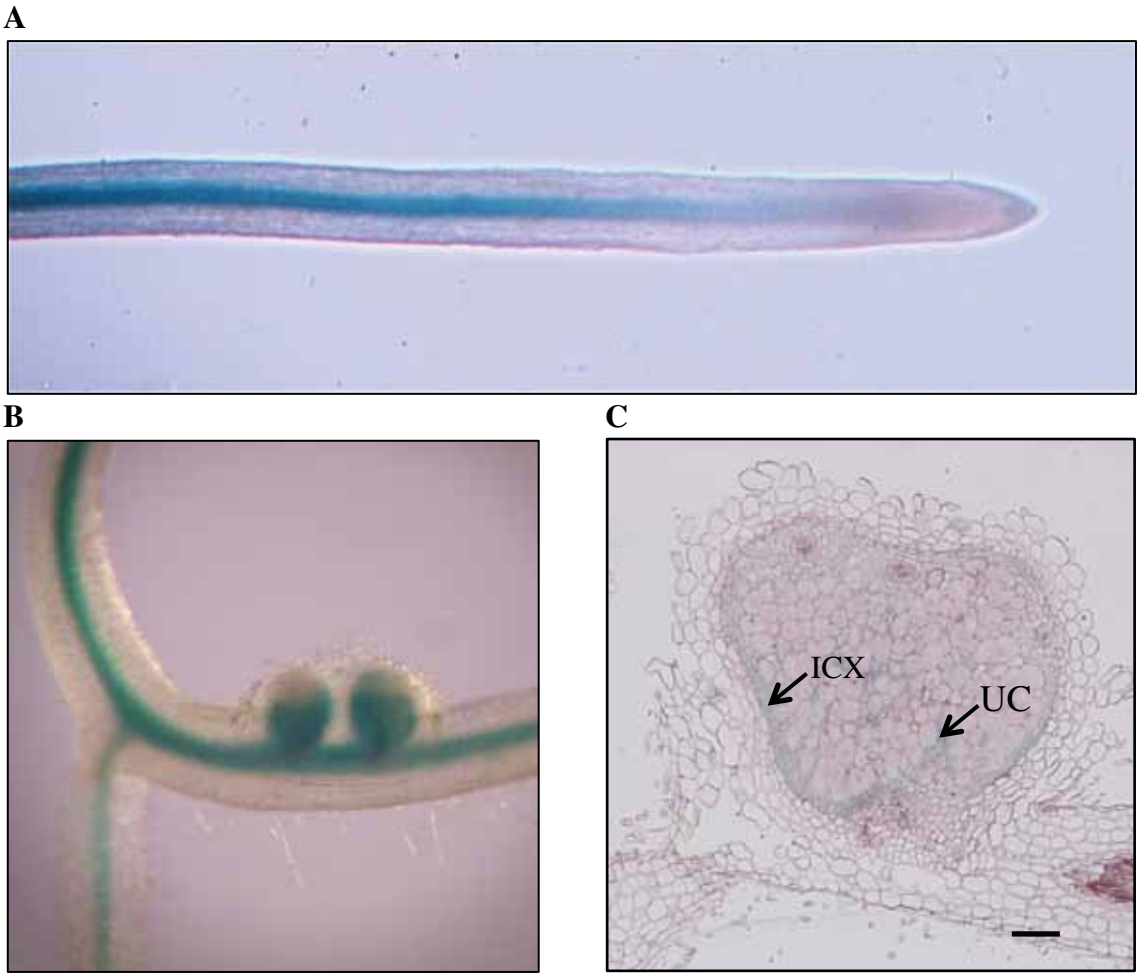


Figure 4.17. Expression of *MtMFS3* in *Medicago* roots and nodules.
GUS activity was detected from roots expressing p243-RedRoot-MtMFS-GUS. Positive roots stained for activity 7 days after transfer to perlite (**A**) and 14 days after addition of *S. meliloti* 2011 (**B**). **C**, Cross-section of a 14 day-old nodule showing expression in the inner cortex (ICX) and uninfected cells (UC). Scale bar = 200 μM.

Primer Name	Sequence
GmMFS1.1qPCRF	TGCATACTGTGGATAGATTTGATG
GmMFS1.1qPCRR	GATTGAGCAGGAAATGCTGATG
GmMFS1.2qPCRF	TTTTGCCCAGTGTGGATAGA
GmMFS1.2qPCRR	GGATTGACCAGATGCTGATG
GmMFS1.3qPCRF	ATGTGCCTTCCTCACCTCTC
GmMFS1.3qPCRR	GACTTACCTATAAACCATATTGCAG
GmMFS1.3 qPCR2 F	GCTAAAATTTGTTTCACACAATGATG
GmMFS1.3 qPCR2 R	CAGAGTAGAGGGCACGTACA
GmMFS1.4qPCRF	AACAACTGGAGAACGACAGCA
GmMFS1.4qPCRR	AACGAGGATTCCTTCTTTGTAGT
GmMFS1.5 qPCR F	ATGTGGGGCTGATCACAAGT
GmMFS1.5 qPCR R	AGCGGTGGAAGAATCAGCAG
GmMFS.1.3.cds.F	ATGGCACAACAACAAGAACATGA
GmMFS.1.3.cds.R	TCATAATGACTTACCTATAAACCAT
GmMFS.1.3.cds.R2	TAGCAAATCAATAGGAAGTTGTCA
GmMFS.1.3.cds.R3	TCAGCATTCTTGGGATTGACC
GmMFS1.3 Promoter F	ACTCTTTATATTACTTGATTTCTCTCAA
GmMFS1.3 Promoter R	TTGTTGTTGTGCCATACCAATATAAT
GmMFS1.3 RNAi F	GTGCATGCTGTTTTGGTGCTTT
GmMFS1.5 cds F	ATGGCATCGGAGTCAGGTCAAA
GmMFS1.5 cds R	TTAAACTTGACTGGATGTAGTGTCAC
MtMFS3 CDS F	ATGGCACAAAAGTCTGAAGATGAACCG
MtMFS3 CDS R	TTAACATTCTTGGGATTTAGGTTCTGC
MtMFS3 RNAi F	GGGAGTTCAAGACGGGGAGGCT
MtMFS3 RNAi R	ACGTTGACGACGAAAACCTTATCCG
MtMFS3 Promoter F	CAACTAAATCTTACAAATGTCTAATCAAAT
MtMFS3 Promoter R	TCAGACTTTTGTGCCATTCC

Table 4.1. List of primers used in this chapter.

4.3 Materials and Methods

4.3.1 Expression of Soybean *MFS* genes

Soybeans were planted in Waikerie sand (ten seeds per cm pot) and inoculated with *Bradyrhizobium japonicum* USDA110 on the day of planting and again the following day (100ml of a 1/10 dilution of a late log phase culture). Plants were grown in a glasshouse and watered daily, substituting nitrogen-free Herridge's nutrient solution (Herridge, 1982) three times per week. Tissue was collected from all organs (root, nodule, de-nodulated root, stem, leaf, and flowers) and frozen in liquid nitrogen. RNA was extracted using a RNeasy kit (Qiagen) and cDNA was synthesized from 1.25 μg of RNA using Superscript III (Invitrogen). For RT-PCR and qPCR, 2 μl of a 1/10 dilution of the cDNA was used per reaction. Primers are listed in Table 4.1. *GmMFS1.3* qPCR was carried out as described in section 2.3.5.

4.3.2 Cloning of *GmMFS1.3* and *GmMFS1.5*

Initially, a primer set was designed to clone *GmMFS1.3* based on the Phytozome genome prediction, using the primers GmMFS1.3 CDS F and R, however a product could not be amplified. RT-PCR using the CDS F primer in conjunction with the GmMFS1.3 qPCR R primer did yield a product (not the case for CDS R and qPCR F), therefore a GmMFS1.3 CDS R2 primer was designed but it did not amplify a product. Based on sequence alignments, it was found that there could be a missing exon in the genome prediction. Therefore, the primer GmMFS1.3 CDS R3 was synthesized and used to successfully amplify a full-length product from nodule cDNA using Platinum Taq High Fidelity (Invitrogen). The full-length product was then inserted into pCR8 TOPO, but upon transformation into the *E. coli* strain TOP10 there were few colonies (grown at 37°C). Sequencing of the isolated plasmids found mutations in the *GmMFS1.3* coding sequence. Therefore, the pCR8 TOPO reaction was transformed into the strain XL1-Blue and grown at room temperature on LB media containing a reduced concentration of NaCl (5 g L^{-1}). After 4 days growth, numerous colonies appeared and sequencing found that these clones

were mutation-free. *GmMFS1.5* was cloned using the primers GmMFS1.5 cds F and GmMFS1.5 cds R in a similar manner to GmMFS1.3, using XL1-Blue and growing the cells at room temperature.

4.3.3 Localization in Onion

GmMFS1.3-pCR8 TOPO was recombined into the plasmid pYFP-attr (Subramanian et al., 2006) using LR Clonase II (Invitrogen). The plasma membrane marker pm-rk CD3-1007 (Nelson et al., 2007) was digested with SacI and EcoRI and inserted into pBluescript. Plasmids were prepared by a homemade maxi-prep protocol based on the method outlined in Li et al. (2010). For the maxi-prep, 100ml of *E. coli* culture containing the plasmid was grown at 37°C in LB media, pelleted and stored at -20°C. The pellet was resuspended in 2 ml of solution I (50 mM Tris-HCl pH 7.5, 10 mM EDTA, 8 mg ml^{-1} lysozyme, and 100 μg ml^{-1} RNase) incubated at room temperature for 10 min, then mixed with 2 ml solution II (0.2 M NaOH, 1% (w/v) SDS) and incubated on ice for 10 min. 2 ml of solution III (1.32 M KOAc, pH 4.8) was then added and the solution was centrifuged at 20,000 x g for 10 min. The supernatant was then poured through a layer of miracloth, to remove any floating material, into a centrifuge tube containing 4 ml of 100% (v/v) isopropanol and spun at 16, 000 x g for 10 min. The pellet was then washed with 2.5 ml of 70% (v/v) ethanol and respun at 16, 000 x g for 10 min. The pellet was resuspended in 250 μl of solution I containing 10 μg of RNase and incubated at 37 °C for 20 min. 750 μl of 6M NaI was then added, mixed, and followed by the addition of 100 μl of silicon dioxide solution (250 mg/ml, Sigma). The solution was mixed for 5 min and centrifuged at 16, 000 x g for 15 sec. The pellet was then washed twice with solution E (50% (v/v) ethanol, 10 mM Tris-HCl, pH 7.5, 100 mM NaCl, 1 mM EDTA), and dried for 5 min at 50°C. The pellet was then resuspended in 100 μl of sterile water. Yields were generally around 1000 ng/μl. A second elution yielded a similar amount of DNA.

5 μg of plasmid DNA (in 10 μl) was combined with a 50 μl suspension (1.5 mg of 0.6 μm gold macrocarriers (Bio-Rad) in 50% glycerol). Onion epidermal peels were bombarded with a PDS-1000/He particle delivery system (Bio-Rad) using a 1100 psi Rupture Disc. Epidermal peels were maintained on Murashige and Skoog basal medium (with vitamins,

Austratec) supplemented with 9 L^{-1} TC grade agar (Austratec), 120 g L^{-1} sucrose and 500 mg L^{-1} tryptone. After bombardment, the peels were incubated in the dark for 24 hours at room temperature before viewing. Images were obtained using a Zeiss LSM 5 Pascal confocal microscope. YFP fluorescence was monitored by excitation at 515 nm with an argon laser combined with a 535-580 nm bandpass filter. mCherry fluorescence was obtained by excitation at 543 nm with a helium/neon laser combined with 615 nm longpass filter.

4.3.4 Expression of GmMFS1.3 in yeast and ^{14}C-Methylammonium Flux

GmMFS1.3-pCR8 TOPO was recombined into the yeast expression vector pYES3-DEST. The resulting plasmid *GmMFS1.3*-pYES3 was verified by sequencing with T7 and CYC1 primers. *GmMFS1.3*-pYES3 was transformed into the yeast strain 26972c by the lithium acetate/PEG and selected on YNB/Proline/Glucose plates. Positive colonies were re-streaked onto YNB/Proline/Glucose plates and single colonies used for subsequent experiments.

Starter yeast cultures were grown overnight in 20 ml of liquid YNB media (0.67% (w/v) Yeast Nitrogen Base without amino acids (Difco) and 2% (w/v) glucose) in sterile 100 ml glass conical flasks with cotton stoppers wrapped in aluminium foil. The cultures were incubated overnight at 28°C with shaking at 200 rpm, harvested at 4, 000 rpm for 2 minutes and washed twice in 50 ml sterile Milli Q water. The cells were resuspended to an OD_{600nm} of 0.2 in 20 ml of a modified minimal liquid media (Grenson, 1966) at pH 6.5 supplemented with 1mM NH_4Cl_2 and 2% (w/v) D-galactose, and incubated overnight at 28°C with shaking at 200 rpm. The cells were harvested at an approximate OD_{600nm} 1.0, and washed twice with 50 ml of sterile MQ water before being resuspended in room temperature KPO_4 buffer (20mM K_2HPO_4/KH_2PO_4 buffer (pH 6.2) supplemented with 2% (w/v) D-galactose) to a uniform OD_{600nm} 4.0.

The flux experiment consisted of 6 replicates for each transformed cell type staggered by 20 seconds: 100 μl of resuspended cells was added to 100 μl KPO_4 reaction buffer with ^{14}C-methylammonium (Perkin-Elmer) in 1.5 ml microcentrifuge tubes. After 10 minutes, 100 μl of the cells was removed and collected by vacuum filtration on to a 0.45 μM

nitrocellulose filter (Millipore, USA) and washed twice with 5ml of ice-cold KPO_4 buffer to prevent further ^{14}C-methylammonium uptake. The cells collected on filters were washed twice with 5 ml of ice-cold KPO_4 buffer. The filters were carefully placed into scintillation vials (Sarstedt) with 4 ml of scintillation fluid (StarScint-Perkin-Elmer). The radioactivity of the samples was determined with a liquid scintillation counter (Tri-Carb 2100, Beckmann or Packard). Counts were converted to equivalent amount of methylammonium and samples were normalised against total protein according to a modified Lowry method (Peterson, 1977).

4.3.5 Stopped Flow Spectrophotometry

Starter yeast cultures were grown overnight in 20 ml of liquid YNB media (0.67% (w/v) Yeast Nitrogen Base without amino acids and 2% (w/v) glucose) in sterile 100 ml glass conical flasks with cotton stoppers wrapped in aluminium foil. The cultures were incubated overnight at 28°C with shaking at 200 rpm. The cell pellets were washed twice with 50 ml of sterile Milli Q water and then resuspended in 20 ml Grenson's minimal yeast media supplemented with 0.5mM $(NH_4)_2SO_4$ and 2% (w/v) galactose. The cultures were incubated for a further 16 hours at 28°C with shaking at 200 rpm. The cultures were harvested at an OD_{600nm} 0.8-1.0, and the cell pellets were washed with 50 ml of 5mM KH_2PO_4 (pH 7.5), resuspending the cells to an OD_{600nm} of 1.0 in 10ml of 5mM KH_2PO_4 (pH 7.5) solution supplemented with 20 μl of B-mercaptoethanol (98% (v/v)). The cells were incubated at 28°C with shaking at 200 rpm for 30 minutes. The cells were then centrifuged at 4,000rpm for 2 minutes, the supernatant decanted, and resuspended in a buffer containing 2.4M sorbitol, 5mM KH_2PO_4 and 600 units of lyticase (Sigma-Aldrich, USA). The cells were incubated for 45 minutes at 28°C, and then centrifuged at 4,000 rpm for 5 minutes. The supernatant was carefully aspirated and the spheroplasts were washed once and resuspended in 1 ml of resuspension buffer (10mM Trisodium citrate, 1mM EDTA and 0.5M sorbitol and 0.4M K_2SO_4, pH adjusted to 6.0 with MES) and then diluted to a uniform OD_{475} 1.0. The spheroplast suspensions were mixed in a fast kinetics chamber (SFM-300, Biologic) with an equal volume of resuspension solution. Volume changes were recorded at 16°C as light scattering at an angle 90° and 475 nm. The kinetics presented are the smoothed normalized averages of 6-9 trace recordings each over a period of 6 seconds.

4.3.6 Cloning of *GmMFS1.3* and *MtMFS3* promoters

The promoter of *GmMFS1.3* (2046 bp) was amplified using the primers GmMFS1.3 Promoter F and R. Similarly, the promoter of *MtMFS3* (2176 bp) was cloned using the primers MtMFS3 Promoter F and R. The *GmMFS1.3* promoter was inserted into pCR8-TOPO, while the *MtMFS3* promoter was inserted into pENTR-D-TOPO. Both sequences were then recombined into p243-RedRoot-GUS using LR Clonase (Invitrogen). Soybean hairy roots were generated by *Agrobacterium rhizogenes* K599, as outlined in 2.3.6. *Medicago truncatula* hairy roots were generated using *Agrobacterium rhizogenes* MSU440, as outlined in 3.3.3. Positive roots were then identified using a dsRED filter set equipped on a Leica stereoscope. GUS staining was carried out exactly as described in 2.3.6.

4.4 Discussion

4.4.1 Synteny of SATs and MFSs

Having discovered that GmSAT1 induced the expression of *YOR387W*, BLAST searches were undertaken against the soybean genome with YOR378W. Surprisingly, the MFS family that was uncovered contained members that are tightly linked with the *SAT* locus in many sequenced dicots. This result prompted the question: Does GmSAT1 regulate the expression of the newly discovered MFS transporter genes? The evidence presented here does not support direct transcriptional regulation. *GmSAT1* is only expressed in root and nodules, however the MFS family members are expressed in all plant tissues. Although there is a discrepancy in expression, it does not eliminate the possibility that GmSAT1 and the MFS transporters participate in a similar pathway. Based on GUS staining, both *GmSAT1* and *GmMFS1.3* are expressed in similar cell types in nodules (inner cortex). Therefore, the MFSs may be involved in the general transport of ammonium (or another compound) out of source tissues and/or into sink tissues (such as seeds and leaves), whereas GmSAT1 influences the coordination of nitrogen metabolism in overlapping

tissues. Thus, the link between GmSAT1 and GmMFS1.3 could be indirect. Nevertheless, the tight linkage between these two gene families is rather intriguing.

4.4.2 The function of *GmMFS1.3*

Expression analysis indicated that *GmMFS1.1-1.4* are present in all tissues at a relatively low level with a significant enrichment of *GmMFS1.3* in nodules and *GmMFS1.5* in nodules, roots, and flowers. Based on tissue localization (and plasma membrane targeting), it is possible that GMFS1.3 (and MtMFS3) could play a role in exporting fixed nitrogen from the infect region of the nodule. Expression in the inner cortex and in vascular bundles of the nodule is very similar to the recently characterized ureide transporters in soybean (Collier and Tegeder, 2012). The localization also overlaps somewhat with *LjAMT1;1* and *LjAMT2;1*, two characterized *Lotus japonicus* ammonium transporters, however *LjAMT1;1* and *LjAMT2;1* are also highly expressed in other nodule tissues (Simon-Rosin et al., 2003; Rogato et al., 2008). Therefore, GmMFS1.3 could be responsible for scavenging ammonium in the inner cortex and delivering it to the vasculature. A second possibility, based on the fact that GmMFS1.3 has a low affinity for ammonium, is that GmMFSs transport other molecules. There is some overlap in localization with known ureide transporters (Collier and Tegeder, 2012), suggesting GmMFS1;3 may be involved in allatoin or allantoic acid transport. Obtaining a yeast ureide transport mutant strain would enable testing of this hypothesis. A third possibility is that GmMFS1.3 transports a carbon compound or nutrient into the nodule from the vasculature, using ammonium (or a similar compound, such as potassium or sodium) in a symport or antiport mechanism. The most closely related protein in insects to GmMFS1.3 is SPINSTER, a transporter that has been associated with carbohydrate efflux in lysosomes (Dermaut et al., 2005). Further, the most homologous protein to YOR378W in yeast is the boron exporter, ATR1 (Kaya et al., 2009). Conducting transport studies of GmMFS1.3 in *Xenopus* oocytes would be desirable to assess these possibilities.

5 GmCamK1 Manuscript

CHARACTERIZATION OF GmCAMK1, A MEMBER OF A SOYBEAN CALMODULIN-BINDING RECEPTOR-LIKE KINASE FAMILY

Thomas A. DeFalco[a], David Chiasson[b], Kim Munro[c], Brent N. Kaiser[b], Wayne A. Snedden[a]

[a] Department of Biology, Queen's University, Kingston, Ontario, Canada, K7L 3N6
[b] School of Agriculture, Food and Wine, The University of Adelaide, Waite Campus Urrbrae, SA 5064, Australia
[c] Protein Function Discovery Facility, Queen's University, Kingston, Ontario, Canada

FEBS Letters. 2010. 584: 4717-4724

STATEMENT OF AUTHORSHIP

CHARACTERIZATION OF GmCAMK1, A MEMBER OF A SOYBEAN CALMODULIN-BINDING RECEPTOR-LIKE KINASE FAMILY

FEBS Letters. 2010. 584: 4717-4724.

CHIASSON, D.M. (Candidate)

Performed bioinformatics and gene expression analysis, contributed to experimental design and manuscript evaluation.

I hereby certify that the statement of contribution is accurate.

Signed .....*Date*.....

DEFALCO, T.A.

Performed protein screen and biochemical analysis, interpreted data, and prepared manuscript.

I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the paper in the thesis.

MUNRO, K.

Performed ITC experimentation and data analysis.

I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the paper in the thesis.

Signed .....*Date* *August 16, 2012*.....

KAISER, B.N.

Interpreted data and evaluated manuscript.

I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the paper in the thesis.

Signed .....*Date*.....

SNEDDEN, W.S.

Supervised development of work, produced protein expression library, helped in data interpretation and manuscript preparation, acted as corresponding author.

I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the paper in the thesis.

Signed .....*Date* *August 16, 2012*.....

Thomas A. DeFalco, David Chiasson, Kim Munro, Brent N. Kaiser, Wayne A. Snedden (2010) Characterization of GmCaMK1, a member of a soybean calmodulin-binding receptor-like kinase family.
FEBS Letters, v. 584 (23), pp. 4717-4724, December 2010

6 Conclusion and Future Directions

From this study, a number of exciting discoveries were made regarding the membrane-bound transcription factor GmSAT1. First, GmSAT1 was shown to bind DNA, localize to the nucleus, and influence transcription, affirming that it is a *bona fide* transcription factor. The results also corroborate previous work that demonstrated membrane-association of GmSAT1. Localization studies presented here clearly show that GmSAT1 is associated with vesicles *in vivo*, and that processing at both termini is likely required for relocation to the nucleus. From the promoter analysis conducted, it now appears that *GmSAT1* is functioning in additional locations within the nodule, including uninfected cells and the inner cortex. The localization and spatial expression results were also supported by an analysis of *MtSAT1* and *MtSAT2*, demonstrating commonality in two legumes. Further, evidence is presented that multiple transcripts originate from the *GmSAT1* locus, similar to the newly identified gene *GmSAT2*. *GmSAT2* is expressed in similar nodule tissues as *GmSAT1* and shares a common function with GmSAT1, being able to induce growth of the yeast strain 26972c on low ammonium. Taken together, these studies have set the stage for understanding the *in vivo* role of GmSAT1.

From the microarray analysis conducted in soybean nodules, it would appear that GmSAT1 is somehow influencing the circadian clock. As the circadian clock controls many processes, including metabolism, this perturbation may explain the *GmSAT1* RNAi phenotype. To date, there have not been any reports regarding the circadian clock in nitrogen-fixing nodules. From the data presented here, and by analyzing available gene expression data, it would appear that components of the clock are expressed in nodules. Because the fixation of nitrogen is dependent on photosynthesis for carbon skeletons, the supply of photosynthate directly influences fixation (Millhollon and Williams, 1986). Since a previous study indicated that the root circadian clock timing is coupled to the supply of sugars by photosynthesis (James et al., 2008), likely the clock is required in nodules to balance N-fixation with carbon supply. Determining how the circadian clock influences nitrogen fixation would be a very interesting and informative endeavor. By analyzing spatial expression of clock components (such as *GI, PRR5, PRR7*) and the effect of silencing these genes in nodules, the requirement of this system could be determined.

The downregulated genes will also be useful to identify direct targets of GmSAT1 *in vivo*. ChIP-seq experimental workflow was initiated by first purifying infected cell nuclei, however the expression of *GmSAT1* was not detected by western blotting. Since GmSAT1 appears to be more strongly expressed in uninfected cells and the inner cortex, a modified nuclear isolation procedure should be employed. The procedure of Kouchi et al. (1988) could be used to first fractionate the nodule into cortical, uninfected, and infected cells before nuclear isolation. This experiment would also be useful to identify the location of the GmSAT1 cleavage events. This study has demonstrated that N- and C-terminal tags are cleaved from GmSAT1 *in vivo*, and that internal tagging renders the protein non-functional (not shown). Therefore, new polyclonal antibodies generated to recognize non-cleaved GmSAT1 regions would be useful for the ChIP-seq and future western blots.

This thesis has also uncovered an interesting relationship between GmSAT1 and a novel subfamily of major facilitator transporters (MFS). These MFS transporters were discovered by searching for sequences related to YOR378W, a MFS transporter upregulated by GmSAT1 expression in yeast. Remarkably, members of this MFS family are found in proximity to *GmSAT1* and *GmSAT1*-like loci in most dicots. The implications of this association remained to be determined, however the fact that GmFMS1.3 transports ammonium and is expressed in similar cell types as GmSAT1 is intriguing. Transport studies in yeast indicate that GmMFS1.3 moves ammonium, but work remains to further characterize this transporter. As MFS transporters are known to couple transport (as antiporters and symporters), an additional molecule may be a substrate of GmMFS1.3, such as an anion or a metabolite. Studies have been initiated in *Xenopus laevis* to determine if such a coupling exists. Determining the mechanism of transport and preferred substrates of the MFS transporters will help in determining their relationship to GmSAT1.

Finally, a novel receptor-like kinase protein was also characterized from soybean nodules. GmCaMK1 was identified in a protein interaction screen using conserved calmodulin (CaM) as bait. The interaction was mapped to a novel 24 amino acid region near the C-terminus of GmCaMK1. The identification of this peptide will hopefully enable the discovery of CaM-binding domains amongst other proteins. The activity of CaM has previously been associated with nodulation, as CaM targets and CaM-like proteins have been found in nodules (Lévy et al., 2004; Liu et al., 2006). Since CCaMK, a CaM-binding

kinase, is critical for nodulation it would be desirable to uncover the role of GmCaMK1. This could be achieved by screening for phosphorylation targets and targeting *GmCaMK1* for reverse genetic analysis.

7 Bibliography

Amor BB, Shaw SL, Oldroyd GE, Maillet F, Penmetsa RV, Cook D, Long SR, Denarie J, Gough C (2003) The NFP locus of Medicago truncatula controls an early step of Nod factor signal transduction upstream of a rapid calcium flux and root hair deformation. Plant J **34:** 495-506

Andrade SL, Dickmanns A, Ficner R, Einsle O (2005) Crystal structure of the archaeal ammonium transporter Amt-1 from Archaeoglobus fulgidus. Proc Natl Acad Sci U S A **102:** 14994-14999

Ane J-M, Kiss GB, Riely BK, Penmetsa RV, Oldroyd GED, Ayax C, Levy J, Debelle F, Baek J-M, Kalo P, Rosenberg C, Roe BA, Long SR, Denarie J, Cook DR (2004) Medicago truncatula DMI1 required for bacterial and fungal symbioses in legumes. Science **303:** 1364-1367

Appleby CA (1984) Leghemoglobin and Rhizobium respiration. Annual Review of Plant Physiology **35:** 443-478

Asamizu E, Shimoda Y, Kouchi H, Tabata S, Sato S (2008) A Positive Regulatory Role for LjERF1 in the Nodulation Process Is Revealed by Systematic Analysis of Nodule-Associated Transcription Factors of Lotus japonicus. Plant Physiol **147:** 2030-2040

Atkins CA, Smith PMC (2007) Translocation in legumes: assimilates, nutrients, and signaling molecules. Plant Physiol **144:** 550-561

Auriac MC, Timmers ACJ (2007) Nodulation studies in the model legume Medicago truncatula: Advantages of using the constitutive EF1α promoter and limitations in detecting fluorescent reporter proteins in nodule tissues. Molecular plant-microbe interactions **20:** 1040-1047

Bailey PC, Martin C, Toledo-Ortiz G, Quail PH, Huq E, Heim MA, Jakoby M, Werber M, Weisshaar B (2003) Update on the basic helix-loop-helix transcription factor gene family in Arabidopsis thaliana. Plant Cell **15:** 2497-2502

Banba M, Gutjahr C, Miyao A, Hirochika H, Paszkowski U, Kouchi H, Imaizumi-Anraku H (2008) Divergence of evolutionary ways among common sym genes: CASTOR and CCaMK show functional conservation between two symbiosis systems and constitute the root of a common signaling pathway. Plant and cell physiology **49:** 1659-1671

Barker DG, Bianchi S, Blondon F, Dattée Y, Duc G, Essad S, Flament P, Gallusci P, Génier G, Guy P (1990) Medicago truncatula, a model plant for studying the molecular genetics of theRhizobium-legume symbiosis. Plant Molecular Biology Reporter **8:** 40-49

Barlow D, Thornton J (1988) Helix geometry in proteins. J Mol Biol **201:** 601-619

Benedito VA, Torres-Jerez I, Murray JD, 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 PX, Tang Y, Udvardi MK (2008) A gene expression atlas of the model legume Medicago truncatula. Plant J **55:** 504-513

Bergmann H, Preddie E, Verma DP (1983) Nodulin-35: a subunit of specific uricase (uricase II) induced and localized in the uninfected cells of soybean nodules. EMBO J **2:** 2333-2339

Blumwald E, Fortin MG, Rea PA, Verma DP, Poole RJ (1985) Presence of Host-Plasma Membrane Type H-ATPase in the Membrane Envelope Enclosing the Bacteroids in Soybean Root Nodules. Plant Physiol **78:** 665-672

Boisson-Dernier A, Chabaud M, Garcia F, Bécard G, Rosenberg C, Barker DG (2001) Agrobacterium rhizogenes-transformed roots of Medicago truncatula for the study of nitrogen-fixing and endomycorrhizal symbiotic associations. Molecular plant-microbe interactions **14:** 695-700

Brown MS, Goldstein JL (1999) A proteolytic pathway that controls the cholesterol content of membranes, cells, and blood. Proc Natl Acad Sci U S A **96:** 11041-11048

Brown S, Oparka K, Sprent J, Walsh K (1995) Symplastic transport in soybean root nodules. Soil Biology and Biochemistry **27:** 387-399

Capoen W, Sun J, Wysham D, Otegui MS, Venkateshwaran M, Hirsch S, Miwa H, Downie JA, Morris RJ, Ane JM, Oldroyd GE (2011) Nuclear membranes control symbiotic calcium signaling of legumes. Proc Natl Acad Sci U S A **108:** 14348-14353

Catalano CM, Lane WS, Sherrier DJ (2004) Biochemical characterization of symbiosome membrane proteins from Medicago truncatula root nodules. Electrophoresis **25:** 519-531

Chabaud M, Larsonneau C, Marmouget C, Huguet T (1996) Transformation of barrel medic (Medicago truncatula Gaertn.) by Agrobacterium tumefaciens and regeneration via somatic embryogenesis of transgenic plants with the MtENOD12 nodulin promoter fused to the gus reporter gene. Plant Cell Reports **15:** 305-310

Charpentier M, Bredemeier R, Wanner G, Takeda N, Schleiff E, Parniske M (2008) Lotus japonicus CASTOR and POLLUX are ion channels essential for perinuclear calcium spiking in legume root endosymbiosis. Plant Cell **20:** 3467-3479

Chen NZ, Zhang XQ, Wei PC, Chen QJ, Ren F, Chen J, Wang XC (2007) AtHAP3b plays a crucial role in the regulation of flowering time in Arabidopsis during osmotic stress. J Biochem Mol Biol **40:** 1083-1089

Cheon CI, Lee NG, Siddique AB, Bal AK, Verma DP (1993) Roles of plant homologs of Rab1p and Rab7p in the biogenesis of the peribacteroid membrane, a subcellular compartment formed de novo during root nodule symbiosis. EMBO J **12:** 4125-4135

Collier R, Tegeder M (2012) Soybean ureide transporters play a critical role in nodule development, function and nitrogen export. Plant J **doi: 10.1111/j.1365-313X.2012.05086.x**

Combier J-P, Frugier F, de Billy F, Boualem A, El-Yahyaoui F, Moreau S, Vernié T, Ott T, Gamas P, Crespi M, Niebel A (2006) MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in Medicago truncatula. Genes & Development **20:** 3084-3088

Covington MF, Maloof JN, Straume M, Kay SA, Harmer SL (2008) Global transcriptome analysis reveals circadian regulation of key pathways in plant growth and development. Genome Biol **9:** R130

Dalchau N, Baek SJ, Briggs HM, Robertson FC, Dodd AN, Gardner MJ, Stancombe MA, Haydon MJ, Stan GB, Gonçalves JM (2011) The circadian oscillator gene GIGANTEA mediates a long-term response of the Arabidopsis thaliana circadian clock to sucrose. Proceedings of the National Academy of Sciences **108:** 5104

Datta DB, Triplett EW, Newcomb EH (1991) Localization of xanthine dehydrogenase in cowpea root nodules: implications for the interaction between cellular compartments during ureide biogenesis. Proceedings of the National Academy of Sciences **88:** 4700-4702

Day DA, Price GD, Udvardi MK (1989) Membrane Interface of the Bradyrhizobium japonicum–Glycine max Symbiosis: Peribacteroid Units From Soyabean Nodules. Functional Plant Biology **16:** 69-84

Dean RM, Rivers RL, Zeidel ML, Roberts DM (1999) Purification and functional reconstitution of soybean nodulin 26. An aquaporin with water and glycerol transport properties. Biochemistry **38:** 347-353

Dermaut B, Norga KK, Kania A, Verstreken P, Pan H, Zhou Y, Callaerts P, Bellen HJ (2005) Aberrant lysosomal carbohydrate storage accompanies endocytic defects and neurodegeneration in Drosophila benchwarmer. The Journal of cell biology **170:** 127-139

Dubois E, Grenson M (1979) Methylamine/ammonia uptake systems in Saccharomyces cerevisiae: multiplicity and regulation. Molecular and General Genetics MGG **175:** 67-76

Ferguson BJ, Indrasumunar A, Hayashi S, Lin MH, Lin YH, Reid DE, Gresshoff PM (2010) Molecular analysis of legume nodule development and autoregulation. J Integr Plant Biol **52:** 61-76

Forde B, Lorenzo H (2001) The nutritional control of root development. Plant and Soil **232:** 51-68

Fortin MG, Morrison NA, Verma DPS (1987) Nodulin-26, a peribacteroid membrane nodulin is expressed independently of the development of the peribacteroid compartment. Nucleic Acids Res **15:** 813-824

Fukushima A, Kusano M, Nakamichi N, Kobayashi M, Hayashi N, Sakakibara H, Mizuno T, Saito K (2009) Impact of clock-associated Arabidopsis pseudo-response regulators in metabolic coordination. Proceedings of the National Academy of Sciences **106:** 7251

Gendron JM, Pruneda-Paz JL, Doherty CJ, Gross AM, Kang SE, Kay SA (2012) Arabidopsis circadian clock protein, TOC1, is a DNA-binding transcription factor. Proceedings of the National Academy of Sciences **109:** 3167-3172

Gepts P, Beavis WD, Brummer EC, Shoemaker RC, Stalker HT, Weeden NF, Young ND (2005) Legumes as a model plant family. Genomics for food and feed report of the cross-legume advances through genomics conference. Plant Physiol **137:** 1228-1235

Gietz RD, Schiestl RH (2007) High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat Protoc **2:** 31-34

Glass ADM, Brito DT, Kaiser BN, Kronzucker HJ, Kumar A, Okamoto M, Rawat S, Siddiqi MY, Silim SM, Vidmar JJ (2001) Nitrogen transport in plants, with an emphasis on the regulations of fluxes to match plant demand. Journal of plant nutrition and soil science **164:** 199-207

Govindarajulu M, Elmore JM, Fester T, Taylor CG (2008) Evaluation of constitutive viral promoters in transgenic soybean roots and nodules. Molecular plant-microbe interactions **21:** 1027-1035

Grønlund M, Gustafsen C, Roussis A, Jensen D, Nielsen LP, Marcker KA, Jensen EØ (2003) The *Lotus japonicus ndx* gene family is involved in nodule function and maintenance. Plant Mol Biol **52:** 303-316

Groth M, Takeda N, Perry J, Uchida H, Draxl S, Brachmann A, Sato S, Tabata S, Kawaguchi M, Wang TL, Parniske M (2010) NENA, a Lotus japonicus homolog of Sec13, is required for rhizodermal infection by arbuscular mycorrhiza fungi and rhizobia but dispensable for cortical endosymbiotic development. Plant Cell **22:** 2509-2526

Guenther JF, Chanmanivone N, Galetovic MP, Wallace IS, Cobb JA, Roberts DM (2003) Phosphorylation of soybean nodulin 26 on serine 262 enhances water permeability and is regulated developmentally and by osmotic signals. The Plant Cell Online **15:** 981-991

Haan C, Behrmann I (2007) A cost effective non-commercial ECL-solution for Western blot detections yielding strong signals and low background. J Immunol Methods **318:** 11-19

Hakoyama T, Niimi K, Watanabe H, Tabata R, Matsubara J, Sato S, Nakamura Y, Tabata S, Jichun L, Matsumoto T (2009) Host plant genome overcomes the lack of a bacterial gene for symbiotic nitrogen fixation. Nature **462:** 514-517

Hanks JF, Tolbert N, Schubert KR (1981) Localization of enzymes of ureide biosynthesis in peroxisomes and microsomes of nodules. Plant Physiol **68:** 65

Hayashi T, Banba M, Shimoda Y, Kouchi H, Hayashi M, Imaizumi-Anraku H (2010) A dominant function of CCaMK in intracellular accommodation of bacterial and fungal endosymbionts. Plant J **63:** 141-154

Herrero E, Davis SJ (2012) Time for a Nuclear Meeting: Protein Trafficking and Chromatin Dynamics Intersect in the Plant Circadian System. Mol Plant **5:** 554-565

Herridge DF (1982) Relative abundance of ureides and nitrate in plant tissues of soybean as a quantitative assay of nitrogen fixation. Plant Physiol **70:** 1

Huq E, Tepperman JM, Quail PH (2000) GIGANTEA is a nuclear protein involved in phytochrome signaling in Arabidopsis. Proceedings of the National Academy of Sciences **97:** 9789

Imaizumi-Anraku H, Takeda N, Charpentier M, Perry J, Miwa H, Umehara Y, Kouchi H, Murakami Y, Mulder L, Vickers K, Pike J, Downie JA, Wang T, Sato S, Asamizu E, Tabata S, Yoshikawa M, Murooka Y, Wu G-J, Kawaguchi M, Kawasaki S, Parniske M, Hayashi M (2005) Plastid proteins crucial for symbiotic fungal and bacterial entry into plant roots. Nature **433:** 527-531

Ivanov S, Fedorova EE, Limpens E, De Mita S, Genre A, Bonfante P, Bisseling T (2012) Rhizobium-legume symbiosis shares an exocytotic pathway required for arbuscule formation. Proc Natl Acad Sci U S A **109:** 8316-8321

Iwata Y, Koizumi N (2005) An Arabidopsis transcription factor, AtbZIP60, regulates the endoplasmic reticulum stress response in a manner unique to plants. Proc Natl Acad Sci U S A **102:** 5280-5285

James AB, Monreal JA, Nimmo GA, Kelly CL, Herzyk P, Jenkins GI, Nimmo HG (2008) The circadian clock in Arabidopsis roots is a simplified slave version of the clock in shoots. Science's STKE **322:** 1832

Ji H, Vagner B, Mingyi W, Jeremy M, Patrick Z, Yuhong T, Michael U (2009) The Medicago truncatula gene expression atlas web server. BMC Bioinformatics **10**

Jones KM, Kobayashi H, Davies BW, Taga ME, Walker GC (2007) How rhizobial symbionts invade plants: the Sinorhizobium-Medicago model. Nat Rev Micro **5:** 619-633

Kaiser BN, Finnegan PM, Tyerman SD, Whitehead LF, Bergersen FJ, Day DA, Udvardi MK (1998) Characterization of an ammonium transport protein from the peribacteroid membrane of soybean nodules. Science **281:** 1202-1206

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

Kaiser BN, Rawat SR, Siddiqi MY, Masle J, Glass AD (2002) Functional analysis of an Arabidopsis T-DNA "knockout" of the high-affinity NH4(+) transporter AtAMT1;1. Plant Physiol **130:** 1263-1275

Kaló P, Gleason C, Edwards A, Marsh J, Mitra RM, Hirsch S, Jakab J, Sims S, Long SR, Rogers J, Kiss GB, Downie JA, Oldroyd GED (2005) Nodulation Signaling in Legumes Requires NSP2, a Member of the GRAS Family of Transcriptional Regulators. Science **308:** 1786-1789

Kanamori N, Madsen LH, Radutoiu S, Frantescu M, Quistgaard EMH, Miwa H, Downie JA, James EK, Felle HH, Haaning LL, Jensen TH, Sato S, Nakamura Y, Tabata S, Sandal N, Stougaard J (2006) A nucleoporin is required for induction of Ca2+ spiking in legume nodule development and essential for rhizobial and fungal symbiosis. Proc Natl Acad Sci U S A **103:** 359-364

Karimi M, Inz D, Depicker A (2002) GATEWAY (TM) vectors for Agrobacterium-mediated plant transformation. Trends Plant Sci **7:** 193-195

Kaya A, Karakaya HC, Fomenko DE, Gladyshev VN, Koc A (2009) Identification of a novel system for boron transport: Atr1 is a main boron exporter in yeast. Mol Cell Biol **29:** 3665-3674

Kereszt A, Li D, Indrasumunar A, Nguyen CDT, Nontachaiyapoom S, Kinkema M, Gresshoff PM (2007) Agrobacterium rhizogenes-mediated transformation of soybean to study root biology. Nat Protoc **2:** 948-952

Kim S-G, Kim S-Y, Park C-M (2007) A membrane-associated NAC transcription factor regulates salt-responsive flowering via FLOWERING LOCUS T in Arabidopsis. Planta **226:** 647-654

Kim S-Y, Kim S-G, Kim Y-S, Seo PJ, Bae M, Yoon H-K, Park C-M (2007) Exploring membrane-associated NAC transcription factors in Arabidopsis: implications for membrane biology in genome regulation. Nucleic Acids Res **35:** 203-213

Kim WY, Fujiwara S, Suh SS, Kim J, Kim Y, Han L, David K, Putterill J, Nam HG, Somers DE (2007) ZEITLUPE is a circadian photoreceptor stabilized by GIGANTEA in blue light. Nature **449:** 356-360

Kim WY, Geng R, Somers DE (2003) Circadian phase-specific degradation of the F-box protein ZTL is mediated by the proteasome. Proceedings of the National Academy of Sciences **100:** 4933

Kim Y-S, Kim S-G, Park J-E, Park H-Y, Lim M-H, Chua N-H, Park C-M (2006) A membrane-bound NAC transcription factor regulates cell division in Arabidopsis. Plant Cell **18:** 3132-3144

Kim Y-S, Park C-M (2007) Membrane regulation of cytokinin-mediated cell division in Arabidopsis. Plant Signal Behav **2:** 15-16

Konvicka K, Campagne F, Weinstein H (2000) Interactive construction of residue-based diagrams of proteins: the RbDe web service. Protein Eng **13:** 395-396

Kouchi H, Fukai K, Katagiri H, Minamisawa K, Tajima S (1988) Isolation and enzymological characterization of infected and uninfected cell protoplasts from root nodules of Glycine max. Physiol Plant **73:** 327-334

Krusell L, Krause K, Ott T, Desbrosses G, Krämer U, Sato S, Nakamura Y, Tabata S, James EK, Sandal N (2005) The sulfate transporter SST1 is crucial for symbiotic nitrogen fixation in Lotus japonicus root nodules. The Plant Cell Online **17:** 1625-1636

Kushnirov VV (2000) Rapid and reliable protein extraction from yeast. Yeast **16:** 857-860

Lavin M, Herendeen PS, Wojciechowski MF (2005) Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the Tertiary. Systematic Biology **54:** 575-594

Law CJ, Maloney PC, Wang DN (2008) Ins and Outs of Major Facilitator Superfamily Antiporters. Annu. Rev. Microbiol. **62:** 289-305

Lévy J, Bres C, Geurts R, Chalhoub B, Kulikova O, Duc G, Journet E-P, Ané J-M, Lauber E, Bisseling T, Dénarié J, Rosenberg C, Debellé F (2004) A Putative Ca2+ and Calmodulin-Dependent Protein Kinase Required for Bacterial and Fungal Symbioses. Science **303:** 1361-1364

Li JF, Li L, Sheen J (2010) Protocol: a rapid and economical procedure for purification of plasmid or plant DNA with diverse applications in plant biology. Plant Methods **6:** 1

Libault M, Thibivilliers S, Bilgin D, Radwan O, Benitez M, Clough S, Stacey G (2008) Identification of Four Soybean Reference Genes for Gene Expression Normalization. The Plant Genome **1:** 44-54

Limpens E, Franken C, Smit P, Willemse J, Bisseling T, Geurts R (2003) LysM domain receptor kinases regulating rhizobial Nod factor-induced infection. Science **302:** 630-633

Limpens E, Ivanov S, van Esse W, Voets G, Fedorova E, Bisseling T (2009) Medicago N2-fixing symbiosomes acquire the endocytic identity marker Rab7 but delay the acquisition of vacuolar identity. Plant Cell **21:** 2811-2828

Limpens E, Mirabella R, Fedorova E, Franken C, Franssen H, Bisseling T, Geurts R (2005) Formation of organelle-like N2-fixing symbiosomes in legume root nodules is controlled by DMI2. Proc Natl Acad Sci U S A **102:** 10375-10380

Limpens E, Ramos J, Franken C, Raz V, Compaan B, Franssen H, Bisseling T, Geurts R (2004) RNA interference in Agrobacterium rhizogenes - transformed roots of Arabidopsis and Medicago truncatula. J Exp Bot **55:** 983-992

Liu J, Miller SS, Graham M, Bucciarelli B, Catalano CM, Sherrier DJ, Samac DA, Ivashuta S, Fedorova M, Matsumoto P (2006) Recruitment of novel calcium-binding proteins for root nodule symbiosis in Medicago truncatula. Plant Physiol **141:** 167-177

Liu J-X, Srivastava R, Che P, Howell SH (2007) An endoplasmic reticulum stress response in Arabidopsis is mediated by proteolytic processing and nuclear

relocation of a membrane-associated transcription factor, bZIP28. Plant Cell **19:** 4111-4119

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2-[Delta][Delta] CT method. Methods **25:** 402-408

Locke JCW, Kozma-Bognár L, Gould PD, Fehér B, Kevei E, Nagy F, Turner MS, Hall A, Millar AJ (2006) Experimental validation of a predicted feedback loop in the multi-oscillator clock of Arabidopsis thaliana. Molecular Systems Biology **2**

Loqué D, Yuan L, Kojima S, Gojon A, Wirth J, Gazzarrini S, Ishiyama K, Takahashi H, Von Wirén N (2006) Additive contribution of AMT1; 1 and AMT1; 3 to high - affinity ammonium uptake across the plasma membrane of nitrogen - deficient Arabidopsis roots. The Plant Journal **48:** 522-534

Lorenz MC, Heitman J (1998) The MEP2 ammonium permease regulates pseudohyphal differentiation in Saccharomyces cerevisiae. EMBO J **17:** 1236-1247

Loughlin PC (2007) Elucidation of a peribacteroid membrane-bound bHLH transcription factor required for legume nitrogen fixation. Ph.D. Thesis. The University of Adelaide, Australia

Lu SX, Knowles SM, Webb CJ, Celaya RB, Cha C, Siu JP, Tobin EM (2011) The Jumonji C domain-containing protein JMJ30 regulates period length in the Arabidopsis circadian clock. Plant Physiol **155:** 906-915

Madsen EB, Madsen LH, Radutoiu S, Olbryt M, Rakwalska M, Szczyglowski K, Sato S, Kaneko T, Tabata S, Sandal N, Stougaard J (2003) A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals. Nature **425:** 637-640

Marini AM, Matassi G, Raynal V, Andre B, Cartron JP, Cherif-Zahar B (2000) The human Rhesus-associated RhAG protein and a kidney homologue promote ammonium transport in yeast. Nat Genet **26:** 341-344

Marini AM, Soussi-Boudekou S, Vissers S, Andre B (1997) A family of ammonium transporters in Saccharomyces cerevisiae. Mol Cell Biol **17:** 4282-4293

Marini AM, Springael JY, Frommer WB, Andre B (2000) Cross-talk between ammonium transporters in yeast and interference by the soybean SAT1 protein. Mol Microbiol **35:** 378-385

Marini AM, Vissers S, Urrestarazu A, André B (1994) Cloning and expression of the MEP1 gene encoding an ammonium transporter in Saccharomyces cerevisiae. EMBO J **13:** 3456

Marsh JF, Rakocevic A, Mitra RM, Brocard L, Sun J, Eschstruth A, Long SR, Schultze M, Ratet P, Oldroyd GED (2007) Medicago truncatula NIN is essential for rhizobial-independent nodule organogenesis induced by autoactive calcium/calmodulin-dependent protein kinase. Plant Physiol **144:** 324-335

Masalkar P, Wallace IS, Hwang JH, Roberts DM (2010) Interaction of cytosolic glutamine synthetase of soybean root nodules with the C-terminal domain of the symbiosome membrane nodulin 26 aquaglyceroporin. Journal of Biological Chemistry **285:** 23880-23888

Masson-Boivin C, Giraud E, Perret X, Batut J (2009) Establishing nitrogen-fixing symbiosis with legumes: how many rhizobium recipes? Trends Microbiol **17:** 458-466

Matsushima R, Fukao Y, Nishimura M, Hara-Nishimura I (2004) NAI1 gene encodes a basic-helix-loop-helix-type putative transcription factor that regulates the formation of an endoplasmic reticulum-derived structure, the ER body. Plant Cell **16:** 1536-1549

Matsushima R, Hayashi Y, Kondo M, Shimada T, Nishimura M, Hara-Nishimura I (2002) An endoplasmic reticulum-derived structure that is induced under stress conditions in Arabidopsis. Plant Physiol **130:** 1807-1814

Mergaert P, Uchiumi T, Alunni B, Evanno G, Cheron A, Catrice O, Mausset AE, Barloy-Hubler F, Galibert F, Kondorosi A, Kondorosi E (2006) Eukaryotic control on bacterial cell cycle and differentiation in the Rhizobium-legume symbiosis. Proc Natl Acad Sci U S A **103:** 5230-5235

Messinese E, Mun JH, Yeun LH, Jayaraman D, Rouge P, Barre A, Lougnon G, Schornack S, Bono JJ, Cook DR, Ane JM (2007) A novel nuclear protein interacts with the symbiotic DMI3 calcium- and calmodulin-dependent protein kinase of Medicago truncatula. Mol Plant Microbe Interact **20:** 912-921

Miao G, Hong Z, Verma D (1992) Topology and phosphorylation of soybean nodulin-26, an intrinsic protein of the peribacteroid membrane. The Journal of cell biology **118:** 481-490

Middleton PH, Jakab J, Penmetsa RV, Starker CG, Doll J, Kaló P, Prabhu R, Marsh JF, Mitra RM, Kereszt A, Dudas B, VandenBosch K, Long SR, Cook DR, Kiss GB, Oldroyd GED (2007) An ERF Transcription Factor in Medicago truncatula That Is Essential for Nod Factor Signal Transduction. The Plant Cell Online **19:** 1221-1234

Mikkelsen MD, Thomashow MF (2009) A role for circadian evening elements in cold-regulated gene expression in Arabidopsis. Plant J **60:** 328-339

Miller A, Cramer M (2005) Root nitrogen acquisition and assimilation. Root physiology: from gene to function**:** 1-36

Millhollon EP, Williams LE (1986) Carbohydrate partitioning and the capacity of apparent nitrogen fixation of soybean plants grown outdoors. Plant Physiol **81:** 280

Mitra RM, Gleason CA, Edwards A, Hadfield J, Downie JA, Oldroyd GED, Long SR (2004) A Ca2+/calmodulin-dependent protein kinase required for symbiotic nodule development: Gene identification by transcript-based cloning. Proc Natl Acad Sci U S A **101:** 4701-4705

Moreau S, Thomson RM, Kaiser BN, Trevaskis B, Guerinot ML, Udvardi MK, Puppo A, Day DA (2002) GmZIP1 encodes a symbiosis-specific zinc transporter in soybean. Journal of Biological Chemistry **277:** 4738-4746

Morey KJ, Ortega JL, Sengupta-Gopalan C (2002) Cytosolic glutamine synthetase in soybean is encoded by a multigene family, and the members are regulated in an organ-specific and developmental manner. Plant Physiol **128:** 182-193

Nagano AJ, Matsushima R, Hara-Nishimura I (2005) Activation of an ER-body-localized beta-glucosidase via a cytosolic binding partner in damaged tissues of Arabidopsis thaliana. Plant Cell Physiol **46:** 1140-1148

Nelson BK, Cai X, Nebenfuhr A (2007) A multicolored set of in vivo organelle markers for co-localization studies in Arabidopsis and other plants. Plant J **51:** 1126-1136

Newcomb E, Kaneko Y, VandenBosch K (1989) Specialization of the inner cortex for ureide production in soybean root nodules. Protoplasma **150:** 150-159

Nguyen T, Zelechowska M, Foster V, Bergmann H, Verma D (1985) Primary structure of the soybean nodulin-35 gene encoding uricase II localized in the peroxisomes of uninfected cells of nodules. Proceedings of the National Academy of Sciences **82:** 5040-5044

Ninnemann O, Jauniaux J, Frommer W (1994) Identification of a high affinity NH4+ transporter from plants. EMBO J **13:** 3464

Nishimura R, Ohmori M, Fujita H, Kawaguchi M (2002) A Lotus basic leucine zipper protein with a RING-finger motif negatively regulates the developmental program of nodulation. Proceedings of the National Academy of Sciences **99:** 15206-15210

Nusinow DA, Helfer A, Hamilton EE, King JJ, Imaizumi T, Schultz TF, Farré EM, Kay SA (2011) The ELF4-ELF3-LUX complex links the circadian clock to diurnal control of hypocotyl growth. Nature **475:** 398-402

Ogasawara K, Yamada K, Christeller JT, Kondo M, Hatsugai N, Hara-Nishimura I, Nishimura M (2009) Constitutive and inducible ER bodies of Arabidopsis thaliana accumulate distinct beta-glucosidases. Plant Cell Physiol **50:** 480-488

Oldroyd GE, Downie JA (2008) Coordinating nodule morphogenesis with rhizobial infection in legumes. Annu Rev Plant Biol **59:** 519-546

Ott T, Sullivan J, James EK, Flemetakis E, Günther C, Gibon Y, Ronson C, Udvardi M (2009) Absence of symbiotic leghemoglobins alters bacteroid and plant cell differentiation during development of Lotus japonicus root nodules. Molecular plant-microbe interactions **22:** 800-808

Owen A, Jones D (2001) Competition for amino acids between wheat roots and rhizosphere microorganisms and the role of amino acids in plant N acquisition. Soil Biology and Biochemistry **33:** 651-657

Panter S, Thomson R, de Bruxelles G, Laver D, Trevaskis B, Udvardi M (2000) Identification with proteomics of novel proteins associated with the peribacteroid membrane of soybean root nodules. Mol Plant Microbe Interact **13:** 325-333

Pantoja O (2012) High affinity ammonium transporters: molecular mechanism of action. Front Plant Sci **3:** 34

Pao SS, Paulsen IT, Saier Jr MH (1998) Major facilitator superfamily. Microbiology and Molecular Biology Reviews **62:** 1-34

Park DH, Somers DE, Kim YS, Choy YH, Lim HK, Soh MS, Kim HJ, Kay SA, Nam HG (1999) Control of circadian rhythms and photoperiodic flowering by the Arabidopsis GIGANTEA gene. Science **285:** 1579-1582

Pessi G, Ahrens CH, Rehrauer H, Lindemann A, Hauser F, Fischer HM, Hennecke H (2007) Genome-wide transcript analysis of Bradyrhizobium japonicum bacteroids in soybean root nodules. Molecular plant-microbe interactions **20:** 1353-1363

Peterson GL (1977) A simplification of the protein assay method of Lowry et al. which is more generally applicable. Anal Biochem **83:** 346-356

Pfeil BE, Schlueter JA, Shoemaker RC, Doyle JJ (2005) Placing Paleopolyploidy in Relation to Taxon Divergence: A Phylogenetic Analysis in Legumes Using 39 Gene Families. Systematic Biology **54:** 441-454

Popp C, Ott T (2011) Regulation of signal transduction and bacterial infection during root nodule symbiosis. Curr Opin Plant Biol **14:** 458-467

Prell J, White J, Bourdes A, Bunnewell S, Bongaerts R, Poole P (2009) Legumes regulate Rhizobium bacteroid development and persistence by the supply of branched-chain amino acids. Proceedings of the National Academy of Sciences **106:** 12477-12482

Radutoiu S, Madsen LH, Madsen EB, Felle HH, Umehara Y, Gronlund M, Sato S, Nakamura Y, Tabata S, Sandal N, Stougaard J (2003) Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. Nature **425:** 585-592

Rawat SR, Silim SN, Kronzucker HJ, Siddiqi MY, Glass ADM (1999) AtAMT1 gene expression and NH4+ uptake in roots of Arabidopsis thaliana: evidence for regulation by root glutamine levels. The Plant Journal **19:** 143-152

Reddy VS, Shlykov MA, Castillo R, Sun EI, Saier MH (2012) The major facilitator superfamily (MFS) revisited. FEBS Journal **279:** 2022-2035

Rogato A, D'Apuzzo E, Barbulova A, Omrane S, Stedel C, Simon-Rosin U, Katinakis P, Flemetakis M, Udvardi M, Chiurazzi M (2008) Tissue-specific down-regulation of LjAMT1; 1 compromises nodule function and enhances nodulation in Lotus japonicus. Plant Mol Biol **68:** 585-595

Saalbach G, Erik P, Wienkoop S (2002) Characterisation by proteomics of peribacteroid space and peribacteroid membrane preparations from pea (Pisum sativum) symbiosomes. Proteomics **2:** 325-337

Saier MH, Jr., Beatty JT, Goffeau A, Harley KT, Heijne WH, Huang SC, Jack DL, Jahn PS, Lew K, Liu J, Pao SS, Paulsen IT, Tseng TT, Virk PS (1999) The major facilitator superfamily. J Mol Microbiol Biotechnol **1:** 257-279

Saito K, Yoshikawa M, Yano K, Miwa H, Uchida H, Asamizu E, Sato S, Tabata S, Imaizumi-Anraku H, Umehara Y, Kouchi H, Murooka Y, Szczyglowski K, Downie JA, Parniske M, Hayashi M, Kawaguchi M (2007) NUCLEOPORIN85 is required for calcium spiking, fungal and bacterial symbioses, and seed production in Lotus japonicus. Plant Cell **19:** 610-624

Schaffer R, Ramsay N, Samach A, Corden S, Putterill J, Carré IA, Coupland G (1998) The late elongated hypocotyl mutation of Arabidopsis disrupts circadian rhythms and the photoperiodic control of flowering. Cell **93:** 1219-1229

Schlueter JA, Lin JY, Schlueter SD, Vasylenko-Sanders IF, Deshpande S, Yi J, O'Bleness M, Roe BA, Nelson RT, Scheffler BE, Jackson SA, Shoemaker RC (2007) Gene duplication and paleopolyploidy in soybean and the implications for whole genome sequencing. BMC Genomics **8:** 330

Schmittgen TD, Livak KJ (2008) Analyzing real-time PCR data by the comparative CT method. Nat Protoc **3:** 1101-1108

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang X-C, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the palaeopolyploid soybean. Nature **463:** 178-183

Sciotti MA, Chanfon A, Hennecke H, Fischer HM (2003) Disparate oxygen responsiveness of two regulatory cascades that control expression of symbiotic genes in Bradyrhizobium japonicum. Journal of Bacteriology **185:** 5639-5642

Seo PJ, Kim MJ, Park JY, Kim SY, Jeon J, Lee YH, Kim J, Park CM (2010) Cold activation of a plasma membrane-tethered NAC transcription factor induces a pathogen resistance response in Arabidopsis. Plant J **61:** 661-671

Seo PJ, Kim S-G, Park C-M (2008) Membrane-bound transcription factors in plants. Trends Plant Sci **13:** 550-556

Severin A, Woody J, Bolon Y-T, Joseph B, Diers B, Farmer A, Muehlbauer G, Nelson R, Grant D, Specht J, Graham M, Cannon S, May G, Vance C, Shoemaker R (2010) RNA-Seq Atlas of Glycine max: A guide to the soybean transcriptome. BMC Plant Biol **10:** 160

Simon-Rosin U, Wood C, Udvardi MK (2003) Molecular and cellular characterisation of LjAMT2;1, an ammonium transporter from the model legume <i>Lotus japonicus</i&gt. Plant Mol Biol **51:** 99-108

Singh S, Parniske M (2012) Activation of calcium- and calmodulin-dependent protein kinase (CCaMK), the central regulator of plant root endosymbiosis. Curr Opin Plant Biol **http://dx.doi.org/10.1016/j.pbi.2012.04.002**

Smit P, Raedts J, Portyanko V, Debellé F, Gough C, Bisseling T, Geurts R (2005) NSP1 of the GRAS Protein Family Is Essential for Rhizobial Nod Factor-Induced Transcription. Science **308:** 1789-1791

Smith PM, Atkins CA (2002) Purine biosynthesis. Big in cell division, even bigger in nitrogen assimilation. Plant Physiol **128:** 793-802

Somers DE, Schultz TF, Milnamow M, Kay SA (2000) ZEITLUPE encodes a novel clock-associated PAS protein from Arabidopsis. Cell **101:** 319-329

Streitner C, Danisman S, Wehrle F, Schoning JC, Alfano JR, Staiger D (2008) The small glycine-rich RNA binding protein AtGRP7 promotes floral transition in Arabidopsis thaliana. Plant J **56:** 239-250

Subramanian C, Woo J, Cai X, Xu X, Servick S, Johnson CH, Nebenfuhr A, von Arnim AG (2006) A suite of tools and application notes for in vivo protein interaction assays using bioluminescence resonance energy transfer (BRET). Plant J **48:** 138-152

Sullivan S, Jenkins GI, Nimmo HG (2004) Roots, cycles and leaves. Expression of the phosphoenolpyruvate carboxylase kinase gene family in soybean. Plant Physiol **135:** 2078

Swensen SM (1996) The evolution of actinorhizal symbioses: evidence for multiple origins of the symbiotic association. American Journal of Botany**:** 1503-1512

Tadege M, Wen J, He J, Tu H, Kwak Y, Eschstruth A, Cayrel A, Endre G, Zhao PX, Chabaud M (2008) Large - scale insertional mutagenesis using the Tnt1 retrotransposon in the model legume Medicago truncatula. The Plant Journal **54:** 335-347

Tominaga-Wada R, Iwata M, Nukumizu Y, Wada T (2011) Analysis of IIId, IIIe and IVa group basic-helix-loop-helix proteins expressed in Arabidopsis root epidermis. Plant Sci **181:** 471-478

Tusnady GE, Simon I (2001) The HMMTOP transmembrane topology prediction server. Bioinformatics **17:** 849-850

Udvardi MK, Day DA (1997) Metabolite transport across symbiotic membranes of legume nodules. Annual Review of Plant Physiology and Plant Molecular Biology **48:** 493-523

Udvardi MK, Price GD, Gresshoff PM, Day DA (1988) A dicarboxylate transporter on the peribacteroid membrane of soybean nodules. FEBS Lett **231:** 36-40

Van de Velde W, Zehirov G, Szatmari A, Debreczeny M, Ishihara H, Kevei Z, Farkas A, Mikulass K, Nagy A, Tiricz H, Satiat-Jeunemaitre B, Alunni B, Bourge M,

Kucho K, Abe M, Kereszt A, Maroti G, Uchiumi T, Kondorosi E, Mergaert P (2010) Plant peptides govern terminal differentiation of bacteria in symbiosis. Science **327:** 1122-1126

Van den Bosch K, Newcom E (1986) Immunogold localization of nodule-specific uricase in developing soybean root nodules. Planta **167:** 425-436

Vasse J, De Billy F, Camut S, Truchet G (1990) Correlation between ultrastructural differentiation of bacteroids and nitrogen fixation in alfalfa nodules. Journal of Bacteriology **172:** 4295-4306

Veereshlingam H, Haynes JG, Penmetsa RV, Cook DR, Sherrier DJ, Dickstein R (2004) nip, a Symbiotic Medicago truncatula Mutant That Forms Root Nodules with Aberrant Infection Threads and Plant Defense-Like Response. Plant Physiol **136:** 3692-3702

Venkateshwaran M, Cosme A, Han L, Banba M, Satyshur KA, Schleiff E, Parniske M, Imaizumi-Anraku H, Ane JM (2012) The Recent Evolution of a Symbiotic Ion Channel in the Legume Family Altered Ion Conductance and Improved Functionality in Calcium Signaling. Plant Cell **24:** 2528-2545

Verma DP, Hong Z (1996) Biogenesis of the peribacteroid membrane in root nodules. Trends Microbiol **4:** 364-368

Vinardell JM, Fedorova E, Cebolla A, Kevei Z, Horvath G, Kelemen Z, Tarayre S, Roudier F, Mergaert P, Kondorosi A, Kondorosi E (2003) Endoreduplication Mediated by the Anaphase-Promoting Complex Activator CCS52A Is Required for Symbiotic Cell Differentiation in Medicago truncatula Nodules. The Plant Cell Online **15:** 2093-2105

Vincill ED, Szczyglowski K, Roberts DM (2005) GmN70 and LjN70. Anion transporters of the symbiosome membrane of nodules with a transport preference for nitrate. Plant Physiol **137:** 1435-1444

von Wirén N, Merrick M (2004) Regulation and function of ammonium carriers in bacteria, fungi, and plants
Molecular Mechanisms Controlling Transmembrane Transport. *In*, Vol 9. Springer Berlin / Heidelberg, pp 95-120

Walsh K, McCully M, Canny M (1989) Vascular transport and soybean nodule function: nodule xylem is a blind alley, not a throughway. Plant Cell Environ **12:** 395-405

Wang H, Moore MJ, Soltis PS, Bell CD, Brockington SF, Alexandre R, Davis CC, Latvis M, Manchester SR, Soltis DE (2009) Rosid radiation and the rapid rise of angiosperm-dominated forests. Proceedings of the National Academy of Sciences **106:** 3853-3858

Wang MY, Siddiqi MY, Ruth TJ, Glass ADM (1993) Ammonium uptake by rice roots (II. Kinetics of 13NH4+ influx across the plasmalemma). Plant Physiol **103:** 1259-1267

Wang ZY, Tobin EM (1998) Constitutive Expression of the CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) Gene Disrupts Circadian Rhythms and Suppresses Its Own Expression. Cell **93:** 1207-1217

Watanabe S, Xia Z, Hideshima R, Tsubokura Y, Sato S, Yamanaka N, Takahashi R, Anai T, Tabata S, Kitamura K (2011) A map-based cloning strategy employing a residual heterozygous line reveals that the GIGANTEA gene is involved in soybean maturity and flowering. Genetics **188:** 395-407

Weaver CD, Shomer NH, Louis CF, Roberts DM (1994) Nodulin 26, a nodule-specific symbiosome membrane protein from soybean, is an ion channel. Journal of Biological Chemistry **269:** 17858

Werner AK, Sparkes IA, Romeis T, Witte CP (2008) Identification, biochemical characterization, and subcellular localization of allantoate amidohydrolases from Arabidopsis and soybean. Plant Physiol **146:** 418-430

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

Xie F, Murray JD, Kim J, Heckmann AB, Edwards A, Oldroyd GE, Downie JA (2012) Legume pectate lyase required for root infection by rhizobia. Proc Natl Acad Sci U S A **109:** 633-638

Yamada K, Hara-Nishimura I, Nishimura M (2011) Unique defense strategy by the endoplasmic reticulum body in plants. Plant Cell Physiol **52:** 2039-2049

Yano K, Yoshida S, Muller J, Singh S, Banba M, Vickers K, Markmann K, White C, Schuller B, Sato S, Asamizu E, Tabata S, Murooka Y, Perry J, Wang TL, Kawaguchi M, Imaizumi-Anraku H, Hayashi M, Parniske M (2008) CYCLOPS, a mediator of symbiotic intracellular accommodation. Proc Natl Acad Sci U S A **105:** 20540-20545

Yazdanbakhsh N, Sulpice R, Graf A, Stitt M, Fisahn J (2011) Circadian control of root elongation and C partitioning in Arabidopsis thaliana. Plant Cell Environ **34:** 877-894

Ye J, Rawson RB, Komuro R, Chen X, Dave UP, Prywes R, Brown MS, Goldstein JL (2000) ER stress induces cleavage of membrane-bound ATF6 by the same proteases that process SREBPs. Mol Cell **6:** 1355-1364

Yoon H-K, Kim S-G, Kim S-Y, Park C-M (2008) Regulation of leaf senescence by NTL9-mediated osmotic stress signaling in Arabidopsis. Mol Cells **25:** 438-445

Young ND, Debelle F, Oldroyd GED, Geurts R, Cannon SB, Udvardi MK, Benedito VA, Mayer KFX, Gouzy J, Schoof H, Van de Peer Y, Proost S, Cook DR, Meyers BC, Spannagl M, Cheung F, De Mita S, Krishnakumar V, Gundlach H, Zhou S, Mudge J, Bharti AK, Murray JD, Naoumkina MA, Rosen B, Silverstein KAT, Tang H, Rombauts S, Zhao PX, Zhou P, Barbe V, Bardou P, Bechner M, Bellec A, Berger A, Berges H, Bidwell S, Bisseling T, Choisne N,

Couloux A, Denny R, Deshpande S, Dai X, Doyle JJ, Dudez A-M, Farmer AD, Fouteau S, Franken C, Gibelin C, Gish J, Goldstein S, Gonzalez AJ, Green PJ, Hallab A, Hartog M, Hua A, Humphray SJ, Jeong D-H, Jing Y, Jocker A, Kenton SM, Kim D-J, Klee K, Lai H, Lang C, Lin S, Macmil SL, Magdelenat G, Matthews L, McCorrison J, Monaghan EL, Mun J-H, Najar FZ, Nicholson C, Noirot C, O/'Bleness M, Paule CR, Poulain J, Prion F, Qin B, Qu C, Retzel EF, Riddle C, Sallet E, Samain S, Samson N, Sanders I, Saurat O, Scarpelli C, Schiex T, Segurens B, Severin AJ, Sherrier DJ, Shi R, Sims S, Singer SR, Sinharoy S, Sterck L, Viollet A, Wang B-B, Wang K, Wang M, Wang X, Warfsmann J, Weissenbach J, White DD, White JD, Wiley GB, Wincker P, Xing Y, Yang L, Yao Z, Ying F, Zhai J, Zhou L, Zuber A, Denarie J, Dixon RA, May GD, Schwartz DC, Rogers J, Quetier F, Town CD, Roe BA (2011) The Medicago genome provides insight into the evolution of rhizobial symbioses. Nature **480:** 520-524

Zanetti ME, Blanco FA, Beker MP, Battaglia M, Aguilar OM (2010) A C subunit of the plant nuclear factor NF-Y required for rhizobial infection and nodule development affects partner selection in the common bean-Rhizobium etli symbiosis. Plant Cell **22:** 4142-4157

Zeilinger MN, Farré EM, Taylor SR, Kay SA, Doyle FJ (2006) A novel computational model of the circadian clock in Arabidopsis that incorporates PRR7 and PRR9. Molecular Systems Biology **2**

Zheng L, Kostrewa D, Berneche S, Winkler FK, Li X-D (2004) The mechanism of ammonia transport based on the crystal structure of AmtB of Escherichia coli. Proc Natl Acad Sci U S A **101:** 17090-17095

Zheng Y, Ren N, Wang H, Stromberg AJ, Perry SE (2009) Global identification of targets of the Arabidopsis MADS domain protein AGAMOUS-Like15. Plant Cell **21:** 2563-2577