2019

Characterization of nodule-specific, SLAC and MATE membrane transporters in Medicago truncatula

lina yang

Follow this and additional works at: https://researchrepository.wvu.edu/etd
Part of the Molecular Genetics Commons
Characterization of nodule-specific, SLAC and MATE membrane transporters in *Medicago truncatula*

Lina Yang

Dissertation submitted to the Davis College of Agriculture, Natural Resources and Design at West Virginia University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In Genetics and Developmental Biology

Vagner A. Benedito, Ph.D., Chair
Daniel Panaccione, Ph.D.
Jianbo Yao, Ph.D.
Nicole Waterland, Ph.D.
Jonathan Cumming, Ph.D.

Division of Plant and Soil Science

Morgantown, West Virginia

2019

**Keywords:** Auto regulation of nodulation (AON), Fabaceae, Gene Duplication, Metabolism Tellurite-resistance/Dicarboxylate transporter (TDT) family, Metabolite Transport, Mutualism, Secondary transporter

Copyright 2019 Lina Yang
ABSTRACT

Characterization of nodule-specific, SLAC and MATE membrane transporters in *Medicago truncatula*

Lina Yang

Legumes play a crucial role in sustainable agriculture because of their intrinsic ability to reduce atmospheric $\text{N}_2$ into $\text{NH}_3$ via symbiotic nitrogen fixation (SNF). SNF is carried out in the symbiosome, a quasi-organelle containing endosymbiotic rhizobial bacteria in the cytoplasm of infected cells of the nodule. Therefore, the bacteria are surrounded by a symbiosome membrane (SM), which is derived from the plant plasma membrane during infection. SNF requires constant nutritional exchanges between symbionts, including reduced carbon (dicarboxylates) from the plant for reduced nitrogen ($\text{NH}_4^+$) from the bacteroids. This exchange of nutrients and signals is fundamental to SNF and occurs through various transporters in several membranes, and critically through the SM. However, despite the fact that many nodule-specific or high-expressed putative transport genes have been identified in nodules at the genetic level, little is still known about their biochemical and physiological roles for SNF. Thus, we functionally characterized two transporter genes exclusively or highly expressed in the nodule of *Medicago truncatula*, a well-established legume model species: MtSLAH1 (*Medtr4g049640, TCDB 2.A.16.5*) of the Tellurite-resistance/Dicarboxylate transporter family, and MtMATE30 (*Medtr7g082810, TCDB 2.A.66.1*) of the Multidrug and toxic compound Extrusion family. There are seven SLAC (Slow Anion Channel associated) and 70 MATE (Multidrug and Toxic compound Extrusion) family members in the *M. truncatula* genome. MtSLAH1 is the SLAC transporter with highest expression value in nodules among seven members. Its expression is reduced 5-fold in 2 days after application of nitrate, a known SNF repressor. MtSLAH1 is highly expressed in infected cells and requires bacteroid differentiation for induction. MtSLAH1 channel permeability to dicarboxylates was tested by patch-clamp of *Xenopus* oocytes, but no current was detected. MtSLAH1 either requires a cofactor for activation or facilitates the efflux a different anion (e.g., nitrate, chloride). On the other hand, MtMATE30 is the MATE gene with highest expression in nodules. It is induced 100-fold by low nitrogen while strongly repressed by nitrate, suggesting a role in symbiotic nitrogen fixation. It starts to express in immature nodules (6 days post-inoculation, dpi), reaches its peak in young, mature nodules (10 dpi) and maintains consistent expression in older nodules. It is expressed in all nodule zones, except the meristem. MtMATE30 canonically belongs to a phylogenetic clade that includes transporters with affinity to alkaloids. Trigonelline is a widely distributed alkaloid and commonly found in legumes. Previous research showed that rhizobia are able to catabolize trigonelline by the trc gene located in the rhizobial pSym megaplasmid. We confirmed the presence of trigonelline in nodules of *M. truncatula* and showed MtMATE30 affinity to trigonelline in a heterologous bacteria system. Altogether, MtMATE30 may be participating in alkaloid metabolism of nodule cells, although its precise physiological role in symbiotic nitrogen fixation still requires further investigation. In the big picture, these two membrane transporters studied here are only two examples among thousands of transporters exclusively expressed in nodule. The legume research community needs to focus more efforts to understand how legumes and
rhizobia communicate and cooperate to fix nitrogen in order to enhance SNF efficiency in legume crops, and possibly to eventually extend it into non-legume crops.
Table of Contents

List of Figures........................................................................................................................................vi
List of Tables........................................................................................................................................vii
Supplementary Tables and Figures........................................................................................................vii

Chapter 1

Literature Review......................................................................................................................................2
Nitrogen as a plant nutrient and symbiotic nitrogen fixation in legumes ..............................................2
Early steps of nodule development and establishment of the endosymbiosis ........................................4
Nodule organogenesis .................................................................................................................................6
The symbiosome membrane .......................................................................................................................8
Membrane transporters in the nodule .........................................................................................................9
Tools for studying rhizobia-legume symbiosis .........................................................................................12
Conclusion..................................................................................................................................................18
References...................................................................................................................................................18

Chapter 2

Genomic analysis of Multidrug And Toxic Compound Extrusion (MATE) transporters in the model legume Medicago truncatula .........................................................................................42
Introduction ................................................................................................................................................43
Results and Discussion...............................................................................................................................46
  Genome-wide identification of MATE transporters in M. truncatula ....................................................46
  Phylogenetic analyses of the M. truncatula MATE family .....................................................................46
  Phylogeny of MATE transporters coupled with gene expression, subcellular localization and mutant analyses are useful to broadly predict potential physiological roles ........................................51
  Chromosomal distribution and structural analysis of MATE genes .......................................................52
  Duplication patterns of MATE genes in the M. truncatula genome .....................................................55
  Expression patterns of the MATE transporter gene family .....................................................................61
Conclusions...............................................................................................................................................62
Methods and materials.............................................................................................................................63
References..................................................................................................................................................64

Chapter 3

Functional characterization of a nodule-specific MATE membrane transporter in Medicago truncatula .............................................................................................................................88
Summary ....................................................................................................................................................88
Introduction ...............................................................................................................................................89
Results and Discussion..............................................................................................................................92
  MtMATE30 is closely related to MATE transporters with affinity to alkaloids ........................................92
  MtMATE30 is a nodule-specific transporter expressed throughout the whole indeterminate nodule, except the meristem .....................................................................................................95
MtMATE30 is expressed in all of the zones of the indeterminate nodule, except the meristem .................................................................................................................. 98
Nodule phenotype of MtMATE30 RNAi hairy roots ...................................................... 100
MtMATE30 is an alkaloid-specific transporter .............................................................. 100

Material and Methods .............................................................................................. 106
References .................................................................................................................. 112
Supplementary figure and table ................................................................................. 139

Chapter 4

An interspecific genomic analysis of the SLAC gene family between Arabidopsis thaliana and Medicago truncatula ................................................................. 142
Summary ..................................................................................................................... 142
Introduction ................................................................................................................ 143
Results ......................................................................................................................... 150
Genome-wide identification of SLAC channels in A. thaliana and M. truncatula genomes .................................................................................................................. Error! Bookmark not defined.
Phylogeny and structure analyses of plant SLAC genes ............................................. 151
Structural analysis of SLAC genes in the Rosids species under analysis .................... 155
Duplication patterns of SLAC genes in plant genomes .............................................. 159
Expression patterns of the SLAC genes in A. thaliana and M. truncatula ................. 162
Conclusion ................................................................................................................... 167
Material and Methods .............................................................................................. 168
References .................................................................................................................. 170
Supplemental figures and table ................................................................................. 194

Chapter 5

Functional characterization of the nodule-specific SLAH1 membrane transporter in Medicago truncatula ................................................................. 199
Summary ..................................................................................................................... 199
Introduction ................................................................................................................ 200
Results and Discussion ............................................................................................. 203
MtSLAH1 is a nodule-specific TDT/SLAC transporter ............................................... 203
MtSLAH1 is expressed in all of the zones of the indeterminate nodule ...................... 207
MtSLAH1 knockdown nodules .................................................................................. 208
MtSLAH1 likely requires an interacting protein to function ...................................... 210
Conclusion ................................................................................................................... 211
Material and Methods .............................................................................................. 212
Supplementary table ................................................................................................. 237
Conclusion and perspective ....................................................................................... 240
## List of Figures

### Chapter 2

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Phylogenetic tree of MATE transporters</td>
<td>p39</td>
</tr>
<tr>
<td>2</td>
<td>Phylogenetic analysis, expression analysis, and gene architecture of MATE transporter family in the <em>Medicago truncatula</em> genome</td>
<td>p40</td>
</tr>
<tr>
<td>3</td>
<td>Synteny analysis within the <em>Medicago truncatula</em> genome</td>
<td>p43</td>
</tr>
<tr>
<td>4</td>
<td>Synteny analysis between <em>Medicago truncatula</em> and <em>Arabidopsis thaliana</em> genomics.</td>
<td>p43</td>
</tr>
</tbody>
</table>

### Chapter 3

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MtMATE30 genomic organization and predicted membrane topology</td>
<td>p63</td>
</tr>
<tr>
<td>2</td>
<td>The phylogenetic tree of MtMATE30 with other 44 functionally characterized plant MATE transporters.</td>
<td>p64</td>
</tr>
<tr>
<td>3</td>
<td>Temporal and spatial expression profile of MtMATE30 in <em>Medicago truncatula</em></td>
<td>p66</td>
</tr>
<tr>
<td>4</td>
<td>Expression profile of MtMATE30 in inoculated roots by quantitative RT-PCR.</td>
<td>p66</td>
</tr>
<tr>
<td>5</td>
<td>MtMATE30 promoter activity analysis by GUS staining.</td>
<td>p67</td>
</tr>
<tr>
<td>6</td>
<td><em>In situ</em> hybridization of nodule section using a MtMATE30 probe</td>
<td>p69</td>
</tr>
<tr>
<td>7</td>
<td>Phenotype of MtMATE30 knockout nodules.</td>
<td>p70</td>
</tr>
<tr>
<td>8</td>
<td>Effects of tetrabutylammonium (TBA) on bacterial growth.</td>
<td>p71</td>
</tr>
<tr>
<td>9</td>
<td>Identification of the alkaloid trigonelline in mature nodules of <em>M. truncatula</em></td>
<td>p72</td>
</tr>
<tr>
<td>10</td>
<td>MtMATE30-mediated trigonelline transport in bacteria cells.</td>
<td>p74</td>
</tr>
<tr>
<td>11</td>
<td>Possible metabolic pathway of trigonelline in infected cells of nodules.</td>
<td>p75</td>
</tr>
</tbody>
</table>

### Chapter 4

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Phylogenetic analysis of SLAC family in the <em>Medicago truncatula</em> and <em>Arabidopsis thaliana</em> genomes.</td>
<td>p106</td>
</tr>
<tr>
<td>2</td>
<td>Predicted SLAC gene structures of the SLAC members</td>
<td>p108</td>
</tr>
<tr>
<td>3</td>
<td>Predicted membrane topology of plant SLAC channels.</td>
<td>p110</td>
</tr>
<tr>
<td>4</td>
<td>Microsynteny analysis within the <em>Medicago truncatula</em> genome</td>
<td>p113</td>
</tr>
<tr>
<td>5</td>
<td>Synteny analysis between the <em>Medicago truncatula</em> and <em>Arabidopsis thaliana</em> genomes.</td>
<td>p115</td>
</tr>
<tr>
<td>6</td>
<td>Expression profile of SLAC genes in the Rosids model species, <em>Arabidopsis thaliana</em> and <em>Medicago truncatula</em>.</td>
<td>p115</td>
</tr>
</tbody>
</table>

### Chapter 5

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MtSLAH1 genomic organization and putative transmembrane domains.</td>
<td>p136</td>
</tr>
<tr>
<td>2</td>
<td>Temporal and spatial expression profiles of MtSLAH1.</td>
<td>p138</td>
</tr>
<tr>
<td>3</td>
<td>Expression profile of MtSLAH1.</td>
<td>p139</td>
</tr>
<tr>
<td>4</td>
<td>MtSLAH1 <em>in situ</em> hybridization in nodules of <em>Medicago truncatula</em></td>
<td>p140</td>
</tr>
<tr>
<td>5</td>
<td>Phenotype of MtSLAH1 mutant plants</td>
<td>p141</td>
</tr>
<tr>
<td>6</td>
<td>Prediction of phosphorylation sites in MtSLAH1</td>
<td>p142</td>
</tr>
</tbody>
</table>
# List of Tables

<table>
<thead>
<tr>
<th>Chapter 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1. Tandem duplications of MATE transporter genes in the <em>M. truncatula</em> genome</td>
<td>p44</td>
</tr>
<tr>
<td>Table 2. Syntenic paralogs of MATE transporter genes in the <em>M. truncatula</em> genome</td>
<td>p46</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 4</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1. Comparison of general properties of SLAC family in two species</td>
<td>p103</td>
</tr>
<tr>
<td>Table 2. Percent identity matrix of all SLAC members of two species. Matrix was created by Clustal Omeg</td>
<td>p107</td>
</tr>
</tbody>
</table>

## Supplementary Tables and Figures

<table>
<thead>
<tr>
<th>Chapter 3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplemental Table S1 List of primers used in these studies</td>
<td>p92</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 4</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplemental Table S1 List of primers used in these studies</td>
<td>p130</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 5</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplemental Table S1 List of primers used in these studies</td>
<td>p151</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplementary Figure S1. Binary vector maps</td>
<td>p91</td>
</tr>
<tr>
<td>Supplementary Figure S2 Temporal and spatial expression of MtMATE30 in different zones of mature nodules.</td>
<td>P92</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 4</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplementary Figure S1. Alignment of SLAC channels in A. thaliana and M. truncatula.</td>
<td>p127</td>
</tr>
<tr>
<td>Supplementary Figure S2. Phylogenetic analysis of SLAC family in the Medicago truncatula, Arabidopsis thaliana plus other four plant species.</td>
<td>p129</td>
</tr>
<tr>
<td>Supplementary Figure S3. Promoter activity of SLAC1 ortholog in Medicago truncatula.</td>
<td>p130</td>
</tr>
<tr>
<td>Supplementary Figure S4. Binary vector map.</td>
<td>p130</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 5</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplementary Figure S1. Temporal and spatial expression profile of MtSLAH1 in nodules of Medicago truncatula</td>
<td>P151</td>
</tr>
</tbody>
</table>
Characterization of nodule-specific, SLAC and MATE membrane transporters in *Medicago truncatula*

In order to support a growing human population, the demand for nitrogen fertilizer in agricultural fields around the world is steadily increasing every year. However, the excessive nitrogen fertilizer application has a great impact on natural environments. Thus, modern approaches to sustainable agriculture lean on optimizing the use of biological resources to enhance agricultural productivity while minimizing environmental impact. Among these biological resources, the symbiotic nitrogen fixation (SNF) established between rhizobia and legumes is a low-cost, safe and efficient process that supplies 40 million tonnes of fixed nitrogen into agricultural systems each year, which is equivalent to almost 40% of the total world’s demand for nitrogen fertilizer (Young and Udvardi, 2009; FAO, 2016). This capability enables legumes to grow well in nitrogen-poor soils without the addition of organic or industrial nitrogen fertilizer, as well as providing the nutrient to subsequent non-legume crops. Therefore, understanding how legumes and rhizobia communicate and cooperate to fix nitrogen is key to enhance SNF efficiency in legumes and possibly extend it into non-legume crops.

SNF, the species-specific mutualism between rhizobia and legumes, is a complex process that requires intensive communication between the symbionts through small molecules, receptors, and transporters on and across several types of cell membranes. Although many nodule-specific or high-expressed putative transporter genes have been identified in legume nodules at the genetic level, little is still known about their biochemical and physiological roles for nodule development and SNF. Thus, the aim of this study is characterize two electrochemical potential-driven membrane transporters exclusively expressed in the nodule of model legume *Medicago truncatula* that may play a role in nodule development and function. MtSLAH1 (*Medtr4g049640*, TCDB 2.A.16.5) belongs to the Tellurite-resistance/Dicarboxylate transporter family, while MtMATE30 (*Medtr7g082810*, TCDB 2.A.66.1) is member of the Multi-Antimicrobial Extrusion transporter family.

We, therefore aimed to:

- Fine-tune the spatial-temporal expression domains of both genes;
- Determine the biochemical and physiological properties of these transporters;
• Make progress towards understanding the physiological roles of these transporters for SNF by using RNA interference (RNAi) knockdown.

Chapter 1

Literature Review

Nitrogen as a plant nutrient and symbiotic nitrogen fixation in legumes

Nitrogen is a primary element in all organisms, including plants and is a constituent of proteins, nucleic acids, and other biologically active substances, such as chlorophyll, alkaloids, and ATP. It represents 1-5% of plant’s dry matter and stimulates the aboveground growth of plants. Although it is the most abundant element in the atmosphere (78% as nitrogen gas, N₂), nitrogen is often the most limiting resource in the soil for plant growth. This limitation is due to the N₂ has an extremely strong triple covalent bond in N₂, making it quite stable, which cannot be directly assimilated by the plants, unless it is fixed (reduced) into the biological active forms: ammonia (NH₃) or nitrate (NO₃⁻). There are three main ways to fix N₂: the industrial (Haber-Bosch) process, natural chemical processes (e.g., lightening), and biological nitrogen fixation by diazotrophic prokaryotes.

Industrial process based on the Haber-Bosch process reduces N₂ into ammonia (NH₃) under the presence of a catalytic and extremely high pressure (around 200 atm) and temperature (at least 400° C) (Witschi, 2000). It was invented in the early 20th century and has solved the shortage of nitrogen compounds for human needs. Currently, half of global food production relies on synthetic nitrogen fertilizers as the consequence of human population explosion of the early 20th century (Smil, 1997). The world demand for nitrogen fertilizer in 2015 was estimated at 116 million metric tons, with a successive growth of 2% per year (FAO, 2016). Although vital for sustaining almost half of the Earth’s population, the industrial nitrogen process and excessive nitrogen fertilizer application leads to several serious problems. First, it is the largest single cost in conventional cropping systems. Second, nitrogen loss from the field causes soil pollution as well as water table contamination and eutrophication of water bodies. Nitrate and ammonium, the forms typically present in fertilizers, are very mobile in the soil, and crops can utilize only 30-40%
of the applied fertilizer (Kant et al., 2010; Li et al., 2015a). This means that more than 60% of N fertilizer is lost through a combination of leaching, surface run-off, denitrification, volatilization, and microbial consumption. The most pronounced results of eutrophication are the increasingly noxious production of green algae, and phytoplankton that decreases water quality and impairs water clarity. Last, but not least, the industrial energy demand to ammonia production has severe effects on the environment as it requires fossil fuels, and generates greenhouse gases, such as CO₂ and nitrous oxide (Tilman et al., 2002; D’haeseleer et al., 2010).

Due to the drawbacks of industrial nitrogen fixation, alternative processes are needed. The natural chemical process neither can provide enough nitrogen for agricultural production nor be easily manipulated. Therefore, the viable alternative to replenishing soil nitrogen by the addition of commercial fertilizers is the symbiotic nitrogen fixation (SNF). The ability to fix nitrogen is only found in prokaryotes. Diazotrophic bacteria can reduce N₂ by using an enzymatic complex called nitrogenase. Some diazotrophic groups, such as Methanococcus and Erwinia, are free-living organisms. However, their contribution of SNF is limited, encompassing only 2-3% of the total N fixed biologically (Giller and Day, 1985; Do Vale Barreto Figueiredo et al., 2013). Other diazotrophs can fix nitrogen in symbiotic associations with legumes as well as a few non-legumes plant species. Symbiotic associations are favored because of the constant availability of metabolic energy provided by the host plant. There are three major types of nitrogen-fixing symbioses with plant species. The first and most important is established between rhizobial bacteria (Rhizobium, Bradyrhizobium, Azorhizobium, Mesorhizobium, and Sinorhizobium) with roots of over 1700 legume plant species in three subfamilies (Mimosoideae, Ceasealpinoideae, and Papilionoideae) (Wall, 2000). The second one occurs between the filamentous actinomycete Frankia and actinorhizal plants of over 25 genera of selected species in the orders Cucurbitales, Fabales, Fagales, and Rosales (e.g., Alnus spp.) (Wall, 2000; Pawlowski and Sirrenberg, 2003; Santi et al., 2013). The last one takes place with cyanobacteria and diverse plant groups, including dicots, cycads, ferns, liverworts, and hornworts (Meeks, 1998; Adams et al., 2013).

Among these symbiotic associations, the mutualism between rhizobia and legumes is very significantly economically. Legumes (Fabaceae family) are the third largest family of flowering plants, and the second most important crop family. Many studies indicate that nodulation has arisen independently multiple times within the Rosid I clade
of angiosperms. Most remarkably, however, around 90% of the species in the legume family can associate with more than 70 species of rhizobia, and fix nitrogen symbiotically (Brewin, 2004). Degradation of legume tissues and excesses of fixed nitrogen released by legumes roots enrich agricultural soils for nearby plants or subsequent crops. SNF contributes 40 million tons of fixed nitrogen into agricultural systems each year, which is equivalent to almost 40% of the total world’s demand for nitrogen fertilizer without any adverse effects to the environment (Rome, 2011; FAO, 2016). This mutualistic relationship makes legumes unique among crops, and a key component of sustainable and environmentally safe agricultural practices.

In order to break the triple covalent bond in the \( \text{N}_2 \) molecule, nitrogenase utilizes 16-18 molecules of ATP to catalyze the six-electron reduction of \( \text{N}_2 \) to \( \text{NH}_3 \) along with an associated reduction of 2 H\(^+\) to \( \text{H}_2 \). Expectedly, this is an energetically very expensive reaction. The respiration of nitrogen-fixing bacteria requires a constant flux of \( \text{O}_2 \), which must be achieved at a very low concentration of free oxygen condition as oxygen binds to nitrogenase making it permanently nonfunctional (Thorneley and Ashby, 1989; Mosher et al., 1994; Soupène et al., 1995; Lodwig and Poole, 2003; Xie et al., 2011). Therefore, in most cases, the nitrogen fixation machinery is spatially segregated in specialized cells of organs like the root-derived nodule, which creates ideal environmental conditions for SNF by the microsymbiont. Therefore, understanding how legumes and rhizobia communicate and cooperate to achieve effective SNF will allow us to enhance SNF efficiencies and perhaps extend this process to non-legume crops.

**Early steps of nodule development and establishment of the endosymbiosis**

In legumes, SNF occurs in an organ developed from root tissues, called the nodule. The development of nodules is initially triggered by a molecular dialogue between the two symbionts that culminates with nodule development, the engulfing of bacteria into specific cells, and the establishment of SNF. Although the steps in nodule development and the final histological organization of the organ may differ considerably among legumes, the establishment of the *Rhizobium*-legume symbiosis is promoted physiologically in the plant by nitrogen deficiency.
In response to nitrogen limitation, legume roots secrete flavonoids (luteolin, methoxychaicone, naringenin, genistein, daidzein) as well as non-flavonoid (trigonelline and stachydrine) into the rhizosphere, which specify which rhizobial species will interact. Differently from the generalistic mycorrhizal mutualism, the legume-Rhizobium mutualism is highly species-specific. When the legume signals bind to the rhizobial transcriptional activator \textit{NodD}, it triggers the expression of \textit{Nod} (nodulation) genes located on the symbiotic plasmids \textit{pSymA} and \textit{pSymB} (Long, 1996). \textit{NodA}, \textit{NodB}, and \textit{NodC} gene products are responsible for synthesizing the backbone of nodulation signals known as Nod factors (NFs). NFs generally consist of lipochitooligosaccharides (LCOs) comprised of a chitin backbone with three to five residues of beta-(1,4)-linked N-acetyl-D-glucosamine (GlcNAC), in which the non-reducing sugar moiety is substituted with an acetyl chain at the C2 position (Long, 1996; Brewin, 2004). Specific bacterial nodulation genes control the presence and identity of these substitutions to produce unique molecules. For example, the nodulation genes \textit{NodH}, \textit{NodP}, \textit{NodQ}, \textit{NodE}, \textit{NodF}, and \textit{NodL} are responsible for the decoration of the LCO side chain, which involves host specificity and may be essential to trigger the formation of the infection thread (Barnett et al., 2001; Smit et al., 2007; Wang et al., 2012). Additionally, \textit{NodI} and \textit{NodJ} are involved in NF secretion (Cárdenas et al., 1996).

Once synthesized and exuded to the rhizosphere, NFs bind to specific, heterodimeric lysine-motif (LysM) receptor-like kinases, such as NFP1 and NFR5 in \textit{Lotus japonicus}, and LYK3 and LYK4 in \textit{Medicago truncatula} and \textit{Medicago sativa} (Gough, 2003; Smit et al., 2007; Streng et al., 2011). These receptors are localized on the plasma membrane of root epidermal cells and trigger signal transduction cascades that induce a cellular response in root hairs within minutes (Gough, 2003). These include root hair deformation (formation of shepherd’s hook) (Dudley and Long, 1989; Heidstra et al., 1997; Amor et al., 2003), cytoplasmic Ca$^{2+}$ spiking (Mitra et al., 2004; Kanamori et al., 2006), induction of early nodulin genes (Heidstra et al., 1997; Sauviac et al., 2005), and rearrangement of cytoskeletal components (Cárdenas et al., 2003; Brewin, 2004; Yokota et al., 2009). Hair deformation is induced within 3 hours (Heidstra et al., 1997), while induction early nodulin gene expression occurs around 6 hours post-inoculation. Mutations leading to loss of function in either of these receptors eliminate almost all Nod factor-inducible responses and cause a specific block of infection thread formation (Smit
et al., 2007). Thus, NFs that initiate signal transduction signaling cascades are the foundation for rhizobial invasion and symbiotic nitrogen fixation in legumes.

An important morphological effect associated with Nod factors is root hair deformation and curling, creating the so-called “shepherd’s hook,” which entraps rhizobia. Rhizobia inside the hook lead to deforming a small susceptible region that invites the plasma membrane and deposit of new cell wall material, forming a plant-derived tube-like structure known as the infection thread (IT). The IT elongates and penetrates the root cortical cell layers towards an already dividing nodule primordium. Therefore, IT acts as a cytoplasmic passage for rhizobia across several cell layers of the root. Once inside the IT, rhizobia are surrounded by a plant-derived fluid extracellular matrix and undergo active division within approximately at 60 µm from the tip of the root hair. Beyond this point, the extracellular matrix glycoprotein is cross-linked driven by hydrogen peroxide, which solidifies the matrix and halts cell division of the bacteria (Brewin, 2004).

Rhizobia are subsequently released into the cytoplasm of specific cells of the developing nodule primordium as infection droplets, which are endocytic vesicles enclosed by membrane derived from the plasma membrane of the plant cell. The newly formed endocytic vesicle is called the symbiosome, an organelle-like compartment enclosing symbiotic bacteria. Inside of the symbiosome, the bacteria differentiate morphologically and physiologically into rod-like cells called bacteroids (Brewin, 2004; Catalano et al., 2007; Udvardi and Poole, 2013). As rhizobia colonize the plant cell cytoplasm, the nodule primordium develops into the root nodule. Nodule organogenesis occurs as a result of reprogramming of differentiated root cells within the epidermis, cortex, and the pericycle. This new organ is connected to the nutrient transport system of the host plant by xylem and phloem vasculature, which contrary to the root, is located in the periphery of the nodule.

**Nodule organogenesis**

Depending on the species, legume nodules can be determinate or indeterminate. Mature determinate nodules (e.g., Lotus japonicus, Glycine max) present a spherical-shape and lack a persistent nodule meristem or distinct developmental zones in the nodule cortex. On the other hand, the indeterminate nodule (e.g., M. truncatula, Trifolium repens) is
formed as it establishes an apical meristem and ITs invade the cells below it. The meristem remains active in the mature indeterminate nodule, making it acquire an elongated form. Therefore, indeterminate nodules show a well-defined histology with a gradient of distinct developmental stages, which are divided into five zones: a persistent nodule meristem (zone I), an invasion zone (zone II), a bacteroid maturation or Interzone (zone II-III), a nitrogen fixation zone (zone III), and a senescence zone (zone IV).

In zone I, the nodule primordium arises in the uninfected inner layer of the root cortex adjacent to the pericycle near the xylem pole. These cells, together with the adjacent pericycle cells, are stimulated to divide and constitute the apical meristem of the incipient nodule (Brewin, 1991). This process is triggered by plant hormones, especially cytokinin (Frugier et al., 2008; Desbrosses and Stougaard, 2011), and happens as root hairs perceive NFs and IT starts to develop, but before IT arrives at the forming nodule. The nature of this translocated signal remains elusive to date (Van Spronsen et al., 2001). As the IT arrives at the nodule, bacteria are released and endocytically engulfed by plant cells in infection zone II. This zone can be further divided (in reference to the root system) into distal and proximal regions. In the distal infection zone, about four cell layers below the nodule meristem, bacteria are released from ITs and individually pinched off into the host’s cell cytoplasm through an endocytosis-like process that forms the plant-derived symbiosome membrane from the plant cell plasma membrane. Once engulfed by the plant cell, the bacteria undergo DNA endoreduplication, in certain legumes, they divide a few times. In the model legume *M. truncatula*, each symbiosome only contains a single bacteria (Limpens et al., 2013), while in other legumes, symbiosomes are capable of hosting a few bacteroids each. In the proximal zone II, which encompass about the next four cell layers, the bacteria stop dividing and start to terminally differentiate by enlarging 10-fold. During this stage, they lose their free-living status and start completely depending metabolically on the host (Vasse et al., 1990). In zone III, the bacteria are fully differentiated into a functional symbiotic form, with nitrogen fixation taking place inside the infected cells of the nodule. It is noteworthy that some cells in zone III remain uninfected. These uninfected cells may play an essential role in metabolite transport in functional nodules. Carbon-derived molecules generated from photosynthesis are transported mainly as sucrose via the phloem to roots and nodules, which are thought to be converted in uninfected cells of the nodule into dicarboxylic acids (e.g., malate and succinate) via the TCA cycle and further transferred to feed infected cells (Chen et al., 2010a). In *M.
truncatula, uninfected cells are also thought to play a role in the export of fixed nitrogen in the form of amides, especially asparagine, to the rest of the plant. Between zone II and zone III, there is an intermediate zone called Interzone II-III consisting of two to three layers of amyloplast-rich cells, which separate the pre-fixing zone II from the N₂-fixing zone. This zone is an important landmark in nodule differentiation, with several key symbiotic genes from both species, including nif, fix, and leghemoglobin, turned on very precisely at this stage (Soupène et al., 1995). In zone IV, after about two weeks of nodule development, symbiosomes and infected cells go through programmed cell death and establish a senescence zone (IV) proximal to zone III. Upon aging, this zone gradually moves in a proximal-distal direction until it reaches the organ apex and the whole nodule degenerates. All nodule zones (I-IV), except for the meristem, are surrounded at the periphery by the nodule parenchyma (inner cortex), vascular bundles, and the nodule endodermis. Furthermore, the entire nodule is surrounded by an outer cortex and an epidermis covered with a wax-like substance, which developed from nodule primordia (Vasse et al., 1990; Limpens et al., 2013). The outer cortex contains two or more peripheral vascular bundles, which connect to root vascular bundles, and is responsible for the exchange of nutrients between roots and nodules (Guan et al., 2013). Interesting, vacuolated outer cortical cells are critical for the formation of infection threads via the pre-infection thread structures forming in the cortex (Van Spronsen et al., 2001).

The symbiosome membrane

The symbiosome membrane that surrounds the bacteroid is derived from the host’s cell plasma membrane during the endocytic infection process. However, it undergoes rapid transformation into a specialized endosymbiotic membrane having properties that are common with the vacuole, such as the presence of V-type H⁺-ATPase pumps (Mylona et al. 1995, Catalano, et al. 2004). The formation of the symbiosome membrane (SM) involves the embedding of unique proteins and molecules in coordination with the synthesis of membrane components with bacterial cell division during the early phases of nodule development and cell infection (Verma and Hong, 1996). The SM serves two major purposes. First, it provides a physical barrier between the bacteria and the host cell cytoplasm, thus avoiding any direct contact between host cell and the invading prokaryote, which otherwise could trigger the host’s defense response and halt the symbiosis.
Furthermore, the SM also provides a shield for rhizobia from the exterior environment, making it ideal for nitrogen fixation. As free-living organisms, rhizobia are strictly aerobic, but they become microaerobes as bacteroids develop in the nodule (Xie et al., 2011). Additionally, the SM controls the exchange of metabolites and signal molecules between the two partners. The SM allows for the transport of metabolites (e.g., carbon skeletons, amino acids), minerals and other chemical signals to the bacteroid, the establishment of a proton motive force that enables secondary metabolite transport, the export of fixed nitrogen as ammonia in addition to the heme cofactor from the bacteroid to the plant's cell cytoplasm for leghemoglobin synthesis, and perhaps the cycling amino acids. Minerals such as zinc, iron, magnesium, and sulfate need to be regularly supplied by the host cell to the microsymbiont. In order to be moved into and out of bacteroids, given their chemical structure of the compounds exchanged, many require specific transporters on the symbiosome membrane (Vincill et al., 2005; Gaude et al., 2012; Kryvoruchko et al., 2016). Most components on the SM are currently unknown.

Membrane transporters in the nodule

The SM serves as a critical interface between the symbiotic partners and represents a regulatory gateway for transporting solutes between symbionts under plant control. Therefore, the SM is proposed to contain an array of transporters and channels to facilitate the movement to and from the endosymbiont (Vincill et al., 2005; Kiirika et al., 2012; Datta et al., 2015).

Two main types of membrane transporter proteins interact with substrates: carriers and channels. Channels are huge proteins, of which multiple subunits are arranged in a cluster with a central hydrophilic pore that facilitates the translocation of molecules or ions across the membrane by energetically favorable electrochemical gradients (Zimmermann and Sentenac, 1999; Ward et al., 2009; Hedrich, 2012). Two factors that affect the flow of ions and water through an open channel are the membrane potential (the proton motive force, pmf) and the concentration gradient. This form of transport is called facilitated diffusion. Properties of channels relate selectivity and gating (Hedrich, 2012). Selectivity depends on the size as well as the chemical property of the channel pore. The movement of molecules through a channel can be controlled by various mechanisms that result in opening or closing of the passageway (Schmidt and Schroeder, 1994; Zimmermann and
Sentenac, 1999; Hedrich, 2012). On the other hand, carriers bind to specific solutes and undergo conformational changes to move substrates across the membrane. Several channels on the SM have been reported. Soybean nodulin (nod26) as an ion channel is located on SM and facilitates the movement of ammonia and water, which was tested in liposomes (Hwang et al., 2010). Another example is GmN70, which acts as anion channels and has great permeability to nitrate in oocytes (Vincill et al., 2005). Depending on the thermodynamics of the system, carriers can be further divided into two categories: passive or active carriers, the latter being further subdivided into primary active carriers or secondary active carriers. Passive carriers or uniporters, like channels, also move substrates favorable to the concentration gradient and do not require direct energy. Primary active carriers, on the other hand, also referred to as ATPase pumps use energy from ATP hydrolysis to move substrates against chemical or electric gradients across the membrane. The SM contains two types of ATPases: K⁺/H⁺-ATPases, which are typical of the plasma membrane, and Mg²⁺/H⁺-ATPases, which are usually present in the Golgi membrane and the tonoplast (Catalano et al., 2004; Limpens et al., 2005). There was also Ca²⁺ pumping ATPase in SM from broad bean root nodule. It belongs to a IIB-type Ca²⁺-ATPases and can translocate Ca²⁺ by coupling to ATP hydrolysis (Krylova et al., 2013). Secondary active carriers, which can be uniporters, antiporters or symporters, transport substrates across membranes by utilizing energy forms other than ATP – i.e., they use an electric potential and/or chemical gradient of another compound to move the substance against its concentration gradient.

Membrane transporters are systematically classified. To better understand membrane transporter evolution and classification, the International Union of Biochemistry and Molecular Biology (IUBMB) formally modified the transporter Classification (TC) system and created a freely accessible reference database called the Transporter Classification (TCDB) database (http://www.tcdb.org/) (Saier and Paulsen, 2001; Saier et al., 2006; Saier et al., 2009; Saier et al., 2014; Saier et al., 2016). It consists of approximately 10,000 transporters from all kingdoms of life classified into more than 1,000 families, according to sequence similarities, structure, and evolutionary relationships. As our understanding of transporters progresses, and new studies appear, nearly 100 new families are added each year to the database (Saier et al., 2014; Saier et al., 2016). The database also lists structures, functions, mechanisms, as well as evolutionary and disease/medical information. Depending on the family, transporters can be single or multi-
component systems and are categorized in a hierarchical system of classes, subclasses, families, and subfamilies in five levels. The first level is divided into eight classes: channels, secondary carriers, primary active transporters, group translocators, transmembrane electron carriers, auxiliary transport proteins, and putative transporter systems (Saier et al., 2006; Saier et al., 2009; Saier et al., 2014). The second level represents subclasses, with variable numbers for each class. The third level is represented by over 1000 transporter families, following by subfamilies. The last level represents a single transporter with a distinctive range of substrates or energy sources (Saier et al., 2006; Saier et al., 2009; Saier et al., 2014; Saier et al., 2016). The classes and subclasses distinguish functionally distinct types of the transporter, whereas the families and subfamilies provide a phylogenetic basis for classification (Saier et al., 2006).

TCDB is not limited to the classification of well-characterized transport proteins. It also attempts to classify and help elucidate the roles of many transporters that are poorly understood, even when minimal information available in the published literature. TCDB also serves as a genome transporter-annotation tool. The TCDB research group keeps an up-to-date list of software for computational analysis of transport proteins, which allow researchers to search, view, compare, and download relevant information. Thus, this database is an excellent resource for researchers asking fundamental biological questions and determining the routes of the evolution of these proteins, identifying superfamily relationships, and allowing structural, functional and mechanistic predictions.

Membrane transporters contain multiple hydrophobic membrane-spanning domains (α-helical as well as β-barrel segments) formed by hydrophobic domains interacting with the phospholipid bilayer of biological membranes. These interactions create a hydrophilic pore in the membrane, which enables the selective transport of solutes from one side of the membrane to the other based on solute properties (Benedito et al., 2010). Several software packages have been developed to predict transmembrane domains (TMD) and 3D structures on the basis of hydrophobicity scales: HMMTOP (www.enzim.hu/hmmtop), Predictprotein (www.predictprotein.org), SOSUI (harrier.nagahama-i-bio.ac.jp/sosui), TMHMM (www.cbs.dtu.dk/services/TMHMM-2.0/).

Early work on biochemical assays with isolated symbiosomes provided the foundation on which molecular and genomic research has expanded. Along with the complete genome sequence and the advance of bioinformatic approaches, more and
more SM transport proteins have been identified at the molecular level. Three soybean and ten *M. truncatula* transcriptome studies have been published with thorough gene expression profiles on nodules, including nodule development stage or under different treatments (Lee et al., 2004; Benedito et al., 2008; He et al., 2009; De Michele et al., 2009; Maunoury et al., 2010; Severin et al., 2010; Moreau et al., 2011; Gaude et al., 2012; Cabeza et al., 2013; Limpens et al., 2013; Cabeza et al., 2014; Roux et al., 2014; Kunert et al., 2016). Comparing these transcriptome databases facilitates the identification of the nodule-specific gene. To date, although much nodule-specific or highly expressed potential membrane transporters have been identified in the nodule at the genetic level, little is still known about their biochemical and physiological roles in nodule initiation and development in legumes (Benedito et al., 2010). Our research expands this knowledge by exploring the activity of nodule-specific transporters in the model legume, *M. truncatula*.

**Tools for studying rhizobia-legume symbiosis**

*M. truncatula*, a model legume species

Legumes (Fabaceae) are the second major crop family (Young and Bharti, 2012). This family includes many economically important species, such as soybean (*G. max*), beans (*Phaseolus* spp. and *Vigna* spp.), alfalfa (*M. sativa*), peas (*Pisum sativum*), clover (*Trifolium*), and peanuts (*Arachis hypogaea*). These species represent important components of the human daily diet and provide nutrients lacking in cereal and vegetables (Bouchenak and Lamri-Senhadji, 2013). Three major species are regarded as model legume species: the major tropical species with determinate nodule development (*Glycine max*), the temperate legume with determinate nodules (*Lotus japonicus*), and the temperate legume with indeterminate nodule development (*M. truncatula*).

*M. truncatula*, a Mediterranean annual forage has some desirable features, which make it an excellent model species for studying various aspects of legume biology. The relatively small diploid genome (390 Mb), autogamous pollination system, a relatively short life cycle (around three months) and ease of transformation make this close relative of alfalfa a good model system (Frugoli, 2001; Lozano-Baena et al., 2007; Colditz and Braun, 2010; Young et al., 2011; Tang et al., 2014). *M. truncatula* belongs to the temperate Galegoid tribe of the legume subfamily Papilionoideae. It was reported that 54% of the
soybean BACs exhibited some degree of synteny with *M. truncatula*, of which 70% exhibited extensive gene order conservation (Young and Udvardi, 2009; Young et al., 2011; Young and Bharti, 2012; Tang et al., 2014). *M. truncatula* is served by a large germplasm collection of ecotypes ([www.genome.ou.edu/medicago.html](http://www.genome.ou.edu/medicago.html)) as well as a large-scale insertion mutagenesis collection maintained by the Noble Foundation (Wang et al., 2014). The completion of genome sequence, a comprehensive HapMap analysis ([www.medicagohapmap.org](http://www.medicagohapmap.org)) and a large insertional mutant collection ([medicago-mutant.noble.org/mutant](http://medicago-mutant.noble.org/mutant)) (Tadege et al., 2005; Tadege et al., 2008; Cheng et al., 2011; Pislariu et al., 2012) have provided an enormous boost for the *M. truncatula* research as a whole. The strictly autogamous behavior of *M. truncatula* is reflected in high genetic homogeneity with in individual ecotypes. Due to the optimal combination of useful features, a few ecotypes have been selected for routine applications in the laboratory. Most of the studies have been based on two ecotypes, Jemalong A17 and R108. Being chronologically a primary ecotype, Jemalong (A17) was chosen for sequencing by the Medicago Genome Sequencing Consortium. R108 is regarded as superior for whole plant transformation than A17. *M. truncatula* can establish a symbiotic relationship with *R. meliloti* within nodules under conditions of nitrogen limitation in the soil, and with mycorrhizal fungi under low P conditions (Gallusci et al., 1991).

**Completion of the genome-sequencing project**

The diploid (2n=16) genome of *M. truncatula* is organized into eight small chromosomes. With 412 Mb, it is roughly four times larger than *A. thaliana* genome but much smaller than soybean genome (1.15 Gb). The sequencing and annotation of the *M. truncatula* (A17) genome were initiated by the IMGAG (International Genome Annotation Group) in 2001 and published in 2011 (Young et al., 2011). It combined NextGen 454 sequences with BAC-based assemblies and optical mapping to produce the current Mt4.0v2 version (Tang et al., 2014). Mt4.0 pseudomolecules now encompass approximately 384.5 Mb, with 330 Mb align accurately with the optical map, covering nearly all the euchromatin and capturing ~86% of all genes (Young et al., 2011; Tang et al., 2014). *M. truncatula* is estimated to contain 16.8 genes per 100 kb (gene density) throughout the euchromatin, with the average gene being 2,211 bp in length and containing four introns. The counts of high confidence genes are 50,894 while *G. max* also has similar gene counts (46,430).
However, the length of the *G. max* euchromatin is about 400 Mb. Apparently, *M. truncatula* has a higher gene density than *G. max*, which means *M. truncatula* has better space efficiency and has more genes locally duplicated than *G. max* (Cannon, 2013). The *M. truncatula* genome sequence creates a platform of easily accessible DNA markers (single nucleotide polymorphisms [SNPs], copy number variants [CNV]) and insertions/deletions [INDELS] among legume species, which accelerates physical mapping assessable, gene functional characterization, as well as comparative genomics. It also reveals the scale and scope of those legume-exclusive gene families, one of many biologically important gene families making legumes special (Young and Udvardi, 2009). Altogether, the availability of sequence of *M. truncatula* is facilitating legume studies, and especially advancing the root nodule biology field.

**Reverse genetics approach for analysis of gene functions**

The identification of important genes related to symbiotic nitrogen fixation can also be accomplished by analyzing the transcriptome profile of nodules and studying phenotypes due to loss of gene function. To facilitate these studies, important reverse genetics tools have been developed for *M. truncatula*, such as the Affymetrix gene expression atlas (Benedito, 2008) and other transcriptomics tools, as well as a large mutant collection of *Tnt1* retrotransposon insertions with indexed flanking sequence tag (FST) catalog ([medicago-mutant.noble.org/mutant/](medicago-mutant.noble.org/mutant/)) (Revalska et al., 2011; Pislariu et al., 2012; Cheng et al., 2014).

**Transcriptomics**

Transcriptomics is the study of the transcriptome complete set of RNA transcripts (mRNA to rRNA, tRNA, incRNA, and others) produced under specific circumstances or in a specific cell using high throughput techniques: Microarray platforms or more recently high throughput next generation sequencing (Dunwell et al., 2001; Morozova et al., 2009; Wang et al., 2009). By comparing transcriptomes of various types of cells under specific condition allows researchers identify novel genes that are differentially expressed in distinct cell types or in response to different treatments. This provides both quantitative and qualitative information, which helps gene function determination, speeds the biological discovery, and enhances the molecular breeding efforts.
The Medicago gene expression atlas

Based on the purpose of transcriptomes, the Medicago gene expression Atlas (MtGEA) project, which initially provided transcriptome date for the different organs under normal growth condition and time-series of nodules and seed development by Affymetrix GeneChip (Bout et al., 1990; Benedito et al., 2008; He et al., 2009; Limpens et al., 2013). Transcriptome data has been expanded from plants under various kinds of abiotic and biotic stress, as well as from specific cell and tissue types. A compendium of gene expression profiles developed for *M. truncatula* ([mtgea.noble.org/v3/](http://mtgea.noble.org/v3/)) contains 50,900 probe sets designed from transcripts. This tool provides a platform for analyzing the *M. truncatula* transcriptomics and identifying organ-specific genes. The Medicago Gene Atlas currently hosts normalized data from 739 chips from 274 different experiments with (Benedito et al., 2008; He et al., 2009). The libraries contain not only most major plant tissue but also the developmental stages as well as treatments with microbial pathogens, bacterial and fungal symbionts, and abiotic stressors. Also, detailed time-courses of nodule and different nodule zones were profiled to obtain greater detail into the transcriptional programs that underline the development of nodule (Benedito et al., 2008; Limpens et al., 2013). Expression of 86% of all annotated genes was detected in at least one organ. Furthermore, the Affymetrix dataset shows differences between plant organs resulting mainly from quantitative, rather than qualitative, changes in global gene expression. There about 26,000 genes are expressed during nodule development, with 30.2% of these increased or decreased more than 2-fold compared with roots. Approximately 42% (23,880) of all expressed gene was detected in all organs. A relatively small number of genes are organ-specific; and subsets of legume-specific genes are expressed exclusively or preferentially in nodules (Benedito et al., 2008; He et al., 2009). Only 3% of all expressed genes are exclusively expressed in nodules, and 6% of genes are co-expressed only in roots and nodules (Benedito et al., 2008). There are 3,830 genes identified to encode membrane transporters, with 2,673 of those classified into 113 families and 146 subfamilies. Among these, 129 showed organ-specific expression manner, including 50 that are nodule specific; 196 transporters showed at least 5-fold up-regulation during nodule development (Benedito et al., 2010).

RNA-Seq data for nodule dissection
Although the advent of microarray and next generation sequencing technology and availability of genome sequences of several model organisms have facilitated the analysis of global gene expression studies, the challenge of tissue specific expression studies remained due to variation in the complexity of both the genome and concerned tissue. Organs and tissues are made up mostly of a heterogeneous mix of spatially distributed cell types which differ not only in term of function and specialization but also in size, biochemistry, and morphology (Gaude et al., 2012; Datta et al., 2015; de Almeida and Strömvik, 2016; Anjam et al., 2016). In response to physiological cues, responses of each cell type in a given tissue may be unique. Therefore, isolation of specific tissue or cell types without significant contamination with unwanted cells or triggering artefactual gene expression during sample preparation is an essential step to understanding specific developmental processes. Laser capture microdissection (LCM) is a method that allows for the isolation of homogenous cell populations from their native niches in the tissue. It was first described in the early 20th century, and it has been used in medical science for accurate separation of tumor, stromal, and normal cells within a single biopsy specimen (Datta et al., 2015). This approach combined with subsequence RNA-seq provided spatial information on simultaneous gene expression for both plant and bacteria and allowed for the identification of new candidate genes involved in nodule development (Roux et al., 2014; Podgorny and Lazarev, 2017). LCM has been used for the isolation of specific nodule zones followed by RNA-sequencing to obtain a dynamic and integrated view of gene expression in different zones of the nodule (Roux et al., 2014). They examined gene expression associated with three developmental milestones of nodule (meristem activity, cell differentiation and selected signaling processes between bacterial Nod factors and redox status) (Roux et al., 2014). Although a meristem of nodule is a small region containing a few layers of cells, it is the zone where transcriptional and post-transcriptional regulators associated with different developmental transitions strongly induced in nodules. Nevertheless, whole-organ or hand-section of transcription analyses cannot pinpoint the regulator in the meristem, nor the comprehensive transcriptome of both partners in different zones of nodule (Roux et al., 2014). With highly sensitive and spatial information in hand, we can advance to identify plant genes specific to nodule development and nitrogen fixation and to understand their functions.

**Tnt1 insertional mutant collection and database**
Insertional mutagenesis provides a rapid way of linking gene function to phenotypes. It has great advantages for both forward and reverse genetics. The transposable element of *Nicotiana tabacum* cell type 1 (*Tnt1*), a retrotransposon that replicates via an RNA intermediate, was originally isolated from tobacco (Lucas et al., 1995; Cui et al., 2013) and one of the very few active plant long terminal repeat (LTR) retrotransposons. It has been already successfully used in several heterologous hosts, including *A. thaliana*, rice, potato, soybean, lettuce, as well as *M. truncatula* (Lucas et al., 1995; Tadege et al., 2005; Pérez-Hormaeche et al., 2008; Pisliaru et al., 2012; Cui et al., 2013; Duangpan et al., 2013). Collectively, these studies demonstrate that *Tnt1* preferentially transpose into gene-rich regions, making them highly mutagenic. Also, *Tnt1* appears to be inactive in normal seed propagation but can be reactivated by tissue culture (Cheng et al., 2011; Pisliaru et al., 2012), which means that relatively few primary transgenic lines can propagate to large populations of mutants through tissue culture. Thus, the Samuel Roberts Noble Foundation and European collaborators have developed an extensive *Tnt1* mutant collection for *M. truncatula* ([medicago-mutant.noble.org/mutant/](http://medicago-mutant.noble.org/mutant/)) with a parental starter line, Tnk88-7-7, having about five copies of the *Tnt1* retroelement inserted into the wild-type ‘R108’ by *Agrobacterium tumefaciens*-mediated transformation. Nowadays, nearly 22,000 *Tnt1* insertion lines have been posted on the Noble Foundation server, with an estimated 520,000 random insertions within *M. truncatula* genome covering almost 90% of all genes (Pisliaru et al., 2012; Sun et al., 2019). To make utilization of these *Tnt1* mutants, there are two methods successfully developed for identifying insertion mutants in *M. truncatula*: PCR-based DNA pool screening and direct database BLAST search of insertion flanks. Currently, 44,238 flanking sequence tags (FSTs) from 3,436 mutant lines have been generated and are BLAST-searchable in this database, making it easy to recover the flanking sequences and identify insertion position in genes. In addition to FSTs, images and phenotype descriptions for most of those mutant lines are included in the database after phenotypic screening (Cheng et al., 2014). Blast searches of the whole database against a gene of interest can easily identify insertion lines of interest and place seeds orders of *Tnt1* mutant lines. Because *Tnt1* mutants are stable during regular breeding, so they are independent and can be segregated by genetic crossing in *M. truncatula* mutant lines (Cui et al., 2013; Cheng et al., 2014). However, the *Tnt1* mutant collection also has several limitations. *Tnt1* elements are integrated randomly in the genome. Before being able to analyze the phenotype of mutants, we need to backcross for a few generation and then self a plant with your insertion of interest to obtain a
homozygous plant with as few spurious insertions as possible. It is a time-consuming and labor-intensive process. Sometimes, multiple Tnt1 insertions can be genetically linked, making it difficult to segregate these insertions. Furthermore, Tnt1 might modify the expression of nearby genes, which might cause false phenotype. Therefore, a single transposon insertion of a Tnt1 mutant of the gene of interest could limit the ability to obtain informative phenotypes due to the presence of multi-gene families (Lucas et al., 1995; Tadege et al., 2005; Cui et al., 2013; Duangpan et al., 2013).

Conclusion

The genome sequencing efforts of legume species in recent years opened new avenues to explore gene functions relevant to further our understanding of all aspects (genetics, biochemistry, development, and physiology) of symbiotic nitrogen fixation in legumes. Additional databases along with genetics and molecular tools dedicated to legume models, especially in *M. truncatula*, allows us to fast track the essential functions of gene products dedicated to legume nodules. It does make sense to start by conducting studies on nodule-specific genes, and more so on membrane transporters that may be involved with molecular exchanges between the symbionts.

This dissertation is an attempt towards this goal. Herein, I describe my efforts to functionally characterize two nodule-specific secondary membrane transporters in *M. truncatula*. My hypotheses are drawn from their nodule-specific pattern of expression as proxy for possible essential functions dedicated to SNF, as described in the following chapters.

References


Adnane B, Mainassara ZA, Mohamed F, Mohamed L, Jean-Jacques D, T. Rim M,


Brown MH, Paulsen IT, Skurray RA (1999) The multidrug efflux protein NorM is a


exchange and nutrition of pathogens. Nature **468**: 527–32


Diener AC (2001) Arabidopsis ALF5, a Multidrug Efflux Transporter Gene Family Member, Confers Resistance to Toxins. PLANT CELL ONLINE **13**: 1625–1638


Molecular Evolution of Slow and Quick Anion Channels (SLACs and QUACs/ALMTs). Front Plant Sci 3: 263


of the Ca2+ pumping ATPase in the peribacteroid membrane from broad bean root nodules. Protoplasma 250: 531–538


Roelfsema MRG, Hedrich R, Geiger D (2012) Anion channels: Master switches of


transport activities in planta. Plant Signal Behav 8: e22813


Streeter JG (1985) Nitrate Inhibition of Legume Nodule Growth and Activity : II. Short
Term Studies with High Nitrate Supply. Plant Physiol 77: 325–328


Sun SJ, Qi GN, Gao QF, Wang HQ, Yao FY, Hussain J, Wang YF (2016) Protein kinase OsSAPK8 functions as an essential activator of S-type anion channel OsSLAC1, which is nitrate-selective in rice. Planta 243: 489–500


(version Mt4.0) for the model legume Medicago truncatula. BMC Genomics 15: 312


Zhao J, Dixon RA (2009) MATE transporters facilitate vacuolar uptake of epicatechin 3′-O-glucoside for proanthocyanidin biosynthesis in Medicago truncatula and

Zhao M, Chen Y, Qu D, Qu H (2011b) TSdb: a database of transporter substrates linking metabolic pathways and transporter systems on a genome scale via their shared substrates. Sci China Life Sci 54: 60–4


Chapter 2

Genomic analysis of Multidrug and Toxic Compound Extrusion (MATE) transporters in the model legume *Medicago truncatula*

Summary

Multidrug and Toxic Compound Extrusion (MATE) transporters comprise a family of efflux proteins with affinity to chemically diverse substrates. Although in most cases, their physiological functions are not fully understood, they are widely distributed in all kingdoms of living organisms. Plants present a strikingly higher diversity of MATE transporters than bacteria and animals. Recent studies showed that plant MATEs are involved in efflux as well as transport between organelles, cells, and organs of the plant of a broad range of substrates, such as organic acids, plant hormones, and secondary metabolites. So far, there has been no genome-wide analysis of the MATE family in the model legume *Medicago truncatula*. In this species, The MATE family is composed of 70 members, which are unevenly distributed across its eight chromosomes with both tandem and segmental duplications detected. *M. truncatula* MATE genes can be classified into four subfamilies comprised 14 small subgroups with diverse physiological functions, such as transport and accumulation of flavonoids and alkaloids, extrusion of plant-derived or xenobiotic compounds, regulation of disease resistance, and response to abiotic stresses. Most genes showed tissue-specific expression patterns. Altogether, our phylogenetic and transcriptomic analyses of the *M. truncatula* genome provide a framework for genome-wide association mapping targeting phenotypes of economic and environmental importance in legumes.

**Keywords**: Gene Duplication, Phylogeny, Synteny, Transcriptional Regulation.
Introduction

The Multidrug and Toxic Compound Extrusion (MATE) family of secondary membrane transporters is comprised of effluxers that couple the electrochemical gradient of a cation (H⁺ or Na⁺) as the force used to translocate substrates across a biological membrane. It is the most recently discovered family of multidrug resistance transporter proteins (Brown et al., 1999; Eckardt, 2001; Saier and Paulsen, 2001; Kuroda and Tsuchiya, 2009; Magalhaes, 2010). The 12-transmembrane (TMD) topology of NorM from Vibrio cholera, the first discovered MATE member, revealed that MATE transporters are distinct from the other four known multidrug resistance transporter families: Major Facilitator Superfamily (MFS), Small Multidrug Resistance (SMR) family, Resistance/Nodulation/Cell Division (RND) family, and ATP-Binding Cassette (ABC) superfamily (Brown et al., 1999; Saier and Paulsen, 2001; Moriyama et al., 2008).

Although their function is not fully understood, MATE transporters occur ubiquitously in all three primary kingdoms of life (Archaea, Bacteria, and Eukarya). Based on amino acid sequence similarity, about 900 MATE transporters have already been identified in species of all three kingdoms and classified into three large subfamilies (bacterial, eukaryotic, and bacterial/archaeal) comprising 14 smaller subgroups (Omote et al., 2006). All plant MATE transporters are grouped into a plant-specific clade. In contrast to the relatively small number of MATE transporters per genome in other three eukaryotic clades, the plant-specific clade presents the most abundant number of orthologs per species. So far, 56, 46, 67, and 117 putative MATE transporters have been identified in Arabidopsis thaliana, Oryza sativa (rice), Solanum lycopersicum (tomato), and Glycine max (soybean), respectively (Santos et al., 2017; Li et al., 2002; Liu et al., 2016), while there are only two genes encoding MATE transporters (hMATE1 and hMATE2) in humans (Matsumoto et al., 2008).

MATE transporters are usually comprised of 400-650 amino acid residue proteins with 12 predicted transmembrane domains (TMDs) (Brown et al., 1999; Omote et al., 2006). The X-ray structures of Na⁺-driven NorM transporters from Vibrio cholera (Nor-VC) and Neisseria gonorrhoea (Nor-NG), as well as H⁺-driven MATE transporters from Pyrococcus faiosus (PfMATE) and Bacillus halodurans (DinF-BH), showed an outward-facing conformation with two bundles of six transmembrane helices (TMDs 1-6 and 7-12) to form a large internal cavity open to the extracellular space (He et al., 2010; Lu et al.,
All published reports describing MATE transporters in their extracellular-open state show they release substrates to the periplasm. The relatively higher degree of conservation in several regions, such as in the vicinity of TMDs as well as the pseudo-twofold symmetry along an axis modules of these transporters, suggest that MATE proteins may have evolved from a common ancestral gene that encoded 6 TMDs and then underwent a duplication event (Omote et al., 2006; Moriyama et al., 2008; He et al., 2010; Lu, 2016).

From the energetics perspective, MATE transporters export a wide range of substrates using an electrochemical gradient of H⁺ or Na⁺ across the cell membrane. The structural features and transport mechanisms of bacterial MATE transporters have been investigated in detail. However, eukaryotic MATEs have only been explored in the last ten years. Most of the eukaryotic MATE transporters characterized so far are antiporters localized to the plasma membrane and use the H⁺ gradient (proton motive force, pmf) to transport solutes out of the cell (Moriyama et al., 2008). On the other hand, many bacterial MATE proteins, including NorM from Neisseria gonorrhoeae, use instead the Na⁺ gradient across the membrane to catalyze the transport (Morita et al., 1998; Kuroda and Tsuchiya, 2009). Although NorM-NG and DinF-BH share similar protein conformation, the ligand-sharing arrangement of cation- and/or substrate-binding sites is quite distinct as are the interactions between substrate and cations. NorM-NG coordinates Na⁺ through a cation-π interaction to trigger protein conformational change from outward- to inward-facing for more favorable to substrate binding, whereas H⁺ and certain drugs directly compete with the DinF-BH binding sites due to differences in amino acids interacting with Na⁺ (Lu et al., 2013b; Lu et al., 2013a; Lu, 2016). Since prokaryotic and eukaryotic MATE transporters share relatively low levels of amino acid similarity, the transport mechanisms used by eukaryotic MATE transporters most likely differ from those of known prokaryote structures.

Current research shows that most bacterial MATEs export organic cations for multidrug resistance, while eukaryotic MATEs serve other physiological functions (Moriyama et al., 2008). In yeast (Saccharomyces cerevisiae), the Erc1 confers resistance to ethionine, a potent carcinogenic compound analog to methionine (Shiomi et al., 1990). Mammalian MATE transporters excrete metabolic waste and xenobiotic organic cations in the kidney and liver (Omote et al., 2006). On the other hand, plant MATE transporters are involved in diverse physiological and metabolic processes of plant growth and development (Burko et al., 2011), including: nutrient homeostasis (Liu et al., 2009);
hormone transport, such as abscisic acid (ABA), salicylic acid (SA) and auxin (Serrano et al., 2013; Yamasaki et al., 2013; Zhang et al., 2014); primary and secondary metabolisms (Zhao and Dixon, 2009; Zhao et al., 2011a); and responses to biotic and abiotic stresses (Yamasaki et al., 2013). Although still limited, the recent progress made on functional characterization of several MATE transporters demonstrates their essentiality to plant physiology and metabolism.

Barrel medic (*Medicago truncatula*) is a model legume with high-quality genomic data as well as molecular and genetic resources available. It is an ideal tool for studying important processes, which are not present in *Arabidopsis thaliana*, such as symbiotic root interactions (mycorrhization and rhizobial nitrogen fixation), development of a complex leaf pattern, and a rich secondary metabolism of (iso)flavonoids (Zhao and Dixon, 2009; Young et al., 2011). With the publicly available genome sequence (Young et al., 2011; Tang et al., 2014) and gene expression data (Benedito et al., 2008), it is possible to identify all MATE genes in the *M. truncatula* genome and investigate their expression patterns and possible functions. Compared to *Arabidopsis thaliana*, little work has been done on MATE transporters in *M. truncatula*. To date, only five MATE transporters were reported to play important physiological roles in a wide range of biology process in *M. truncatula*. MtMATE1 and MtMATE2 are involved in vacuolar sequestration of anthocyanins and proanthocyanidins in the seed coat and leaf, respectively (Zhao and Dixon, 2009; Zhao et al., 2011a). MtMATE55, MtMATE65, and MtMATE69 have very recently been characterized as citric acid transporters, with MtMATE55 and MtMATE69 being involved in Fe accumulation and translocation, and MtMATE66 implicated in A$^{3+}$ stress response (Wang et al., 2017). Phylogenetic analysis can help predict the physiological functions of MATE transporters closely related to those already functionally characterized. Therefore, detailed analyses of MtMATE sequences will help us explore their functions.

In this chapter, we performed a genome-wide search of all putative MATE transporters in the *M. truncatula* genome and assessed their expression profiles by using publicly available gene expression data. We further explored their chromosomal distribution, gene architecture (i.e., intron-exon structure), and analyzed gene duplication events, synteny, and phylogenetic relations. This study provides new insights into understanding MATE transporters in plants and offers useful information for further exploring the functional roles that MATE transporters play in integrating legume physiology and cell metabolism.
Results and Discussion

Genome-wide identification of MATE transporters in *M. truncatula*

There were a total of 70 genes encoding MATE transporters identified in the *M. truncatula* genome by using the TransportTP tool (http://bioinfo3.noble.org/transporter: Li et al., 2009; Benedito et al., 2010). Additionally, four MATE genes [Medtr3g087820 (122 aa), Medtr4g131290 (74 aa), Medtr7g007430 (185 aa), and Medtr8g037130 (68 aa)] coded for too short peptides, possibly encompassing pseudogenes. Twenty of the full-length MATE genes potentially transcribe alternative splicing forms, although only their full forms (with all exons) were considered in this study. These putative 65 MATE members plus 5 experimental proven MATE proteins consist of 319 to 620 amino acid residues, most containing 11-12 TM domains. This result is similar to that found in *O. sativa* and *A. thaliana*, in which protein lengths found for members of the MATE family ranged from 392 to 644 amino acids and 469 to 575 amino acids, respectively, most encompassing 12 TMDs (Li et al., 2002; Zhang et al., 2014; Liu et al., 2016; Wang et al., 2016). The predicted molecular weights of *M. truncatula* MATE proteins ranges from 34 to 68 kDa, and the predicted isoelectric point (pI) values varied broadly between 4.75 and 9.95, while in rice the molecular weights ranged from 41.3 to 65.8 kD with pI values between 5.14 and 10.07, and in *A. thaliana* they varied between 50.8-63.5 kD and pI's of 4.66-8.67 (Liu et al., 2016; Wang et al., 2016). Overall, the discrepancies observed in peptide length and physicochemical properties of MATE transporters in different plant species may have allowed them to assume different functions over the course of evolution.

Phylogenetic analyses of the *M. truncatula* MATE family

To further explore potential homologies, a comprehensive phylogenetic tree containing *M. truncatula*, Arabidopsis, and selected functionally characterized MATE transporters was created to evolutionarily analyze MATE family (Figure 1). MATE transporters were grouped into four distinct clades. As described below, this categorization can be useful in assigning potential metabolic and physiological roles to each *M. truncatula* MATE transporter.
**Clade I: MATE transporters regulating plant organ initiation, growth, and development**

Clade I encompasses three subclades. Subclade I-A is an orphan branch with only a single, poorly functionally characterized bacterial transporter (TsMDR). Subclade I-B contains MtMATE55 as well as nine other *M. truncatula* proteins and seven Arabidopsis MATEs (AtDTX48-52, AtDTX54-55). The functionally characterized Arabidopsis MATEs in this clade are all involved in regulating plant development and growth. MtMATE55 is expressed in flowers, leaves, vegetative buds, and the hypocotyl and is also involved in Fe homeostasis (Wang et al., 2017). MtMATE55 is most closely related to the Arabidopsis DTX48 (BCD1/ZRIZ1/ABS4/ZF14) (**Figure 1**). DTX48 is localized to Golgi bodies and may also have a role in Fe nutrition and organ initiation and development (Burko et al., 2011; Seo et al., 2012). Alternatively, DTX48 was also reported to localize to the late prevacuolar compartment, and its overexpression led to inhibition of hypocotyl cell elongation under light (Wang et al., 2015). AtDTX50 (ABSL2) is another closely related ortholog of MtMATE55. AtDTX50 is a plasma-membrane ABA efflux transporter expressed mostly in guard cells and vascular tissues (Zhang et al., 2014; Wang et al., 2015).

The ALTERED DEVELOPMENT PROGRAM 1 (ADP1/ADS1/DTX51) transporter plays an essential role in regulating lateral organ outgrowth and in maintaining normal architecture in Arabidopsis, likely through fine-tuning local auxin biosynthesis (Li et al., 2015a). More recently, the trans-Golgi membrane localized BIG EMBRYO 1 (ZmBIGE1) from maize was shown to complement adp1 Arabidopsis mutants (Suzuki et al., 2015). Additionally, the Arabidopsis ABNORMAL SHOOT 3-LIKE 1 (ABS3L1/DTX50) and ABS3L2/DTX52, which are localized to the late endosome and prevacuole compartment, were implicated in repressing the elongation of hypocotyl cells (Wang et al., 2015). The small I-C subclade encloses the Arabidopsis RESISTANT TO HIGH CO\textsubscript{2} 1 (RHC1/DTX56) transporter, which localizes to the plasma membrane and participates in responses to elevated CO\textsubscript{2} and stomatal closing through a signaling cascade involving the protein kinase HT1 (HIGH LEAF TEMPERATURE 1) in CO\textsubscript{2}-induced activation of the anion channel SLAC1 (Tian et al., 2015). Furthermore, Medtr5g011540, a close Medicago MATE, with strong expression in all organs but little transcriptional activity in fruit, was identified as homologous to RHC1/DTX56 in this subclade (**Figure 1**). Therefore, the
characterized MATE transporters in this clade directly or indirectly participate in nutrition, plant hormone or signaling molecule translocation from synthetic to action sites, thereby regulating plant development and growth.

Clade II: MATE transporters that secrete organic acids to regulate plant nutrient translocation and metal acquisition

Clade II is diverse and can be divided into three subclades. Subclade II-B encompasses only prokaryotic transporters while II-C holds transporters from fungal and protozoan species as well as animal species. Animal MATE genes are expressed mostly in kidneys and testis to export drugs and waste (Yonezawa and Inui, 2011). Subclade II-A is plant-specific and contains the well-characterized plasma membrane citric acid exporters. It can be further subdivided into three branches. The first branch of subclade II-A contains citric acid exporters that secrete organic acids to the apoplast to chelate Al$^{3+}$ in the rhizosphere and are thus involved in alleviating Al$^{3+}$ toxicity in acidic soils (Magalhaes et al., 2007; Magalhaes, 2010; Zhou et al., 2014; Kochian et al., 2015). They include OsFRDL1 from rice (Green and Rogers, 2004; Durrett et al., 2007; Yokosho et al., 2009), ScFRDL1 from rye (Rogers and Guerinot, 2002), and SbMATE from sorghum (Sivaguru et al., 2013). MtMATE66 appears in this branch next to BoMATE, a citrate exporter from Brassica oleracea implicated in Al$^{3+}$ tolerance (Wu et al., 2014). Furthermore, other members of this branch, such as HvAACT1/HvMATE1 from barley, SbMATE1 from sorghum, ZmMATE1 from maize, and OsFRDL4 from rice were characterized as citric acid exporters and also proven to be key elements of Al$^{3+}$ tolerance (Furukawa et al., 2007; Magalhaes et al., 2007; Liu et al., 2009; Cheng et al., 2011). Globally, Al toxicity is one of the most significant obstacles to crop yields in acidic soils, which comprise around 50% of the world’s arable land (Zhao et al., 2011b). The Al-activated efflux of organic acids from roots into the rhizosphere involves MATE transporters, which are now recognized as a major component of Al tolerance mechanism in plants (Kochian et al., 2015). Overexpression of these MATE transporters confers Al resistance in many crops (Liu et al., 2009; Yokosho et al., 2009; Maron et al., 2010; Yang et al., 2011; Takanashi et al., 2013). Overexpression of MtMATE66 resulted in enhanced tolerance to Al stress in M. truncatula hairy roots, indicating its potential utility in crop breeding for cultivating legumes in acidic soils (Wang et al., 2017).
The second branch of this subclade encompasses transporters involved in Fe homeostasis by secreting citric acid into the vascular system to form a transportable citric acid-ferrous complex that facilitates Fe$^{2+}$ or Zn$^{2+}$ translocation from root to shoot. This branch includes the FERRIC REDUCTASE DEFECTIVE 3 (FRD3/DTX43) from Arabidopsis, GmFRD3a and GmFRD3b from soybean, and MtMATE55. A recent QTL study indicated that FRD3 is essential for Zn$^{2+}$ tolerance in Arabidopsis by regulating Fe homeostasis (Pineau et al., 2012). In addition, LjMATE1, a MATE citrate transporter from the model legume *Lotus japonicus*, assists in Fe translocation to nodule tissues (Takanashi et al., 2013). The third branch contains the Arabidopsis ENHANCED DISEASE SUSCEPTIBILITY 5 (EDS5/DTX47), which is localized to the chloroplast envelope and transports salicylic acid (SA) out of the organelle in epidermal cells (Serrano et al., 2013; Yamasaki et al., 2013). Chloroplasts are the place where SA is synthesized for triggering systemic acquired resistance, thus supporting the EDS5 role on disease tolerance mechanisms in Arabidopsis (Serrano et al., 2013; Yamasaki et al., 2013).

**Clade III: MATE transporters that mediate vacuolar sequestration of secondary metabolites and involved in plant defense**

Clade III can be further split into four subclades. Subclade III-A contains two species-specific branches clearly separating six Arabidopsis (DTX20-25) from six Medicago transporters. Interestingly, four Medicago in-tandem duplicated sequences are clustered together (Medtr7g087370/400/410/430) while the other two in-tandem sequences are located elsewhere in the genome (Medtr6g027190/200). Four of the Arabidopsis sequences are also clustered in-tandem in this subclade (At1g33080/090/100/110) (**Figure 1** and **Table 2**). We further explored whether this gene cluster was already present, at least partially, in the last common ancestor between these two species. Indeed, synteny analysis revealed they fall within the same syntenic group (**Table 2**). Genetic studies of these transporters should take into account possible functional redundancies, especially regarding the tissues where their expression overlaps.

Subclade III-B holds four Medicago and three Arabidopsis transporters (DTX26-28), all yet to be functionally characterized. Subclade III-C comprises several characterized anthocyanin/flavonoid/alkaloid transporters, including the Arabidopsis FLOWER FLAVONOID TRANSPORTER (FFT/DTX35); the grapevine VvAM1, VvAM3,
VvMATE1 and VvMATE2; the tobacco NtJAT2; the uncharacterized ROOT HAIR-SPECIFIC 2 (RHS2/DTX31) from Arabidopsis (Gomez et al., 2009; Won et al., 2009; Thompson et al., 2010; Shitan et al., 2014); and, finally, the M. truncatula MtMATE2, which mediates the vacuolar sequestration of anthocyanins and other flavonoid glycosides in leaf and flower cells (Zhao et al., 2011a). Subclade III-D also contains well-characterized vacuolar transporters of phenolics and alkaloids, such as the nicotine transporters NtMATE1 and NtMATE2 from tobacco (Shoji et al., 2009), the proanthocyanidin transporter AtTT12 (DTX41) from Arabidopsis (Marinova et al., 2007), as well as the epicatechin 3'-O-glucoside transporter MtMATE1 from Medicago (Zhao and Dixon, 2009). While ten Medicago transporters in this subclade have been functionally characterized, nine still await to be studied.

**Clade IV: MATE transporters that play essential roles in cell detoxification**

Clade IV contains mostly uncharacterized MATE transporters and can be further classified into four subclades. Two DETOXIFICATION 1 (DTX1) transporters from Arabidopsis and the nicotine transporter NtJAT1 from tobacco are assigned to subclade IV-A. AtDTX1 is a plasma membrane transporter capable of exporting toxic xenobiotics, such as tetramethylammonium chloride and *ethidium bromide, to the extracellular space* (Li et al., 2002). Among the six Medicago transporters in subclade IV-A, we can identify in-tandem paralogs, such as Medtr1g108840/Medtr1g109060/Medtr1g108990 as well as Medtr6g004220/Medtr7g082800/810. This branch contains Medr7g082810 (MtMATE30), which is functionally characterized in Chapter 3 of this dissertation. Subclade IV-B only contains uncharacterized transporters, including nine duplicated *M. truncatula* homologs arranged in tandem on chromosome 6 (*Table 1*).

Subclade IV-C encompasses ABERRANT LATERAL ROOT FORMATION 5 (ALF5/DTX19). ALF5 is expressed in the root epidermis and is required for protection against toxic compounds, such as polyvinylpyrrolidone and pyrrolidone (Diener, 2001). The closely related DTX18 was characterized as responsible for secretion of the defensive compound, coumaroylagmatine, in response to *Phytophthora infestans* colonization (Dobritzsch et al., 2016). Six Medicago ALF5 orthologs are tandemly arranged on chromosome 2 (*Table 1*), possibly as a result of successive gene duplications. However, their expression patterns are varied, with Medtr2g079030 showing constitutive and overall...
highest expression of all MATE genes in spite of its weak expression in nodules (Figure 1). At last, subclade IV-D encompasses a single transporter, Medtr7g007430. It is intriguing that despite the evolutionary distance and lack of close homologs in other plant species, this gene is constitutively and highly expressed throughout the mature plant (Figure 1).

Phylogeny of MATE transporters coupled with gene expression, subcellular localization and mutant analyses are useful to broadly predict potential physiological roles

Overall, the widely diverse tissue-expression patterns shown by the MATE genes present in the *M. truncatula* genome indicate that they might be involved in very different physiological functions, even, for example, when two groups of plasma membrane MATE transporters perform the efflux of citric acid out of the cell. For example, citrate transporters, including *AtFRD3* (*AtDTX43*), *OsFRDL1*, and *MtMATE65* are expressed in vascular tissues and required for root-to-shoot translocation of not only iron but also zinc (Durrett et al., 2007; Yokosho et al., 2009; Pineau et al., 2012), whereas *AtMATE* (*DTX42*), *OsFRDL4*, and *MtMATE66* are expressed in root epidermal cells but instead are essential for Al tolerance (Liu et al., 2009). Some MATE transporters also mediate vacuolar sequestration of phenolic compounds and alkaloids, such as *NtMATE1* and *NtMATE2* as well as *MtMATE1* and 2 (Shoji et al., 2009; Zhao and Dixon, 2009; Zhao et al., 2011a). Some MATE transporters that participate in plant development and growth have been characterized as hormone transporters in various organs or tissues. Specifically, *AtADS1/ADP1/ABS3L* is involved in plant defense by regulating local auxin biosynthesis, thereby affecting plant architecture (Sun et al., 2011; Li et al., 2015a). *AtEDS5* and *AtDTX50* are both salicylic acid (SA) transporters (Nawrath et al., 2002), as is the substrates of MATE transporters with dramatic effects on different processes of plant growth and development, such as *ZmBIGE1*, *AtADS1*, *AtRHC1*, *AtBCD1*, and *MtMATE55* have yet to be identified (Burko et al., 2011, Wang et al., 2015). *AtRHC1* modulates CO$_2$-induced stomatal closing by linking elevated CO$_2$ concentration to repression of a signal transduction cascade (Tian et al., 2015), while *DTX18/BCD1/ZRIZ1/ABS4* is involved in Fe translocation and apical development (Burko et al., 2011; Seo et al., 2012). Lastly, the maize *ZmBIGE1* regulates organ initiation and development (Suzuki et al., 2015) and
ABS3L1/DTX52 and ABS4L1/DTX50 are involved in plant growth (Sun et al., 2011; Li et al., 2015b). Therefore, plant MATE transporters are functionally versatile, and further analyses are needed in model and crop species to better understand this family and the impact on agronomically relevant traits.

**Chromosomal distribution and structural analysis of MATE genes**

The 70 MtMATE genes found in the *M. truncatula* genome are not randomly distributed on its eight chromosomes. Rather, chromosome 6 contains the highest density of MtMATE genes (17), while chromosome 3 has the least number of members (four). To better understand the characteristics of MATE transporters in *M. truncatula*, their gene structures were plotted alongside the phylogenetic tree (Figure 2C). This analysis provides more insight into the evolution of gene structures of tandem and segmental duplications. MATE genes within the same subfamily share similar length and number of exons-introns, with a few exceptions. MATE genes in clade I have one to three exons. Similarly, most MATE members in clade II have 12 to 14 exons, with just one exception (Mt7g070210.1, which has nine exons). In the clade III and IV, most genes have six to nine exons. The conserved gene architecture shown within each MATE subfamily is evidence of their expansion in the plant lineage from a few genes present in the ancestor genome.
Figure 1. Phylogenetic tree of MATE transporters. All *Medicago truncatula* MATE transporters (n=70) were included, along with all *Arabidopsis thaliana* MATEs (n=54) plus all functionally characterized MATEs from other plant species (n=26) as well as non-plant MATEs represented in the TCDB database (n=36). The tree was drawn to scale with branch lengths measured in a number of substitutions per site. *Medicago truncatula* MATEs are in shown bold while the ones functionally characterized in this species are indicated in red. The four distinct clades are indicated in different colors. Subclades are indicated in the circle.
Figure 2. Phylogenetic analysis, expression analysis, and gene architecture of MATE transporter family in the *Medicago truncatula* genome. (A) The phylogenetic tree was generated with protein sequences by Maximum Likelihood (ML) algorithm with 1,000 bootstraps based on JTT matrix-based model in MEGA v.6.0. Functionally characterized MATE genes are in bold. The tree was drawn to scale with branch lengths measured in some substitutions per site. (B) Heatmap with normalized gene expression data of corresponding MATE genes in mature organs of *Medicago truncatula* growing under optimal conditions. A complete lack of expression is indicated by white, yellow indicates below-median values, median expression (n=5 reads) is indicated in orange, and higher counts are shown in gradations of red (maximum n=208 reads). (C) Predicted gene structure including deduced exons (black boxes) and introns (lines) for each of the MATE members. Bar scale: 500bp.
Duplication patterns of MATE genes in the \textit{M. truncatula} genome

Gene duplication analysis of a genome helps us understand the molecular evolution of a species (Schlueter et al., 2004). Recent studies comparing plant genomes showed that there was a shared whole–genome hexaploidy (pre-rosid WGH) occurring before the Rosids-Asterids split approximately 140-150 Mya (Young et al., 2011). The pre-rosid WGH (a.k.a paleopoloidization) event led to the radiation of many gene families in the genome, with the consequent acquisition of substantial gene redundancies in the eudicot plant lineage. Further, the common ancestor of the legumes \textit{M. truncatula} and \textit{G. max} shared an additional round of whole genome duplication (WGD) about 58-59 Mya, which dates to the origins of the papilionoid lineage (early legume-specific duplication) (Young et al., 2011). This key evolutionary event created many genes in crop legumes that exist as paralogous duplicates, and that may have since acquired distinct functions (Cannon, 2013). After the split of the hologelegina (e.g., \textit{M. truncatula} and \textit{Lotus japonicus}) and the milletioids (e.g., \textit{G. max} and \textit{Phaseolus vulgaris}) at about 54 Mya, the \textit{G. max} ancestor was subject to one additional round of WGD at approximately 13 Mya (Glycine-specific WGD), while \textit{M. truncatula} underwent high rates of local gene duplications (Wojciechowski, 2003; Pfeil et al., 2005; Sprent and James, 2007; Young et al., 2011). Thus, there are two critical evolutionary events which impacted the \textit{M. truncatula} genome: the early pre-rosid WGD and the local \textit{Medicago}-specific duplications. Both events are expected to have had great influence on the expansion of MATE gene family in the \textit{M. truncatula} genome.

Analysis of microsynteny (conservation of gene block order within a species genome) can help to detect duplications of chromosomal fragments within a lineage as well as find potential genetic functional redundancies within a genome. The detection of local, in-tandem duplications is also very helpful to find genetic redundancies in a genome. Therefore, In order to further understanding the evolution of the MATE family in Medicago, we further investigated the microsynteny relationships within the \textit{M. truncatula} genome (Figure 3) as well as the shared synteny between the \textit{A. thaliana} and \textit{M. truncatula} genomes (Figure 4). We also compared \textit{M. truncatula} with the \textit{G. max} genome to gain insight into the evolution pattern of taxon divergence within the legume clade.

In our analysis, 70% (49/70) of the MATE genes in \textit{M. truncatula} showed as part of duplication events, either locally (segmental) or as syntenic blocks elsewhere. Likewise,
57% (40/70) are part of in-tandem duplications (Table 2) and are among the 5,811 duplicated regions found in the whole *M. truncatula* genome (Figure 3). Tandem duplications often are a consequence of uneven crossovers and result in homologous genes that can initially become functionally redundant but later may take paths towards sub/neofunctionalization, or pseudogenization, unless the additional expression level leads to better fitness. The in-tandem duplication ratio of the MATE family is higher in *M. truncatula* than *O. sativa* (37% or 9/45), *A. thaliana* (21% or 20/56) and *G. max* (36% or 25/117) (Liu et al., 2016; Wang et al., 2016). This reveals that the *Medicago*-specific gene duplication event was fundamental to establish the current MATE gene family in this species.

We also found four syntenic paralog pairs in the MATE family (Table 2) within the *M. truncatula* genome. Among the 324 microsyntenic regions identified, only 11.4% of the MATE genes were revealed as derived from segmental duplications (Figure 3, Table 2), compared to 15.6%, 17.9% and 60.68% of segmental duplications in *O. sativa*, *A. thaliana*, and *G. max*, respectively. These results show that segmental duplications contributed to the expansion of the MATE family in plants overall, but less so in the Medicago lineage. Thus, in plants, most duplication events in the MATE family occurred through in-tandem duplication, with a smaller contribution of segmental duplications (Kong et al., 2007). However, *M. truncatula* showed the highest ratio of in-tandem duplications amongst the plant species analyzed, while *G. max* had the highest ratio of segmental duplications, even when considering the soybean-specific WGD event. Indeed, it had already been reported that gene redundancy in *M. truncatula* was typically derived from local gene duplications (Schlueter et al., 2004), whereas in *G. max* it was mainly derived from the WGD event after the papilionoid speciation. The MATE family follows this trend (Tables 2-3, Figures 2-3).

The Arabidopsis and Medicago lineages split around 92 Mya, thus after the pre-rosid WGH event (Cannon, 2013). We found eleven pairs of syntenic paralogs of MATE transporter genes between the *M. truncatula* genome and *A. thaliana* (Figure 4 and Table 3). This analysis helps build strong hypotheses on physiological functions of MATE genes in *M. truncatula* for those genes closely related to functionally analyzed MATE transporters in Arabidopsis.
Figure 3. Synteny analysis within the *Medicago truncatula* genome. The analysis was carried out with SynMap software using coding sequences (JCVI cDNA sequence v.4.0) as input data and default parameters. Syntenic regions in the genome are presented as dots. The four MATE pairs located in syntenic regions are highlighted with red dots and shown in Table 1.

Figure 4. Synteny analysis between *Medicago truncatula* and *Arabidopsis thaliana* genome. The analysis was carried out with SynMap software using coding sequences (Mt4.0v1 and TAIR10 genome release versions for *Medicago* and *Arabidopsis*, respectively) as input data, and default parameters. Syntenic regions in the genome are presented as dots. The 11 MATE pairs located in syntenic regions are highlighted with red dots and shown in Table 1.
<table>
<thead>
<tr>
<th>Tandem Duplication Groups</th>
<th>Ioci</th>
<th>Syntenic Paralog Pairs</th>
<th>Ioci</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>Medtr1g108810.1 Medtr1g108840.1 Medtr1g109060.1</td>
<td>Pair 1</td>
<td>Medtr2g088400.1 Medtr4g048000.1</td>
</tr>
<tr>
<td>Group 2</td>
<td>Medtr2g078930.1 Medtr2g078960.1 Medtr2g078970.1 Medtr2g078990.1 Medtr2g079020.1 Medtr2g079030.1</td>
<td>Pair 2</td>
<td>Medtr3g111020.1 Medtr5g067460.1</td>
</tr>
<tr>
<td>Group 3</td>
<td>Medtr3g099700.1 Medtr3g099720.1</td>
<td>Pair 3</td>
<td>Medtr6g081310.1 Medtr8g106670.1</td>
</tr>
<tr>
<td>Group 4</td>
<td>Medtr4g071370.1 Medtr4g071390.1 Medtr4g071420.1 Medtr4g071445.1 Medtr4g071465.1</td>
<td>Pair 4</td>
<td>Medtr2g097900.1 – MtMATE55 Medtr8g037175.1</td>
</tr>
<tr>
<td>Group 5</td>
<td>Medtr5g032720.1 Medtr5g032750.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 6</td>
<td>Medtr5g090280.1 Medtr5g090290.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>Medtr6g027190.1</td>
<td>Medtr6g027200.1</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>----------------</td>
<td>-----------------</td>
<td></td>
</tr>
<tr>
<td>Group8</td>
<td>Medtr6g047590.1</td>
<td>Medtr6g047610.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Medtr6g047640.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group9</td>
<td>Medtr6g081240.1</td>
<td>Medtr6g081300.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Medtr6g081310.1</td>
<td>Medtr6g081390.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Medtr6g081400.1</td>
<td>Medtr6g081410.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Medtr6g081780.1</td>
<td>Medtr6g081810.1</td>
<td></td>
</tr>
<tr>
<td>Group 10</td>
<td>Medtr7g082800.1</td>
<td>Medtr7g082810.1</td>
<td></td>
</tr>
<tr>
<td>Group 11</td>
<td>Medtr7g087370.1</td>
<td>Medtr7g087400.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Medtr7g087410.1</td>
<td>Medtr7g087430.1</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Syntenic paralogs of MATE transporter genes in the *M. truncatula* genome

<table>
<thead>
<tr>
<th>Syntenic Paralog Pairs</th>
<th><em>Medicago truncatula</em> (Tandem duplication group)</th>
<th><em>Arabidopsis thaliana</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Pair 1</td>
<td>Medtr1g101500.1</td>
<td>At5g52050</td>
</tr>
<tr>
<td>Pair 2</td>
<td>Medtr2g022980.1</td>
<td>At4g21903</td>
</tr>
<tr>
<td></td>
<td></td>
<td>At4g21910(^2)</td>
</tr>
<tr>
<td>Pair 3</td>
<td>Medtr2g078960.1</td>
<td>At5g53870</td>
</tr>
<tr>
<td></td>
<td><em>(tandem duplicate group 2)(^{f})</em></td>
<td></td>
</tr>
<tr>
<td>Pair 4</td>
<td>Medtr4g048000.1</td>
<td>At3g21690</td>
</tr>
<tr>
<td>Pair 5</td>
<td>Medtr4g071370.1</td>
<td>At4g39030 - EDS5</td>
</tr>
<tr>
<td></td>
<td><em>(tandem duplicate group 4)(^{f})</em></td>
<td></td>
</tr>
<tr>
<td>Pair 6</td>
<td>Medtr5g010830.1</td>
<td>At4g23030</td>
</tr>
<tr>
<td>Pair 7</td>
<td>Medtr5g011540.1</td>
<td>At4g22790</td>
</tr>
<tr>
<td>Pair 8</td>
<td>Medtr5g099020.1</td>
<td>At4g38380</td>
</tr>
<tr>
<td>Pair 9</td>
<td>Medtr6g027200.1</td>
<td>At3g03620(^3)</td>
</tr>
<tr>
<td></td>
<td><em>(tandem duplicate group 7)(^{f})</em></td>
<td></td>
</tr>
<tr>
<td>Pair 10</td>
<td>Medtr7g087370.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>(tandem duplicate group 11)(^{f})</em></td>
<td></td>
</tr>
<tr>
<td>Pair 11</td>
<td>Medtr8g102430.1</td>
<td>At4g29140 - ADS1</td>
</tr>
</tbody>
</table>

\(^{f}\) Refers to Supporting Table 1 for *M. truncatula* tandem duplication groups.

\(^2\) Tandem duplicates in the *Arabidopsis* genome.

\(^3\) This *Arabidopsis* gene was identified as syntenic in two different loci of the *M. truncatula* genome, each encompassing tandem duplications in the legume.
Expression patterns of the MATE transporter gene family

The gene expression pattern of a gene is a key piece of information to unveil the physiological roles their proteins play in an organ, tissue, or cell. In order to understand the potential roles of MATE transporters in *M. truncatula*, RNA-seq expression data of different organs of a mature plant grown under optimal conditions were analyzed (Figure 2). MATE genes are expressed in virtually every organ of the *M. truncatula* plant, most showing tissue-specific expression patterns. Two genes (Medtr5g032750 and Medtr2g078970) were only expressed in leaves, whereas none were specifically expressed in vegetative buds. Medtr5g090280 and Medtr2g078990 were only detected in the flower, whereas five genes (Medtr6g027190, Medtr6g027200, Medtr5g090290-MtMATE1, Medtr8g063590, and Medtr6g081300) were exclusively expressed in seeds. Interestingly, none of the MATE genes were specifically expressed in roots. Twelve genes were only expressed in nodules, suggesting a role in nodule formation or function. While 12 genes did not show any expression in the samples tested, 13 genes were constitutively expressed. Twelve genes were mainly expressed in above-ground tissues with basal expression in below-ground organs, whereas on the flip side, seven genes showed high expression mainly in below-ground tissues and basal expression in above-ground samples.

To understand the functional divergence of segmentally duplicated MATE genes, we compared the expression profiles of gene homologs in different tissues. In total, the four paralog pairs derived from segmental duplications exhibited significant divergence in their expression patterns. In the first pair, Medtr2g088400 is highly and constitutively expressed, while Medtr4g048000 also showed constitutive expression, but with high expression limited to leaves, roots, and nodules. The second pair showed some expression overlap. Medtr3g111020 is highly expressed in nodules, with basal expression in seeds and vegetative buds, while Medtr3g5g067460 showed low expression in seeds and vegetative buds. In the third pair, Medtr8g106670 exhibited constitutive expression in all tissues, while Medtr6g081310 showed no expression, indicating it is potentially undergoing pseudogenization unless it went through subfunctionalization and is now expressed under specific conditions. The same case was seen in the fourth pair, which Medtr2g097900 exhibited constitutive expression while Medtr8g037175 showed expression in none of the tissues. Therefore, the genes derived from segmental duplications may have evolved towards functional divergence, at least to some degree.
based on their distinct expression patterns. Similarly, expression divergence was also observed for in-tandem duplication genes. We detected 11 clusters with a maximum of 8 in-tandem duplicated MATE genes. In one case (group seven of Table 1 Medtr6g027190/200), both genes showed a low expression level in seeds. In the remaining ten groups, at least one gene per group revealed a different expression pattern, indicating functional divergence from its original metabolic role prior to the duplication.

**Conclusions**

Previous studies showed the importance and versatility of plant MATEs. In conflict with the early assumption, they are not restricted to efflux of xenobiotics, metabolic waste, and metal detoxification. In plants they are also involved in many other cellular mechanisms, including mediating stress tolerance as well as participating in a wide range of physiological processes, such as regulating plant growth, organ initiation and development; ion homeostasis; translocation of hormones and other signaling molecules; vacuolar sequestration of secondary metabolites; and extracellular efflux of organic acids. So, a comprehensive genome-wide analysis of the MATE family was performed in an important model legume species *M. truncatula*. Here, we identified and analyzed 70 MATE genes in the *Medicago truncatula* genome regarding their expression patterns and phylogenetic relationships to predict potential functions. Segmental and whole genome duplication largely contributed to structuring the MATE family in this species. Tandem duplication of genes, possibly via unequal crossing over, was an important mechanism for the drastic expansion of this gene family in plant genomes.

Indeed, MATE transporters display considerable variation in their evolutionary history, gene structure, and tissue expression pattern, indicating a diversity of physiological functions. Considering the sessile lifestyle of terrestrial plants and their opulent secondary metabolism, the massive expansion of efflux transporters in plant genomes in comparison to other kingdoms is not surprising nor are the novel and essential roles many MATE transporters play in plant physiology and defense mechanisms. Our study sets the stage for a systematic functional characterization of MATE transporters in *M. truncatula*. 
Methods and materials

Identification of MATE transporters in the \textit{M. truncatula} genome

The whole proteins dataset of \textit{M. truncatula} (www.medicagogenome.org) was submitted to the TransportTP transporter prediction tool (http://bioinfo3.noble.org/transporter/) (Li et al., 2009) for identification and classification of membrane transporters into transporter families based on the Transporter Classification system (TCDB, www.tcdb.org) (Saier et al., 2016; Saier et al., 2006). Briefly, the criteria for the MATE family was previously described (Benedito et al., 2010) and the transporters were assessed with the typical features of the MATE transporters in plants (e.g., 12 transmembrane domains, >300 amino acid residues, the similarity to plant MATEs in curated as well as comprehensive databases), and assigned a functionality confidence level >=2 (1 for highest confidence the gene product is functional; 2 for a likely functional gene product, with a few features off or missing; and 3 for possibly pseudogenes and non-functional MATE transporters).

Phylogenetic analyses

A phylogenetic tree was generated using protein sequences by a Maximum Likelihood (ML) algorithm with 1,000 bootstraps based on a JTT matrix-based model in MEGA v.6.0 (Tamura et al., 2013), after alignment of protein sequences was performed with ClustalW (www.ebi.ac.uk/clustalw). It involved 186 full-length MATE genes from \textit{M. truncatula} (70), \textit{A. thaliana} (56), selected characterized plant MATE transporters (26), and all listed MATE genes from TCDB (family 2.A.66.1). The initial tree for the heuristic search was obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using a JTT model. The tree was drawn to scale, with branch lengths measured in a number of substitutions per site.

Synteny analysis

Synteny analysis was carried out using the SynMap tool (http://genomevolution.com/CoGe/SynMap.pl) based on coding sequences (JCVI genome sequence v.4.0: www.medicagogenome.org) as input data and default parameters (Zheng et al., 2015). The in-tandem gene duplicates in a genome (defined here as similar genes
at most ten genes apart with a threshold e-value ≤10e-4) were determined using Blast2raw script in the CoGe toolkit (Haug-Baltzell et al., 2017). Syntenic blocks (microsynteny within a genome or shared synteny between two genomes) were identified with SynMap through the DAGchainer algorithm using the relative gene order option with a minimum of 5 gene pairs aligned and a maximum distance of 20 genes between two matches. QuotaAlign was used to merge adjacent syntetic blocks (Zheng et al., 2015).

References


**Bouchenak M, Lamri-Senhadji M** (2013) Nutritional quality of legumes, and their role in


Cárdenas L, Domínguez J, Santana O, Quinto C (1996) The role of the nodl and nodJ
genes in the transport of Nod metabolites in Rhizobium etli. Gene 173: 183–187


dicarboxylic acids. Yeast 11: 1485–91


Li F, Chung T, Pennington JG, Federico ML, Kaeppler HF, Kaeppler SM, Otegui


Lozano-Baena MD, Prats E, Moreno MT, Rubiales D, Pérez-de-Luque A (2007)


Marinova K, Pourcel L, Weder B, Schwarz M, Barron D, Routaboul J-M, Debeaujon


Oke V, Long SR (1999) Bacterial genes induced within the nodule during the 

as fundamental transporters of metabolic and xenobiotic organic cations. Trends 
Pharmacol Sci 27: 587–93

Park H, Bakalinsky AT (2000) SSU1 mediates sulphite efflux in Saccharomyces 
cerevisiae. Yeast 16: 881–8

Pawlowski K, Sirrenberg A (2003) Symbiosis between Frankia and actinorhizal plants: 

Pérez-Hormaeche J, Potet F, Beauclair L, Le Masson I, Courtial B, Bouché N, 
Lucas H (2008) Invasion of the Arabidopsis genome by the tobacco 
retrotransposon Tnt1 is controlled by reversible transcriptional gene silencing. Plant 
Physiol 147: 1264–78

PCR 87–112

Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT- 
PCR. Nucleic Acids Res 29: e45


Pfeil BE, Schlueter J a, Shoemaker RC, Doyle JJ (2005) Placing paleopolyploidy in 
relation to taxon divergence: a phylogenetic analysis in legumes using 39 gene 
families. Syst Biol 54: 441–454

Phillips DA, Joseph CM, Maxwell CA (1992) Trigonelline and Stachydrine Released 
from Alfalfa Seeds Activate NodD2 Protein in Rhizobium meliloti. Plant Physiol 99: 
1526–31

Pineau C, Loubet S, Lefoulon C, Chalies C, Fizames C, Lacombe B, Ferrand M, 
Loudet O, Berthomieu P, Richard O (2012) Natural Variation at the FRD3 MATE 
Transporter Locus Reveals Cross-Talk between Fe Homeostasis and Zn Tolerance 
in Arabidopsis thaliana. PLoS Genet 8: e1003120

Pislariu CI, Murray JD, Wen J, Cosson V, Muni RRD, Wang M, Benedito V a, 
retrotransposon insertion mutant collection with defects in nodule development and 

Podgorny O V., Lazarev VN (2017) Laser microdissection: A promising tool for


Schroeder JI (1995) Anion channels as central mechanisms for signal transduction in


Sun SJ, Qi GN, Gao QF, Wang HQ, Yao FY, Hussain J, Wang YF (2016) Protein kinase OsSAPK8 functions as an essential activator of S-type anion channel OsSLAC1, which is nitrate-selective in rice. Planta 243: 489–500


Vincill ED, Szczyglowski K, Roberts DM (2005) GmN70 and LjN70. Anion transporters
the symbiosome membrane of nodules with a transport preference for nitrate.

Plant Physiol 137: 1435–44


Young ND, Udvardi M (2009) Translating Medicago truncatula genomics to crop


Zhao M, Chen Y, Qu D, Qu H (2011b) TSdb: a database of transporter substrates linking metabolic pathways and transporter systems on a genome scale via their shared substrates. Sci China Life Sci 54: 60–4


the aluminium tolerance of barley by expressing the citrate transporter genes SbMATE. J Exp Bot 65: 2381–2390


Chapter 3

Characterization of a nodule-specific MATE membrane transporter in *Medicago truncatula*

**Summary**

The legume *Medicago truncatula* is a well-established model organism that allows the study of fundamental features of the Fabaceae family, such as symbiosis nitrogen fixation. We identified a MATE transporter, MtMATE30 (Medtr7g082810) which is exclusively expressed in nodules. The expression of MtMATE30 in nodules is induced 100-fold by low nitrogen while it is strongly repressed by nitrate, suggesting a role in symbiotic nitrogen fixation. So far, most characterized MATE transporters function as H+/Na+-coupled antiporters in the efflux of substances out of the cytoplasm. The MtMATE30 gene spans 4.2 Kb and has seven exons. It encodes a predicted protein of 480 amino acid residues with 12 transmembrane domains. MtMATE30 expression starts in immature nodules (6 days post inoculation, dpi), and reaches a peak in young, mature nodules (10 dpi) and maintains consistent expression in older nodules. It is expressed in all nodule zones except the meristem. MtMATE30 RNAi hairy roots were developed but did not show a clear distinct phenotype compared to wild-type nodules. MtMATE30 canonically belongs to a phylogenetic clade that includes transporters with affinity to alkaloids. Trigonelline is a widely distributed alkaloid and commonly found in legumes. Previous research showed that rhizobia are able to catabolize trigonelline by the *trc* gene located in the pSym megaplasmid of *Rhizobium meliloti* RCR2011 (Boivin et al., 1991). Through LC-MS, the presence of trigonelline was confirmed in nodules of *Medicago truncatula*. Furthermore, MtMATE30 has an affinity to trigonelline as tested by HPLC/MS in heterologous bacteria system expressing MtMATE30. Altogether, MtMATE30 may be participating in alkaloid metabolism of nodule cells although its precise physiological role in symbiotic nitrogen fixation still requires further investigation.

**Keywords:** Fabaceae, Metabolite Transport, Mutualism, Secondary Metabolism
Introduction

Legumes are the third largest plant family in the world and provide essential nutrients for humans and livestock. Many species in this family are able to establish symbiotic nitrogen fixation (SNF) with soil bacteria, rhizobia. This process occurs in a specialized organ-nodule, in which bacteria reduce nitrogen N\textsubscript{2} into ammonia and deliver to the plant for growth and development in exchange for reduced carbon derived from photosynthesis along with other necessary nutrients to the microsymbiont. The legume-rhizobia symbiosis contributes in about 40 million tons of reduced nitrogen into crop fields each year and is an essential component of sustainable agricultural ecosystems (Udvardi and Poole, 2013). Thus, understanding the whole process of SNF is necessary for optimization of legume metabolism to sustainably boost crop yields.

The barrel medic (Medicago truncatula) is a model legume species with well-established genomic data as well as molecular and genetic resources available. It is an autogamous diploid species with a small genome (~390 Mb) distributed in 8 chromosomes, a relatively short life cycle (~3 months), and a large collection of ecotypes and mutants (Lozano-Baena et al., 2007; Colditz and Braun, 2010; Young et al., 2011; Tang et al., 2014). It is an ideal model for studying important biological processes that are not present in Arabidopsis thaliana, such as symbiotic root interactions, development of a complex leaf pattern, and secondary metabolism of isoflavonoids (Zhao and Dixon, 2009; Young et al., 2011). As a member of the Galegoid (Hologalegina) clade (Lavin et al., 2005; Young and Udvardi, 2009), M. truncatula is closely related to important temperate legume crops (e.g., alfalfa, pea, lentils), which made it easier to transfer the research results into the field production.

The establishment of SNF requires comprehensive chemical signals and nutrient exchanges between symbionts. Firstly, nitrogen-limited legumes secrete flavonoids into the rhizosphere and, upon receiving this signal, free-living rhizobial bacteria in the soil activate the synthesis of lipochitooligosaccharides (LCOs) known as nodulation (Nod) factors. Perception of Nod factors by the host plant triggers the reprogramming of root cells to allow bacterial invasion via tube-like infection threads (ITs) that invaginate into the root cortex. Bacteria are subsequently released via endocytosis into the cortical cells of an already dividing nodule primordium, thus forming an organelle-like structure called symbiosome. From then on, most chemical exchanges between the symbionts must
traverse both the symbiosome and bacteriod membranes through substrate-specific membrane transporters. Therefore, membrane transporters play essential roles in nodule development and physiology, including metabolism, ion homeostasis, signal transduction and recognition, and energy economy. Importantly, since the symbiosome membrane (SM) is derived by endocytosis from the plasma membrane (PM), it has an inverted nature (inside-out) to that of the PM (Verma and Hong, 1996; Udvardi and Day, 1997). Despite the extensive advances in legume genomics in recent years, very little is known about the transporters located in the SM that enable SNF.

Due to their sessile nature, higher plants have evolved to their surrounding environment by synthesizing a plethora of secondary metabolites, which are involved in many developmental and growth processes, protection against biotic and abiotic stresses, and plant communication. Alkaloids represent one of the largest groups of structurally diverse nitrogen-containing secondary metabolites with over 27,000 known compounds identified to date (Amirkia and Heinrich, 2014). The functions of plant alkaloids have been mainly associated with the interaction with microorganisms and herbivory (Schrittwieser and Resch, 2013). A growing body of studies has been confirming the importance of alkaloids for bacterial colonization of host plants. For instance, the monoamine alkaloid β-phenethylamine (β-PHA) is produced and highly accumulated in bacteroids and is associated with the differentiation of bacteria into the nitrogen-fixing bacteroids within the plant host cells (Fujihara et al., 2006). Moreover, alkaloids have also been shown to play a trophic role by providing nutrients for the microsymbiont (Boivin et al., 1990; Boivin et al., 1991). Trigonelline is a widely distributed pyridine alkaloid derived from nicotinic acid (niacin) in plants and found in most legumes (Tramontano et al., 1982; Boivin et al., 1990; Allred et al., 2009). It has several biological activities, including growth inhibition of certain mutant strains of Streptococcus (Minorsky Peter, 2002; Rojas-Andrade et al., 2003; Zheng et al., 2005). Trigonelline was found in alfalfa root exudates and activated the expression of nodD2 factor in Sinorhizobium meliloti, thus inducing Nod factor biosynthesis (Phillips et al., 1992). Trigonelline is also catabolized by S. meliloti strain RCR2011 by activating transcription of the trc gene located on the pSym megaplasmid, which is only activated during symbiosis (Boivin et al., 1991). Earlier research proposed that trigonelline might be a nutrient and provided to rhizobia by the plant during SNF, although knockout mutants do not display any significant phenotypes when grown in culture or during nodulation in legumes (Boivin et al., 1990). The symbiosome membrane transporter responsible for effluxing trigonelline into the symbiosome remains unknown. The exact role of trigonelline
in SNF and nodule physiology can only be assessed when the key genes regulating its biosynthesis and transport are found.

In recent years, a growing number of studies have shown that the Multidrug and Toxic Compound Extrusion (MATE) family of membrane transporters are involved in alkaloid and flavonoid exchange. Although MATE transporters were initially identified in prokaryotes (Morita et al., 1998; Brown et al., 1999), they are widely spread in all kingdoms of living organisms, from bacteria to plants and animals. Most members are comprised of 400-700 amino acids spanning 12-13 transmembrane domains (TMDs).

Plant genomes usually contain a high number of MATE genes, in contrast to those of prokaryotes and animals, which harbor just a few. For instance, there are 56 MATE genes in the A. thaliana genome (Liu et al., 2009), 117 in Glycine max (Liu et al., 2016), 46 in Oryza sativa (Liu et al., 2016; Wang et al., 2016), 67 in Solanum lycopersicum (Santos et al., 2017), and 70 in M. truncatula (Wang et al., 2017). Meanwhile, there are only two in humans (HsMATE1 and HsMATE2) (Omote et al., 2006). The radiation of the MATE family in plant genomes may in part be the result of several whole-genome hexaploidy (pre-rosid WGH) as well as local segmental duplications retained during the course of evolution to carry out transport of the diversity of secondary metabolites produced by plants. Although the biochemical properties of MATE transporters have not been fully elucidated so far, their abundance in plant genomes suggests they play diverse and essential physiological and metabolic roles in growth and development (Burko et al., 2011). Indeed, plant MATE transporters have been implicated in nutrient homeostasis (Green and Rogers, 2004, Liu et al., 2009); hormone transport, including abscisic acid (ABA), salicylic acid (SA) and auxin (Serrano et al., 2013; Yamasaki et al., 2013; Zhang et al., 2014); primary and secondary metabolism (Zhao and Dixin, 2010; Yazaki, 2005), and in biotic and abiotic stresses (Yamasaki et al., 2013). Although still limited, one progress made on functional characterization of MATE transporters demonstrates their essentiality to plant physiology and metabolism. The role of MATE transporters in legume nodules has been so far unexplored.

Here we characterized a novel MATE transporter from M. truncatula (Medtr7g082810, MtMATE30) as a nodule-specific alkaloid transporter. MtMATE30 has the highest expression in nodules among all 70 of MATE genes identified in the species (Chapter 2). During nodule development, MtMATE30 starts at four dpi and continues until senescence. Expression analysis in roots infected with S. meliloti mutant strains impaired
in specific steps of the symbiosis revealed that MtMATE30 expression requires bacteroid differentiation. Within the mature nodule, MtMATE30 is expressed throughout the nodule, except the meristem. Transport assays showed MtMATE30 affinity to the alkaloid trigonelline and no affinity to the flavonoid luteolin. We discuss the potential physiological function of MtMATE30 for symbiosis nitrogen fixation in legumes.

**Results and Discussion**

**MtMATE30 is closely related to MATE transporters with affinity to alkaloids**

Using the *M. truncatula* gene expression Atlas (https://mtgea.noble.org/v3/; (Benedito et al., 2010)) and RNAseq data (Young et al., 2011), we identified a putative MATE transporter with a nodule-specific expression pattern (Medtr7g082810; Benedito et al., 2010). Following a previous analysis of the MATE family in *M. truncatula*, this gene has been named *MtMATE30* (see Chapter 2; Wang et al., 2017). Among the 70 members in the MATE family, MtMATE30 is now the sixth MATE gene to be characterized in this species (Zhao and Dixon, 2009; Zhao et al., 2011a; Wang et al., 2017). It is located on chromosome 7 and spans a genomic region of 4,219 bp, consisting of seven exons (1440 nt coding segment) (Figure 1A). The MtMATE30 putative protein contains 480 amino acid residues encompassing 12 transmembrane domains (Figure 1B), with both terminal tails located in the intracellular (cytosolic) space (Figure 1B) suggesting MtMATE30 act as an effluxer. Based on the current version of sequence data available (IMGAG annotation v.4.2), MtMATE30 has no annotated 5’-untranslated region (UTR). Since every gene needs the 5’-UTR to initiate translation, we further verified experimentally and were able to amplify a cDNA containing 133 nt upstream of the start codon. Therefore, the transcription start site (TSS) predicted for MtMATE30 is incorrectly calculated in the official genome annotation, and the actual full-size of it’s 5’-UTR remains unknown. We looked into the RNA-seq read mapping of the closest homologue in the G. max genome on the GBrower (https://phytozome.jgi.doe.gov/ibrowse/) and reads of exons and introns perfectly aligned with MtMATE30, except the 5’-UTR segment. The soybean homologue contains 110-bp 5’-UTR, indicating MtMATE30 might also share a 5’-UTR of similar length.
A phylogenetic tree was generated using only functionally characterized MATE transporters from *M. truncatula*, *A. thaliana*, and other plants to compare MtMATE30 to other plants MATE transporters (Figure 2). All MATEs were grouped into four distinct clades (I-IV), which is consistent with the tree shown in Chapter 2. Given their diverse substrates and considering their varied subcellular localizations, the phylogenetic clades of the MATE family indicate affinity for molecules generally associated with distinct physiological and cellular processes. The phylogenetic analysis displays MtMATE30 clustered within clade IV, which has been identified as involved in cell detoxification. MtMATE30 is closely related to the nicotine effluxer NtJAT1 (Jasmonate-inducible Alkaloid Transporter 1) from *Nicotiana tabacum* (tobacco), sharing 61.7% sequence similarity at the amino acid level. NtJAT1 localizes to the tonoplast of leaf cells and has an affinity to other endogenous tobacco alkaloids, but not flavonoids (Morita et al., 2009). MtMATE30 is also close to Detoxification 1 (AtDXT1), the first MATE transporter characterized in *A. thaliana*. AtDXT1 is a plasma membrane transporter capable of exporting toxic xenobiotics, such as tetramethylammonium chloride and ethidium bromide, to the extracellular space in the bacterial mutant kam3 (Li et al., 2002). Clade IV also encompasses ABERRANT LATERAL ROOT FORMATION 5 (ALF5/DTX19) and DTX18. ALF5 is expressed in root epidermal cells and required for protection against toxic compounds, such as polyvinylpyrrolidone and pyrrolidone (Diener, 2001). AtDXT18 is responsible for secretion of the antimicrobial compound, coumaroylagmatine, in response to *Phytophthora infestans* colonization with *Arabidopsis thaliana* (Dobritzsch et al., 2016). Given the phylogenetic proximity of MtMATE30 with functionally characterized alkaloid transporters,
we hypothesized that it could be an alkaloid transporter and may have an affinity to alkaloids in nodules, not flavonoids.

Figure 2. The phylogenetic tree of MtMATE with other 44 functionally characterized plant MATE transporters. The tree was constructed with MEGA 6, using protein sequence from *Medicago truncatula* (IMGAG v.4.2), *Arabidopsis thaliana* (TAIR10: www.arabidopsis.org/portals/education/aboutarabidopsis.jsp). Node numbers represent the percentage of 1,000 bootstrap replications to assess the robustness of each node. Branch lengths are relative to estimated amino acid change per site (bar = 0.5). *Arabidopsis thaliana* (At), *Glycine max* (Gm), *Lotus japonicus* (Lj), *Medicago truncatula* (Mt), *Oryza sativa* (Os), *Nicotiana tabacum* (Nt), *Triticum aestivum* (Ta), *Eucalyptus camaldulensis* (Ec), *Brassica oleracea* (Bo), *Populus trichocarpa* (Pt), *Brassica rapa* (Br), *Secale cereale* (Sc), *Lupinus albus* (La), *Hordeum vulgare* (Hv), *Vitis vinifera* (Vv), *Sorghum bicolor* (Sc), *Zea mays* (Zm), *Malus domestica* (Md), and *Gossypium hirsutum* (Gh).
MtMATE30 is a nodule-specific transporter expressed throughout the whole indeterminate nodule, except the meristem

Analysis of its gene expression pattern in various organs (e.g., leaf, flower, root, nodule, seed, vegetative bud) (Benedito et al., 2008) indicates MtMATE30 is exclusively expressed in nodules (Figure 3A), which indicated it might be involved in nodule formation or function. We further explored its expression pattern in the different development stage (0, 4, 6, 10, 14 and 28 dpi). MtMATE30 expression was barely detectable at 4 days post-infection (dpi), peaked at 10 dpi (young, mature nodule), and remained steady in the functional nodule at 21 and 28 dpi (Figure 3B). Transcription of MtMATE30 decreased dramatically (10-fold) two days after the application of nitrate, a known repressor of SNF (Streeter, 1985; Van Noorden et al., 2016). This result indicates a potential SNF role of MtMATE30. The MtMATE30 gene expression pattern was also confirmed by a time-course analysis via RT-qPCR on nodulating roots (Figure 4A). These results suggest MtMATE30 may be involved in supporting or maintaining the nodule function instead of nodule formation. To visualize the tissue expression pattern of MtMATE30, a 2.0 kb promoter region upstream of the start codon was fused to the β-glucuronidase (GUS) reporter gene (proMtMATE30::GUS) in hairy roots. GUS expression revealed that MtMATE30 is indeed expressed throughout the indeterminate nodule of M. truncatula with a (perhaps artifactual) leaky expression detected in root vascular bundles (Figure 5).
Figure 3. Temporal and spatial expression profile of *MtMATE30* in *M. truncatula*. (A) The Medicago Gene Atlas revealed *MtMATE30* (Mtr.49406.1.S1_at) is specifically expressed in nodules and nodulating roots ([https://mtgea.noble.org/v3](https://mtgea.noble.org/v3)) (Benedito et al., 2010). (B) Public RNA-Seq data confirms *MtMATE30* specific expression in nodule samples (Young et al., 2011).
Figure 4. Expression profile of MtMATE30 in inoculated roots by quantitative RT-PCR. (A). Time-course of MtMATE30 expression using M. truncatula ‘A17’ roots inoculated with the functional rhizobium strain. (B). Analysis of MtMATE30 expression in M. truncatula ‘A17’ roots at 28 days post-inoculation with rhizobium mutant strains impaired in specific steps of symbiosis establishment (exoA, bacA and fixJ) along with the functional rhizobium strain Sm1021_lacZ. NC represents a negative control. Asterisks indicate a statistically significant difference in the expression in relation to the control (*α<0.05, **α<0.01) according to the one-way ANOVA following by Dunnett’s test. Mean±SE of three biological replicated are presented.

We further examined the MtMATE30 expression in roots infected with S. melliloti mutant strains impaired in specific milestone stages of the symbiosis establishment (Figure 4B). The exoA mutant is deficient in the production of succinoglycan, a known extracellular polysaccharide, and fails to induce the shepherd’s crook formation in the root hair or form infection threads (Leigh et al., 1985; Mitra and Long, 2004). M. truncatula roots infected with exoA rhizobial mutant showed a complete lack of nodule development at 28 dpi (data is not shown). MtMATE30 expression was almost undetectable, indicating the requirement of infection for its expression. The bacA mutant strain shows normal infection thread formation and invasion, but it fails to differentiate into bacteroids
bacA nodules showed arrested development as small white bumps, which cannot perform normal nitrogen fixation. MtMATE30 transcripts were detected but significantly lower than the control with the wild-type bacteria (strain Sm1021). fixJ is a signal transducing regulator in rhizobia that regulates the expression of nitrogen fixation-related genes by sensing oxygen tension. Nodules elicited by fixJ mutant bacteria are able to fully differentiate into bacteroids, and show an intermediate phenotype that looks like the wild-type nodule, which can fix nitrogen but undergoes early senescence (Oke and Long, 1999; Saito et al., 2003; Mitra and Long, 2004; Sinharoy et al., 2013). MtMATE30 was highly expressed in the fixJ nodule and showed no difference than control nodules. Overall, MtMATE30 is expressed throughout the nodule tissues, and its transcription is downstream to rhizobial infection, possibly with an onset during bacteroid development. Given this expression pattern, MtMATE30 might play a role in nodule formation or function.

MtMATE30 is expressed in all of the zones of the indeterminate nodule, except the meristem

As shown in Figure 5, MtMATE30 was expressed in mature nodules. More specifically, it was expressed in all of the zones with dark staining except meristem with light staining. However, the nodule with gus staining was over staining. We explored another technique (in situ hybridization) to visualize MtMATE30 expression in the nodule. The result from in situ hybridization clearly showed that MtMATE30 is expressed throughout all zones except the meristem (Figure 6). Moreover, MtMATE30 is highly expressed in the fixation zone as shown as the dark staining (Figure 6C). This was consistent with the expression of the development stages showing the figure 3B since nodules after ten days of post-inoculation were mature and had fully developed fixation zone without the senescence zone. The expression of MtMATE30 in nodule zones is consisted to the publicly available laser capture microdissection (LCM) coupling with RNA-Seq data (Roux et al., 2014) (Supplementation Figure S2). As shown in Figure, MtMATE30 was not detected in the meristem and the distal infection zone, gradually increased its expression in the proximal infection zone and interzone, and finally peaking in the fixation zone. There were two types of cells in the nodules: infected cells and uninfected cells. The majority of infected cells were located in the fixation zone, with the minority in the distal end of infection zone and
in the Interzone. As expected, *MtMATE30* had a much higher expression in infected cells than in the uninfected cells (Figure 6C, 6E). Additional LCM data based on Affymetrix microarrays on different nodule cell types showed *MtMATE30* is higher expression in infected cells than in uninfected cell (9X) (Limpens et al., 2013). Altogether, *MtMATE30* most likely located in the membrane of infected cells of the fixation zone, which may be responsible to efflux the substrates out of cells. Those expression profiles indicate *MtMATE30* may involve nodule nitrogen fixation instead of nodule formation, bacteria differentiation or nodule senescence.

**Figure 6. In situ hybridization of nodule section using a MtMATE30 probe.** Nodule sections (10-µm thick) were hybridized with antisense (A, C) and sense (B, D, E) 313-nt RNA antisense probe against the *MtMATE30* transcript. Purple precipitate indicates *MtMATE30* gene expression. Notice in (E) that in the nitrogen fixation zone, infected cells show much higher expression than non-infected cells.
Nodule phenotype of *MtMATE30* RNAi hairy roots

*MtMATE30* expression was knocked down in transgenic hairy roots by RNAi interference (RNAi). However, no Nod-symbiotic phenotype was found in the RNAi knockdown mutants (Figure 7A-B). We confirmed that *MtMATE30* expression in the RNAi nodules was significantly lower than the control nodules (Figure 7C). The *MtMATE30* RNAi mutants formed pink nodules, indicating normal development to organ maturity (Figure 7C).

![Figure 7. The phenotype of MtMATE30 knockdown nodules.](image)

Figure 7. The phenotype of *MtMATE30* knockdown nodules. Wild-type (left) and RNAi nodule (right) under light stereomicroscope (A) and fluorescent illumination (B). (C) qPCR of MtMATE30-RNAi and control lines. Asterisks indicate a statistically significant difference in the expression in relation to the control (**α<0.01) according to Student’s test. Mean±SE of three biological replicated are presented. Bars= 2 mm.

*MtMATE30* is an alkaloid-specific transporter

To examine whether MtMATE30 works as an effluxer, we expressed MtMATE30 in the heterologous system of *Escherichia coli* mutant strain *acrB* (Seo et al., 2012). *acrB* is highly susceptible to xenobiotics, such as antibiotics, dyes, bile salts and synthetic compounds. The *acrB* strain was complemented with the pDEST42-MtMATE30 vector and grown in the presence or absence of tetrabutylammonium chloride (TBA). TBA, a typical substrate for MATE transporters, is a noxious synthetic compound that can be universally excreted out of cells through the tripartite efflux system *AcrA/AcrB/TolC* by wild-type bacteria (Seeger et al., 2008). *acrB* cells with empty vector (control) grew slightly faster than the MtMATE30-complemented bacteria (Figure 8A) in a half-strength of LB
liquid medium without TBA. In contrast, the MtMATE30-expressing acrB bacteria showed higher growth rate compared to the control (Figure 8B) with 1mM TBA. As expected, 1 mM TBA inhibited the growth of non-transformed acrB cells, since it lacks the machinery to pump TBA out of the cell, whereas acrB complemented with MtMATE30 grew more quickly, indicating that MtMATE30 is able to act as a TBA effluxer in bacteria by mediating TBA excretion.

An important question during the characterization of a membrane transporter is its substrates. In our phylogenetic analysis, MtMATE30 fell in a branch of Clade IV transporters in which we functionally characterized specifically with alkaloid translocation. So far, there is very little information about which alkaloids are present in legume nodules. Trigonelline (1-methylpyridinium-3-carboxylate) is derived from the pyridine nucleotide cycle and a common alkaloid in plants, including legume nodules, seeds, roots, and root exudates (Vasse et al., 1990; Rojas-Andrade et al., 2003). Although its role remains unknown in nodulation, it can be catabolized by rhizobia via the trc gene product, which is only induced during nodulation and mapped on the pSym symbiotic megaplasmids, along with most genes that are involved in infection, nodulation, and nitrogen fixation (Boivin et al., 1990). The symbiotically-induced catabolism of trigonelline is clear in endosymbiotic rhizobia, but little is known how it is transported to the bacteroids. Therefore, we propose that trigonelline is a potential substrate of MtMATE30. Therefore, we tested and confirmed the presence of trigonelline in mature nodules of M. truncatula via LC-MS/MS (Figure 9A-B).

Figure 8. Effects of tetrabutylammonium (TBA) on bacterial growth. Time-course analysis of bacterial growth without (A) or with 1mM TBA (B) in acrB bacterial lines with empty pDEST42 vector (black lines) or expressing the MtMATE30 transporter (red lines). Cell growth was monitored for 16h by measuring optical absorption at 600nm in 900-µL cell culture incubated in a plate reader at 37°C.
In order to first understand the effect of alkaloids on bacteria, we measured bacterial growth rates in LB media supplemented or not with 1 mM trigonelline or nicotine. Neither of the alkaloids exerted any influence on growth rates of acrB bacteria expressing or not MtMATE30 (Figure 10A-B). Since MtMATE30 is an effluxer, the intracellular alkaloid content in cell lines growing in LB supplemented with trigonelline was quantified by HPLC-MS. MtMATE30 complemented bacterial cells accumulated significantly less trigonelline than the control under all concentrations tested (Figure 10C). A time-course analysis also showed that the intracellular trigonelline content in MtMATE30-expressing cells was consistently lower than in control cells transformed with the empty vector (Figure 10D). Therefore, these results demonstrate that the MtMATE30 transporter mediates the efflux of trigonelline in the bacterial system and is a potential substrate *in planta*.

Since MtMATE30 is an effluxer, the physiological role of MtMATE30 in nodules may depend not only on which substrate it carries, but also in which membrane it is localized, since MATE transporters have been characterized in other plant species to localize to diverse membranes of the cell, such as the plasma membrane, Golgi complex, and tonoplast. Given its unique presence in infected nodules, no MATE transporters have been described that localize to the symbiosome membrane. As effluxers, all MATE transporters so far characterized pump their substrates out of the cytosol, either out of the cell or into organelles. Since trigonelline is produced by the host, it could be transported...
from the cytosol of infected cells into the symbiosome through MtMATE30. To achieve this goal, MtMATE30 must be located in the symbiosome membrane. If MtMATE30 is shown to pump trigonelline into the symbiosome, it may be involved in feeding bacteroids with secondary metabolites, although the biological reason for this activity remains to be verified. It is possible that trigonelline works as a reserve compound supporting NAD$^+$ biosynthesis in the bacteroid (Lima et al., 2014) since it has been experimentally confirmed that trigonelline can convert back to NAD$^+$ in germinating coffee seeds (Shimizu and Mazzafera, 2000). Tramontano et al. (1983) used [carboxy$^{14}$]trigonelline to trace the degree of trigonelline metabolism, and found out the large percentage of radioactivity was detected as NAD, NADP, and NMN besides trigonelline itself, members of pyridine nucleotide pathway (Figure 11) in seedlings of Pisum sativum (Tramontano et al., 1983). This demonstrates that trigonelline can be converted back to NAD$^+$ in these cells. In term of cellular energy economy, trigonelline may be considered as an important reserve molecule for NAD$^+$ in special processes or in a specific tissue (Shimizu and Mazzafera, 2000). For instance, the very beginning of seed germination is an energetic costly process, and the demand of NAD and NADP is vast (Shimizu and Mazzafera, 2000). Similarly to seed germination, the organogenesis of legume nodules is a sink with high demand for carboxylates (Udvardi and Day, 1997). Thus, trigonelline might serve as an NAD$^+$ reserve and be transported into symbiosomes, and then be converted into NAD$^+$ by bacteroid for supporting nitrogenase activity, since NAD$^+$ and NADH in tandem regulate the citric acid cycle and oxidative phosphorylation (Figure 11). This may explain why MtMATE30 knockdown mutants showed no phenotype. There are two ways of producing NAD$^+$ via de nova pathway and salvage pathways, while de nova pathway provides the majority of NAD$^+$. Knockdowning of a minority of salvage pathways of producing NAD$^+$ would not affect NAD$^+$ production in nodules. If trigonelline cannot be converted into NAD$^+$ in symbiosome, trigonelline may provide precursors for the citric acid cycle after metabolism by bacteroids (Boivin et al., 1990; Boivin et al., 1991). Alternatively, it has also been proposed that trigonelline exerts an auxiliary role in oxidative stress in legume nodules (Lima et al., 2014). The content of trigonelline in Catharanthus roseus tissue culture increased under several stresses, suggesting trigonelline may be involved in the induction of defensive metabolism (Berglund et al., 1996). Before N fixation begins, nodules require vast amount of energy from the host plant to support nodule formation, which the process put a lot of stress on the host plant. This may increase the production of trigonelline, which then is transported in bacteroids to relief the stress.
Figure 10. MtMATE30-mediated trigonelline transport in bacteria cells. (A-B) Time-course analysis of MtMATE30 transporter in bacterial cells with 1mM nicotine (A) and with 1mM trigonelline (B). Empty vector line (control, black line) and MtMATE30-expressing line (red line) in acrB bacterial mutant were incubated in half-strength LB medium supplemented with different chemicals and measured OD$_{600}$ value every 20 mins. (C) Intracellular accumulation of trigonelline in acrB bacterial cells incubated with media supplemented with different concentrations of trigonelline. MtMATE30-expressing cell line (gray bar) or empty vector (control, back bar) were incubated in half-strength LB medium with different concentration of trigonelline for 8 hours. Results are mean of five repetitions. Asterisks indicate statistically significantly different compared to control (**$P$<=$0.001$). (D) Time-course analysis of trigonelline transport in MtMATE30-expressed acrB cells (red line) along with the control (black line). Both bacterial line were incubated in half-strength SD medium supplemented with 1mM trigonelline and sampled at the times indicated. Results are mean (n=5) and bars represent standard errors. (E) Intracellular accumulation of the flavonoid luteolin in acrB bacteria after 8 hours of incubation in LB medium supplemented with 1mM luteolin.
Since some MATE proteins in clade IV also have an affinity to flavonoids, we tested luteolin as a potential substrate for MtMATE30. Luteolin is an early inducer of symbiotic nitrogen fixation. Under low nitrogen conditions, this flavonoid is exuded by roots and acts as an effector to activate Nod factor synthesis in rhizobia (Yeh et al., 2002; Nelson et al., 2017). We found no significant difference in the intracellular luteolin content between MtMATE30-expressing bacteria and the control (Figure 11E). Thus, our results indicate that MtMATE30 is a transporter specific to alkaloids. At this point, we cannot rule out that MtMATE30 has an affinity to other alkaloids. To further identify other potential substrates of MtMATE30 as well as to characterize the richness of its secondary metabolism in M. truncatula mature nodules, we are conducting a metabolomics analysis to compare the whole metabolome of wild-type and knockout nodules using LC-MS/MS.

**Perspective**

The lack of a clear phenotype in the knockdown nodule may be explained by RNAi lines often allowing residual levels of expression, which may be enough activity for normal development (Zhang et al., 2017). We would apply CRISPR/Cas9 to achieve complete gene knockout and see the phenotype of nodules. Also, MtMATE30 knockouts via CRISPR/Cas9 system would allow quantification and comparison of trigonelline content in nodules and further explore our hypothesis. Alternatively, MtMATE30 may not play an essential function in nodules, or may display gene redundancy, potentially with other
MATE genes. As we discussed in Chapter 2, MtMATE30 falls in clade IV, along with other 23 MATE transporters in *M. truncatula*. Of these, seven are highly expressed in nodules and may play overlapping functions, possibly through a certain level of substrate promiscuity.

**Material and Methods**

**Plant growth and biological material**

*M. truncatula* (ecotype A17) seeds were scarified with concentrated sulfuric acid for 12min, rinsed several times with cold water, followed by surface sterilization with a 1:3 (v/v) dilution of commercial bleach (6% w/v of ClO-) containing 1-2 drops of Tween 20 for 10min and rinsed three times with sterile water. Seeds were placed in a Petri dish containing damp, sterile filter paper and incubated in the dark at 4°C for vernalization for three days, transferred to a growth chamber (200 μE m⁻² s⁻¹ for a 16-h photoperiod, 24°C day/21°C night, and 40% humidity) for four days. Seedlings were then transferred to 5-inch plastic containers with washed and double-autoclaved vermiculite/turface mixture (1:2) and placed in a growth chamber. Up to the day of inoculation with rhizobia, plants were supplied only with distilled water. After seven days of growth, plants were inoculated with *Sinorhizobium meliloti* strain Sm1021 containing the pXLGD4 plasmid (*lacZ* vector) or *S. meliloti* mutant strains (*bacA*, *fixJ*, *exoA*), which are also in the Sm1021 background (Boivin et al., 1990). The *S. meliloti* culture was grown overnight in yeast mannitol broth (YMB) medium at 28°C with shaking at 250 rpm until *OD₆₀₀*=1.0, pelleted by centrifugation at room temperature (4412 g for 15min), resuspended in low-nitrogen half-strength B&D medium (Broughton and Dilworth, 1971) to *OD₆₀₀*= 0.02. Each plant was supplied with 50 ml of the inoculum. Subsequently, B&D medium with low nitrogen (1mM KNO₃) was added to the tray every week up to the point of harvesting. For seed production, plants were transferred from mixture medium to 4-inch pots with soil after 28 post-inoculation and cultivated in a greenhouse.

**Identification of MtMATE30 in M. truncatula and phylogenetic analysis**

The *MtMATE30* gene (Medtr7g082810: *Medicago truncatula* Genome Project IMGAG v4.2: www.medicagogenome.org) was initially identified as a nodule-specific gene based
on microarray expression profiles (Affymetrix probeset Mtr.49406.1.S1_at) using the Medicago truncatula Gene Expression Atlas (Benedito et al., 2008; He et al., 2009; Benedito et al., 2010) (mtgea.noble.org/v3). Alignment of the MtMATE30 amino acid sequence and other proteins of the MATE family was performed using Clustal Omega (www.ebi.ac.uk/Tools/msa/clustalo). The phylogenetic tree was constructed with MEGA v.6 (Tamura et al., 2013) using the maximum likelihood (ML) method with arithmetic mean and 1,000 bootstrap replicates.

RNA extraction and cDNA synthesis

To measure MtMATE30 expression by qPCR during nodulation, the whole roots with nodules were harvested from 0 to 28 dpi, immediately frozen in liquid nitrogen and kept at -80°C until processing. Total RNA was extracted using the mirVana miRNA isolation kit (ThermoFisher) with the protocol for total RNA extraction and treated with DNA-free DNase I kit (Ambion). RNA was then quantified using Nanodrop 100. cDNA synthesis was performed with SuperScript III reverse transcriptase (ThermoFisher) using 2 μg of DNase-treated total RNA and 1 μl of oligo(dT)20 in a 10-μl reactions.

qRT-PCR analyses

All qPCR reactions were performed in three biological replicates and three technical replicates, and displayed as relative expression ratios. The reactions were carried out in an ABI 7500 System (Applied Biosystems) in 5-μl reactions containing 2.5μl 2X SYBR Green Power Master Mix (Applied Biosystems), 1μl cDNA (diluted 1:10), 0.5μl of each gene-specific primer (final concentration: 200 nM) (Supplemental Table 1) and 0.5μl H2O. Non-template controls were included for each gene pair. PCR cycles were the following: 50°C for 2min; 95°C for 10min; 40 cycles of 95°C for 15s and 60°C for 1min with SYBR Green fluorescence detected continuously. After 40 cycles, melting curves were generated by heating the sample up to 95°C for 15s followed by cooling down to 60°C for 1min and then heating the samples up to 95°C with a reading of fluorescence every 1°C increase. Transcript levels were normalized using the geometric mean of three housekeeping genes: MtEF1a (ethylene response factor, Medtr4g008860), MtUbi (Ubiquitin, Medtr3g091400), and MtMSC27 (Medicago sativa cDNA 27). Primer efficiency for each primer pair was conducted independently using a mixture of all cDNA samples in a dilution curve to calculate the relative transcript abundance of each gene of interest. Relative MtMATE30 expression levels were calculated by the Pfaffl method including
normalization to the constitutive housekeeping genes and control samples (non-inoculated roots) (Pfaffl, 2001; Pfaffl, 2004).

**Cloning the coding and promoter sequence of MtMATE30**

Primers for PCR-based cloning of MtMATE30 coding sequence (CDS) of *M. truncatula* ecotype A17 were designed using the Primer3 tool (bioinfo.ut.ee/primer3) with the official genomic sequence and annotation (www.medicago.org). cDNA was produced from total RNA of 28-dpi nodules of *M. truncatula* ecotype A17 as a template for amplification using the high-fidelity KOD Hot Start DNA Polymerase (Millipore) of the MtMATE30 transcript with stop-codon and 3'-UTR or without the stop codon for fusion with a fluorescent marker. The PCR product was directly cloned into pENTR/D Gateway entry vector (ThermoFisher). For substrate transporter assays in the *E. coli* mutant strain acrB, the entry vector was recombined with pDEST42 destination vector via LR reaction (ThermoFisher).

To examine the spatial distribution of MtMATE30 gene expression in nodules, its promoter (2,775bp upstream of the coding region) was amplified using leaf genomic DNA from ecotype A17 as a template and the high-fidelity KOD Hot Start DNA Polymerase (Millipore). The PCR products were cloned into pENTR/D and recombined using LR reaction (Gateway LR clonase enzyme mix, ThermoFisher) into the destination vector pKGWFS7 (Karimi et al., 2002) with dual reporter genes (GFP-GUS) to create proMtMATE30::GFP-GUS fusion expression vector. The maps representing each of the final constructs used are included as Supplemental Figure S1.

**Escherichia coli transport assays**

*E. coli* strain acrB is highly sensitive to xenobiotic drugs since it modified from AD1-8 strain (Morita et al., 2009; Shitan et al., 2014). For complementation assays, the MtMATE30 cDNA was subcloned into the pDEST42 expression vector (Ghasemi et al., 2014), and the final construct was transformed into acrB mutant cells. Bacteria transformed with the empty pDEST42 vector was used as a control. The bacterial transformants were precultured in 3mL LB medium supplemented with 100 µg/L carbenicillin for 12 hours. The cell cultures were diluted to OD<sub>600</sub>=0.5 with a 10-fold gradient series, and 1ml LB medium containing 1mM TBA, nicotine or trigonelline was distributed in 96-well plates. Absorbance (λ<sub>600nm</sub>) of liquid cultures was monitored every 20min in a microplate reader at 37°C for 16 hours. No alkaloid supplementation in the medium was included as a negative control.
Each treatment was carried out in three biological (independent cultures) and three technical replicates.

For the concentration and the time course analyses of intracellular secondary metabolites for trigonelline and luteolin, the bacteria were precultured in 3ml LB medium with 100 μg/mL carbenicillin for 12 hours. One milliliter of bacterial suspension was transferred to 200ml LB medium with carbenicillin for 8 hours. The culture was spun for 10min at ~4417g (RCF), the pellet was resuspended in half-strength LB medium with different concentrations of trigonelline, luteolin or nicotine for 8h. The cells were harvested by centrifugation of 5ml culture; the pellet was washed twice with cold deionized water. Next, the fresh weight of the pellet was recorded and cells were disrupted with beads (2 mm φ zirconia beads, BioSpec) in extraction buffer (50% ethanol, 50% methanol, and 0.5% acetic acid) using a mini-bead beater (BioSpec) at maximum speed for 2 mins (Morita et al., 2009; Shitan et al., 2014). Samples were spun and the supernatant was then filtered through a syringe filter (0.22μm PES syringe filter, 13mm φ) for HPLC-MS analysis.

### HPLC-MS/MS equipment and detection parameters

The high-performance liquid chromatography (HPLC) system consisted of an Agilent 1100 series binary pump with a 100 vial autosampler (Agilent Technologies). The column oven temperature was maintained at 23°C. The mobile phase consisted of an 85:15 ratio mixture of a 0.1% aqueous ammonia solution in acetonitrile to a 0.1% aqueous ammonia solution in water. The aqueous ammonia solution used to prepare the mobile phase was 25% w/w NH₃. The flow rate was isocratic at 200μL/min through a 100 x 2 mm Chromolith RP-18e column with a 5 x 2 mm guard column (VWR International) directed into a Waters Quattro Micro triple quadrupole mass spectrometer (Waters Micromass). The sample was introduced into the mass spectrometer by an API probe operated in the positive ion mode. Multiple reaction monitoring (MRM) was performed on nicotine (m/z 162.97>131.85) with dwell times of 200msec. The mass spectroscopic conditions were: capillary voltage of 2.97 kV, cone voltage of 30V, source temperature of 105°C, desolvation temperature of 200°C, desolvation gas flow of 500 L/hr, collision gas pressure of 3 x 10⁻³ mBar, and collision energy set at 13 eV.

### Trigonelline identification in *M. truncatula* nodules
Alkaloids were isolated from mature nodules and quantitatively analyzed by LC-MS. Four microliters of nodule extract were injected into an LC System coupled to an Agilent 6530 quadrupole time of flight (Q-ToF) mass spectrometer with electrospray ionization operated in positive ion mode. A sole and clear peak appeared at ~3 minutes in the chromatogram, which MS/MS fragmentation pattern matched the Human Metabolome Database (HMDB: www.hmdb.ca) entry for trigonelline thus confirming its identity.

**Hairy root transformation, histochemical staining, and light microscopy**

Plants with composite transgenic roots were generated as described in the *Medicago truncatula* handbook ([www.noble.org/medicago-handbook/](http://www.noble.org/medicago-handbook/)) (Chabaud et al., 2006) with some modifications. Briefly, the constructs (*proMtMATE30::GUS; MtMATE30-GFP* or *MtMATE30-mRFP*) were transformed into *Agrobacterium rhizogenes* strain ARqua1 using the freeze-thaw method (Jyothishwaran et al., 2007). Transgenic ARqua1 lines were grown in 3mL LB preculture containing proper antibiotics until OD$_{600}$=0.9-1.5. A volume of 500µL inoculum was plated onto LB-agar with antibiotics and cultivated for two days at 28°C in the dark. Concomitantly, seeds of *M. truncatula* ecotype A17 seeds were sacrificed and sterilized as described above. When a dense layer of bacteria was formed on the plate, the seedling radicle was cut ~3mm from the root tip and the wound was immediately used to scrap the plate surface to coat it with bacteria. We planted the coated seedlings in 32-well trays containing autoclaved surface and vermiculite (1:2) mixture. 1.5 L of Low Nitrogen B&D (Benedito et al., 2008; Sinharoy et al., 2013) was added to the tray. The trays with plants were covered with a dome with the hole open for two weeks. Inoculation with *S. melliloti* rhizobial strains with plasmids containing the visual markers lacZ (pXLGD4-Sm1021), GFP (Sm1021-GFP) or mRFP (Sm1021-mRFP) were alternatively used for nodulation. For tissue localization of promoter::GUS assay, the nodules were harvested at 28 dpi and stained with X-Gluc (5-bromo-4-chloro-3-indolyl-β-D-glucuronic acid, cyclohexyl ammonium salt) (Boivin et al., 1990; Veereshlingam et al., 2004). Briefly, the nodules were excised from transgenic hairy roots and vacuum infiltrated for 20 min with fixing solution (0.3% *p*-formaldehyde, 10 mM MES (2-(N-Morpholino)-ethanesulfonic acid) pH 5.6; 0.3 M mannitol). Samples were left at room temperature for 1h followed by washing steps. The fixing solution was subsequently replaced by a washing solution (0.5 M MES and 0.3 M mannitol). Samples were rinsed for 20 min three times with moderate rotation movement and incubated in pre-warmed (37°C) staining solution containing (0.08% X-Gluc, 50 mM potassium ferricyanide, 50 mM potassium ferrocyanide, and 0.1 M PIPES
(pH 7.2)) for 16h at 37°C. Samples were rinsed in 0.1 M PIPES (pH 7.2), cleared with 70% ethanol for 30 min. The stained nodules were mounted in 5% low melting temperature agarose and sectioned into 50 μm with a 1000 Plus Vibratome (Technical Products International). Whole nodule images were taken under the MIF Olympus MVX10 system, a Dual upright microscope, and stereoscope. The sample sections were placed on glass slides and observed under a microscope (Nikon).

**In situ hybridization**

*In situ* hybridization was performed on 28-dpi nodules as previously described (Wiel et al., 1990) with some modifications. Briefly, nodules were fixed with 3% paraformaldehyde and 0.25% glutaraldehyde in 0.1 M sodium phosphate buffer (pH 7.2) at room temperature overnight, followed by graded ethanol series (50%, 75%, 80%, 95% and 100% twice) for 45 mins each. After, samples were put into xylene for 15 min and switched to the mixture of xylene and periplast (1:1) for two hours at 60°C. Finally, the nodules were embedded in paraplast and sectioned into 10-μm slices with a microtome. Sections were dried overnight on coated slices at 37°C. Before deparaffinization, the slices were incubated in 65°C for one hour, submerged twice in xylene (15 min each), and rehydrated by a graded ethanol series. Hybridization was using IsHyb In Situ Hybridization Kit (BioChain) following the manufacture’s protocol with DIG-labeled antisense and sense probes. The plasmid pGEM-T easy was digested with SacII and SpeI to yield templates for DIG-labeled antisense and sense probe production with SP6 and T7 RNA polymerase (New England Biolabs) respectively following DIG NA labeling and detected kit manual. The sectioned samples were placed on slides and observed with a microscope (Nikon).

**Analysis of *MtMATE30* RNAi nodules**

The target region of the MtMATE30 transcript (138-451 nt within first exon) was amplified with high-fidelity KOD DNA polymerase (Millipore), cloned into the pENTR/D entry vector (ThermoFisher), and transferred into the inverted-repeat binary destination vector pK7GWIWG2_II-RedRoot with DsRed fluorescent marker as a constitutive report gene (Jach et al., 2001; Limpens et al., 2004). Root transformation followed as described above and inoculation with *Sm1021*-GFP rhizobia was carried out. Nodules were harvested and phenotypically analyzed at 28 days post-inoculation. At harvest, transgenic roots and nodules were screened for presence of red fluorescence under MIF Olympus MVX10. Nodules samples were collected for RT-PCR analysis and processed as described above.
Statistical analysis

Statistical analyses were performed with one-way ANOVA followed by Student’s t-test.

ACKNOWLEDGEMENTS

Both, the bacterial mutant (acrB) and expression vector (pDEST42) were kindly provided by Dr. Nikola Kovinich. I would like to thank Dr. Yao for kindly providing pGEM-T Easy vector and IsHyb In Situ Hybridization Kit (BioChain). I would like to thank Dr. Terry McManus for his efforts to quantify trigonelline in bacterial cells via HPLC/UV-MS. Another very special word of gratitude goes to Gregory Boyce for helping with trigonelline identification in nodule samples via LC-MS/MS. I also like to thank Sue Meyers and Gail Sikorsky for helping with taking care of the Medicago truncatula plants in the greenhouse.

References


early step of Nod factor signal transduction upstream of a rapid calcium flux and root hair deformation. Plant J 34: 495–506


Fujihara S, Terakado J, Nishibori N (2006) Accumulation of an Aromatic Amine, β-


Kryvoruchko IS, Sinharoy S, Torres-Jerez I, Sosso D, Pislariu CI, Guan D, Murray J, Benedito VA, Frommer WB, Udvardi MK (2016) MtSWEET11, a Nodule-


Limpens E, Mirabella R, Fedorova E, Franken C, Franssen H, Bisseling T, Geurts R


metabolite transporters in tomato. BMC Plant Biol 17: 185


Shimizu MM, Mazafera P (2000) A role for trigonelline during imbibition and
germination of coffee seeds. Plant Biol 2: 605–611


Sun SJ, Qi GN, Gao QF, Wang HQ, Yao FY, Hussain J, Wang YF (2016) Protein kinase OsSAPK8 functions as an essential activator of S-type anion channel OsSLAC1, which is nitrate-selective in rice. Planta 243: 489–500


Plant Cell 21: 267–284


Zhao M, Chen Y, Qu D, Qu H (2011b) TSdb: a database of transporter substrates linking metabolic pathways and transporter systems on a genome scale via their shared substrates. Sci China Life Sci 54: 60–4


Supplementary figures and table

Supplementary Figure S1. Binary vector maps.
Supplementary Figure S2. Temporal and spatial expression of *MtMATE30* in different zones of mature nodules. Laser-capture microdissection (LCM) samples quantified via RNA-Seq (Roux et al., 2014). FI, zone I (nodule meristem); FIId: zone II (distal portion of the infection zone); FIIP: zone II (proximal portion of the infection zone); IZ: interzone (bacteroid differentiation zone); ZIII: zone III (nitrogen fixation zone).

**Supplemental Table S1** List of primers used in these studies.

<table>
<thead>
<tr>
<th>Quantitative RT-PCR</th>
<th>Sequence</th>
<th>Gene ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT-MATE30-fw</td>
<td>GCTGGCTCCATTGTCCAAAC</td>
<td>Medtr7g082810</td>
</tr>
<tr>
<td>RT-MATE30-rv</td>
<td>TGCTCTGGTCGTGTGATCTG</td>
<td></td>
</tr>
<tr>
<td>MtEF1a(RT)-fw</td>
<td>TGACAGGCAGCTGTGTAAGG</td>
<td>Medtr4g008860 (Zhang et al., 2010)</td>
</tr>
<tr>
<td>MtEF1a(RT)-fw</td>
<td>TCAGCGAAGGTCTCAACCAC</td>
<td></td>
</tr>
<tr>
<td>MtUbi(RT)-fw</td>
<td>GCAGATAGACACGCTGGGA</td>
<td>Medtr3g091400 (Kakar et al., 2008)</td>
</tr>
<tr>
<td>MtUbi(RT)-rv</td>
<td>AACTCTTGGGCAGGCAATAA</td>
<td></td>
</tr>
<tr>
<td>MtMSC27-fw</td>
<td>GTTGAAGTAGACATTTGGTGAAC</td>
<td>Medtr2g436620</td>
</tr>
<tr>
<td>MtMSC27-rv</td>
<td>AAAGCAGGTTATCTGAAATCT</td>
<td></td>
</tr>
<tr>
<td>RT-LEG-fw</td>
<td>GCGCCTTACAGGTGTGTAAGAG</td>
<td>Medtr5g081000</td>
</tr>
<tr>
<td>RT-LEG-rv</td>
<td>CAGTGCTCAATTTCTCTGCTCC</td>
<td></td>
</tr>
</tbody>
</table>

**RNAi and in situ hybridization**

<table>
<thead>
<tr>
<th>RNAi and in situ hybridization</th>
<th>Sequence</th>
<th>Gene ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>MtMATE30-RNAi-fw</td>
<td>CACCGTGATGATGTTGGACATCTTAATC</td>
<td></td>
</tr>
<tr>
<td>MtMATE3030-RNAi-rv</td>
<td>CAGGTATTAGCCAAAGTGCAATG</td>
<td></td>
</tr>
<tr>
<td><strong>Promoter cloning</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>------------------</td>
<td></td>
</tr>
<tr>
<td>proMtMATE30-fw</td>
<td><strong>CACC</strong>CAAGCTATTGAAGTAGA CGATC</td>
<td></td>
</tr>
<tr>
<td>ProMtMATE30-rv</td>
<td>GGTTCATTCAAGTATCATTTAGCT TGATC</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>CRISPR/Cas9 cloning</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MATE30cas-g1-fw</td>
<td>ATTGCTTGATAGCATCTTTGAAA</td>
</tr>
<tr>
<td>MATE30cas-g1-rv</td>
<td>AAACTTTCAAGATGCTATACAAG</td>
</tr>
<tr>
<td>MATE30cas-g2-fw</td>
<td>ATTGGTTTGTCATTTCACACTTG</td>
</tr>
<tr>
<td>MATE30cas-g2-rv</td>
<td>AAACACAAGTGTGAATGGACAAAA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Subcellular localization or bacterial complementation</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MtMATE30-fw</td>
<td><strong>CACC</strong>GAAGTAAAAAGTCAAGAAGAACAAGA</td>
</tr>
<tr>
<td>MtMATE30-rv1</td>
<td>CATCAATGTTTTTCTTTAATAGACAA CCTATGAAGAACC</td>
</tr>
<tr>
<td>MtMATE30-rv2-no_stop</td>
<td>TGCTCTGGCTGTGATCTGTAC</td>
</tr>
</tbody>
</table>
Chapter 4

An interspecific genomic analysis of the SLAC gene family between *Arabidopsis thaliana* and *Medicago truncatula*

Summary

In contrast to potential-driven secondary transporters, the knowledge of plant anion channels at the molecular level is still in its infancy. Notwithstanding, understanding of the fundamental roles anion channels play in plant physiology has largely increased in the last decade, including their roles in osmoregulation, turgor regulation, signal cascades, metabolite compartmentalization, and abiotic tolerance. Slow-type (S-type) anion channels (SLAC) were identified in *A. thaliana* and composed a small plant-specific gene family of five members, each functionally characterized molecularly and physiologically except one. The founding member is the guard-cell specific SLAC1 (Slow Anion Channel associated protein 1) participates in the regulation of cell turgor in leaves and associates with CO₂ and abscisic acid (ABA)-dependent stomatal closure. Although much has been published recently on functional roles of SLAC channels in *A. thaliana*, little has been explored in other plant species. In our research, we identified seven SLAC members in the model legume *M. truncatula*. Furthermore, we examined their expression patterns and analyzed gene evolution and phylogenetic relations between *A. thaliana* and *M. truncatula* in order to propose their possible physiological roles. By comparing the SLAC family in model species of the plant lineage, this study provides new insights into understanding the evolution of SLAC channels. This study also contributes to a more complete picture of the functional roles that SLAC channels play in the Rosids clade and helps us propose the possible anion selectivity or the gating mechanisms in legumes.

**Keywords:** gene evolution, transmembrane transporters, nodulation, symbiotic nitrogen fixation, Tellurite-resistance/Dicarboxylate transporter (TDT) family
Introduction

In order to understand anion transport through channels in legumes, and especially in nodule cells, we conducted a comparative study of the SLAC anion channel family in the Rosids clade using the model species *Medicago truncatula* and *Arabidopsis thaliana*. First, we reviewed the role of anion channels in the physiology of plants and what is known about the roles of SLAC channels in *A. thaliana*. Following, we assessed the SLAC family in *M. truncatula* to assign putative roles for them in legume physiology in order to construct testable hypotheses for future research aiming at functional analysis of their molecular functions.

Overview of membrane channels in plant physiology

The physiology of plants makes use of several anions, such as inorganic nitrate, chloride, sulfate, and phosphate as well as small, negatively charged organic molecules (e.g., malate, oxalate, and citrate) and play important roles in electrochemical exchanges within and between cells. These anions move through transporters and when coordinated with cation fluxes, are key players of cell turgor regulation. Indeed, channels allow the transport of anions through selective pores powered by the electrochemical motive force (downhill movement, or facilitated diffusion) without necessarily inducing protein conformational changes and serve wide physiological functions in plants, such as osmo- and turgor regulation, signaling cascades, metabolite compartmentalization, nutrition, and metal tolerance (Roberts, 2006; Barbier-Bryggo et al., 2011; Hedrich, 2012; Zhou et al., 2014). For instance, the opening of an anion channel in the membrane leads to depolarization of the electrochemical potential, and consequently can induce signaling events or the activation of voltage-gated ion channels (Vahisalu et al., 2008; Vahisalu et al., 2010; Drechsler et al., 2015, Brandt et al., 2012, Zheng et al., 2014; Qiu et al., 2016). Moreover, anion channels in roots are involved in anion excretion into the rhizosphere as well as loading into the xylem (Zheng et al., 2014; Jaborsky et al., 2016a; Qiu et al., 2016). Anion exudation into the rhizosphere regulates the uptake rate of several mineral nutrients, or chelates aluminum in acidic soil for detoxification (Magalhaes et al., 2007; Wang et al., 2017). Xylem loading of organic anions is essential for the translocation of metal cations that travel from roots to shoots as chelates with organic acids (Zimmermann and Sentenac, 1999; Furukawa et al., 2007; Yokosho et al., 2009). In legumes, dicarboxylic TCA acids
are secreted to symbiosomes of rhizobial-infected nodule cells to provide the intracellular nitrogen-fixing bacteria with a carbon source during symbiosis (Udvardi and Day, 1989; Schroeder, 1995; Jeong et al., 2004; Takanashi et al., 2013). The transporters mediating this exchange remain largely unknown (Benedito et al., 2010; Udvardi and Poole, 2013).

**Anion channels in plants**

Despite the immense importance anion channels play in plants, their investigation started at the molecular-genetic level just two decades ago (Schroeder, 1995; Ryan et al., 1997; Raman et al., 2005), leading to the discovery of their unique features, such as selectivity and gating mechanisms. Anion channels play specific roles based on their selectivity. For instance, chloride-selective channels may be involved in salt tolerance (Teakle and Tyerman, 2010; Qiu et al., 2016). Nitrate, sulfate, and phosphate-selective channels help carry nutrients, or play as necessary ligands for activation of anion channels, while organic-acid channels are important intermediates in carbohydrate metabolism and intracellular pH regulation (Zimmermann and Sentenac, 1999; Roberts, 2006; Zheng et al., 2014; Drechsler et al., 2015; Jaborsky et al., 2016a). The multiple carboxyl groups of organic acids also make them potent detoxification chelators of divalent and trivalent cations, such as aluminum and heavy metals (Liu et al., 2009; Liu et al., 2016). Beyond ion selectivity, gating is another important feature of anion channels. Many anion channels are regulated by an electromotive force or potential difference (voltage) created between the two faces of a biological membrane (Barbier-Bryggo et al., 2011; Hedrich, 2012; Roelfsema et al., 2012). Some channels are activated by membrane depolarization or hyperpolarization (Zheng et al., 2014; Jaborsky et al., 2016a) while others are regulated by intracellular signaling events, such as phosphorylation, or intra/extracellular metabolites (Vahisalu et al., 2010; Geiger et al., 2011; Brandt et al., 2012; Maierhofer et al., 2014a).

Based on early work on the activation kinetics of anion channel currents in response to voltage pulses through patch-clap experiments, there are two types of anion channels in guard cells: S-type (slow) and R-type (rapid) channels. The transition from S-type channels from open to closed state takes several seconds, while R-type channels only need a few milliseconds to activate (Roberts, 2006; Barbier-Bryggo et al., 2011; Kollist et al., 2011). Also, S-type channels show a weaker voltage dependency, which
means they activate over a much broader range of voltages than R-type channels (Roberts, 2006; Barbier-Brygoo et al., 2011; Kollist et al., 2011). The anion species and concentration of S-type channels are determined by permeation (selectivity) and gating mechanism.

**The SLOW ANION CHANNEL ASSOCIATED 1 (SLAC1) protein in A. thaliana**

The physiological role of the S-type anion channels has been elucidated in guard cells. SLAC1, the first and now well-characterized slow anion channel in *A. thaliana*, was identified from three independent mutant screens (*slac1-1, slac1-2 and slac1-3*) by two different labs. SLAC1 shares homology to the tellurite-resistance multidrug transporter TehA of *E. coli*, a member of the Tellurite-resistance/C₄-Dicarboxylate Transporter (TDT) family (Negi et al., 2008; Vahisalu et al., 2008; Vahisalu et al., 2010). SLAC mutant plants failed closing their stomata in respond to ozone, CO₂, drought stress, ABA or the air humidity with opening stomata (Negi et al., 2008; Vahisalu et al., 2008). In Arabidopsis, SLAC1 is expressed specifically in guard cells and is involved in stomatal movement. SLAC1 conducts selective anions, with greater permeability for nitrate than chloride and a much-reduced permeability for malate, bicarbonate, or sulfate (Negi et al., 2008; Vahisalu et al., 2008) without obvious anion-binding sites. The relative anion permeability sequence of SLAC1 was I⁻>NO₃⁻>Br⁻>Cl⁻ (Chen et al., 2010). This permeability sequence negatively correlates with the hydration energies of monovalent anions, since the lower the hydration energy, the greater the channel permeability (Chen et al., 2010; Vahisalu et al., 2010; Brandt et al., 2012).

Stomatal closure involves a complex mechanism to release ions (K⁺) from guard cells to the apoplast, which requires the coordinated efflux of anions induced by endogenous signals (ABA, Ca²⁺, NO) and environmental stimuli (CO₂, light intensity, humidity, ozone). ABA-induced activation of SLAC channels in the plasma membrane is of prime importance for stomatal closure. The dual regulation of SLAC1 currents by calcium-dependent and -independent phosphorylation is controlled directly through the ABA signaling pathway (Geiger et al., 2011; Brandt et al., 2012; Maierhofer et al., 2014a). SLAC1 can be activated by three classes of protein kinases: the calcium-independent protein kinase OPEN STOMATA 1 (OST1), calcium-dependent protein kinases (CPKs), and calcineurin B–like (CBL) calcium sensors with their interacting serine-threonine
protein kinases (CIPKs) (Vahisalu et al., 2010; Geiger et al., 2011; Maierhofer et al., 2014a). In the absence of ABA, ABI1 and ABI2 (ABA-insensitive proteins ABI1 and 2), a type 2C protein phosphatase (PP2C) directly interacts with protein kinases OST1, CPKs and CIPKs, rendering those kinases inactive. Upon reception of ABA by the cytosolic ABA receptors, PCAR/PYR/PCAR with ABI1/2 to form a ternary complex relieve those kinases from ABI1/2. Consequently OST1, CPKs, and CIPKs activate SLAC1 through phosphorylation. As a result, anions, mainly chloride and nitrate, escape the guard cells through SLAC1, which leads to membrane depolarization and triggers the activation of the guard-cell outward rectifying K\(^+\) channel (GORK) that directs potassium out of cells. Thereby, the guard cell loses the osmotic potential, driving water out of the guard cell. As a consequence, the guard cell loses turgor, finally leading to stomatal closure (Chen et al., 2010b; Brandt et al., 2012; Kollist et al., 2014). In addition to the activation of anion channels, ABA also inhibits plasma membrane H\(^+\)-ATPases, preventing stomata from reopening (Kollist et al., 2011). Interestingly, SLAC1 along with SLAH3 (Slow Anion Channel Homologue 3) inhibits the activity of inward-rectifying K\(^+\) channel KAT1 in A. thaliana by protein-protein interaction to prevent stomatal reopening under drought stress (Zhang et al., 2016). This direct coupling mechanism is independent of the functions of SLAC1 and SLAH3 as anion channels and is more energy-efficient than the convoluted ABA-SLAC multistep mechanisms described above. Therefore, guard cells have evolved more than one mechanism to fine-tune stomatal movements.

Therefore, ABA and SLAC channels work in concert in two signaling pathways to induce stomatal closure and prevent reopening. First, ABA and ABA receptor kinases lead to phosphorylation and activation of SLAC proteins, which induces anions and K\(^+\) efflux, leading to turgor loss and eventually stomata closure (Brandt et al., 2012; Kollist et al., 2014). Secondly, KAT1 activity is inhibited by SLAC1 and SLAH3 in A. thaliana by the protein-protein interaction (Zhang et al., 2016). ABA triggers the up-regulation of SLAC1 and SLAH3 expression, and these proteins both are to be activated by protein kinases to accelerate and facilitate anion efflux or used to inhibit KAT1 activity by protein-protein interaction, further leading to stomatal closure (Zhang et al., 2016). These steps provide a basic and fundamental mode of activation of SLAC1/SLAH3 via Ca\(^{2+}\)-dependent or -independent regulation of stomatal closure. We also know that direct phosphorylation of SLAC transmembrane domains is critical, although the molecular determinants of anion selectivity and the multiple phosphorylation sites in SLAC1/SLAH3 channels are not yet
fully understood.

A 3-D molecular model of SLAC1 channel was predicted with high confidence based on the structure of TehA channel from the bacteria *Haemophilus influenzae*, excluding the soluble C-terminal domain, which is not found in bacteria (Chen et al., 2010b). The model revealed three critical features. First, SLAC1 is trimeric and composed of quasi-symmetrical subunits, which may account for a possible mode of phosphorylation activation at the interface between the subunits (Chen et al., 2010b; Kollist et al., 2011). Second, within the trimer, each subunit has ten transmembrane helices arranged from helical hairpin pairs to form an inner five-helix transmembrane structure through the membrane with a central pore. These independent pores have a constant diameter of about 5Å, although no distinct anion binding sites were found (Chen et al., 2010b). Third, the pore is blocked by the phenyl ring of a well-conserved phenylalanine residue (F450 in SLAC1) on 9th TMD located in the middle of the pore.

**Other SLAC channels in A. thaliana**

In addition to SLAC1, four other SLAC1 homologs (SLAH1-4) were identified in *A. thaliana*. SLAH1–3 are localized to the plasma membrane and display different expression patterns and account for slow anion channel activity in diverse cell types (Barbier-Brygoo et al., 2011), while SLAH4 has not been functionally characterized yet. SLAH1-2 are expressed in roots, and SLAH3 is expressed throughout the whole plant, but with a dominant expression in roots as well. However, complementation studies in *A. thaliana* revealed that only SLAH1 and SLAH3 can functionally complement SLAC1 when expressed in *slac1* guard cells (Negi et al., 2008). This indicates that SLAH2 may primarily permeate different anions and play a distinct function in roots.

SLAH3 has been characterized as a nitrate channel because of its permeability to NO$_3^-$ in *Xenopus* oocytes is 20-fold higher than that of SLAC1 (Geiger et al., 2011). Although SLAC1 and SLAH3 are both expressed in guard cells, and activated by two types of calcium-sensitive kinases (CPK21/23, CIPK23/CBL1) associated with ABA signaling (Maierhofer et al., 2014a) involving guard cells closure, *slah3* mutants did not show any stomatal phenotype. In contrast to the leakage current-voltage curve (i.e., weak voltage dependency) of SLAC1, SLAH3 is activated with a positive voltage (i.e., voltage-dependent, toward positive potential) and modulated by a presence of extracellular NO$_3^-$.
concentration, therefore facilitating SLAH3 activation by shifting its activation threshold towards the guard cell resting membrane potential (Geiger et al., 2011; Hedrich, 2012). Thus, NO$_3^-$ acts as both a gate opener as well as the major substrate of SLAH3 (Geiger et al., 2011; Zheng et al., 2014). Additional to guard cells, SLAH3 also exhibits high expression in roots, while of SLAC1 is restricted in guard cells. These differences between SLAC1 and SLAH3 lead the two S-type anions channels to play distinct roles within different tissues under different stimuli with minimum overlap. SLAC1 is primarily involved in the release of Cl$^-$ and NO$_3^-$ from guard cells during stomatal closure, whereas SLAH3 mostly extrudes NO$_3^-$ and plays a role in the nitrate-dependent alleviation of ammonium toxicity in the plant (Zheng et al., 2014), as well as in regulating the NO$_3^-$/Cl$^-$ ratio in roots in concert with SLAH1 during salt stress.

SLAH1 is expressed in the xylem-pole pericycle of the root and modulates shoot accumulation of Cl$^-$ involving in salt tolerance. SLAH1 knockdown mutants exhibit the reduction of shoot Cl$^-$ accumulation without significantly affecting the permeability of NO$_3^-$ (Qiu et al., 2016). Interestingly, it is co-localized with SLAH3 in the plasma membrane of root cells. Salt stress remarkably reduces the expression of SLAH1 (60X) but has slightly affect on SLAH3 (5X). Although SLAH3 is still able to deliver NO$_3^-$ from roots to shoots, the transport of Cl$^-$ is compromised (Cubero-Font et al., 2016). Interesting, SLAH1 able to activate SLAH3 without kinases and the presence of extracellular NO$_3^-$ by forming SLAH3/SLAH1 heteromer, that shows a vast increase of Cl$^-$ conductance (Cubero-Font et al., 2016). Therefore, SLAH1 and SLAH3 act in tandem by heteromerization to regulate NO$_3^-$/Cl$^-$ ratio in roots under salinity (Cubero-Font et al., 2016; Qiu et al., 2016).

SLAH2 has been identified as a nitrate-specific channel, which is activated by CPK21/23 protein kinases (Maierhofer et al., 2014b). Like SLAH3, it requires nitrate in the extracellular bath solution for ion conduction. Since Cl$^-$ is the primary anion that counterbalances K$^+$ influx during stomatal opening, SLAH2 was not able to rescue the slac1-2 mutant under the SLAC1 promoter because of its impermeability to chloride (Schmidt and Schroeder, 1994; Zhang et al., 2016). The physiological function of SLAH2 has not been elaborated yet. Structure-guided site-directed mutation analysis of SLAH2 showed that the specificity of SLAC anion channels is resolved by the presence and constellation of the side chain polarity located on the alpha helix 3, which acts together with two pore-occluding phenylalanine residues (Maierhofer et al., 2014).
SLAC homologues in other organisms

In addition to plants, remote homologs of SLAC channels also appear in bacteria, archaea, and fungi. The prokaryotic homologs only contain the conserved transmembrane SLAC domain, while some fungi homologs also have the N- and C-terminal extensions. Mae1 from the fission yeast *Schizosaccharomyces pombe* works as a malate uptake transporter (Grobler et al., 1995). Ssu1 and Ssu2 from *Saccharomyces cerevisiae* are characterized as sulfite, and nitrate, and nitrite efflux pumps, respectively (Park and Bakalinsky, 2000; Cabrera et al., 2014). TehA from the bacterium *Escherichia coli* is an anion channel involved in tellurite resistance and a component of the TehA/TehB operon (Turner et al., 1997)(Chen et al., 2010b). So far, most microbial SLAC homologs have been confirmed to be permeated to malate, sulfite, nitrate, and nitrite, while plant SLAC channels and TehA from *Haemophilus influenza* are highly permeable to nitrate and/or chloride, but not malate. Importantly, the roles of anions in plant physiology are quite distinct to those in animals, implying that the original functions of SLAC channels probably account for a novel gene family in plants (Ward et al., 2009). A study even suggested plant SLAC channels might have an (unknown) evolutionary origin that differs from the microbial SLAC proteins (Dreyer et al., 2012).

Regarding the evolution of the SLAC channel family in plants, it is important to understand how it evolved in the Rosid clade. We understood that a genome-wide comparative approach between Arabidopsis and the model legume, *Medicago truncatula* (barrel medic), which shared a common last ancestor circa 90 million years ago (Mudge et al., 2005). As a legume, *M. truncatula* is an ideal genetic/genomic model for important crops (e.g., soybean, beans, peas, peanut, cover crops) and show important traits not present in *A. thaliana*, such as symbiotic root interactions (mycorrhization and nodulation for nitrogen fixation) and tolerance to drought and salinity (Keller et al., 1989; Zhao and Dixon, 2009; Adnane et al., 2015).

Our study identified seven SLAC genes in the *M. truncatula* genome. We explored their chromosomal distribution, gene structures, and expression patterns along with five SLAC channels in *A. thaliana*. By comparing the sequences of the SLAC family between *A. thaliana* and *M. truncatula*, possible anion selectivity or gating mechanisms of SLAC channels in *M. truncatula* can be proposed, thus helping us explore their functions. We
further analyzed gene duplication in the family and phylogenetic relations among ortho- and paralogs. This study also provides new insights into understanding how SLAC channels evolved in plants and an overall picture of the functional roles that SLAC transporters may play in symbiotic nitrogen fixation.

Results

Identification of SLAC channels in the Rosid model genomes, Arabidopsis thaliana and Medicago truncatula

A. thaliana and M. truncatula shared the whole–genome hexaploidization (pre-rosid WGH) event that split the Rosid-Asterid clade approximately 140-150 Mya (Young et al., 2011). Following, they split into different Rosid lineages around 92 Mya. A comparative analysis of the SLAC genes between the two species can offer an insight of how this family evolved in the Rosids clade. We identified seven SLAC members in the M. truncatula genome and five in A. thaliana (Negi et al., 2008; Vahisalu et al., 2008). By comparing the general properties of SLAC proteins in both species (Table 1), we found that they have similar ranges of protein length, similar physicochemical properties and number of transmembrane domains (TMDs). These general properties fit the TDT family description, which indicates that the SLAC family in Rosids are highly conserved. Also, the wide range of SLAC properties shown within each species suggests the ancestor gene(s) might have evolved to play different functions in order to meet special physiological needs of plants. The five genes encoding the SLAC channels in A. thaliana were unevenly distributed on three chromosomes: Chromosome 1 with three genes, one on chromosome 4, and another on chromosome 5. Likewise, in M. truncatula, the SLAC genes were mainly located on three chromosomes: Chromosome 4 (three genes) as well as chromosomes 6 (two genes) and 2 (one gene). Additionally, one channel remains unmapped and has been artificially located on a scaffold (Medtr0549s0020).
Table 1. Comparison of general properties of SLAC family in two species

<table>
<thead>
<tr>
<th>Parameters analyzed</th>
<th>A. thaliana</th>
<th>M. truncatula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of SLAC genes in the genome</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Range of protein sequence lengths (aa)</td>
<td>385-635</td>
<td>368-614</td>
</tr>
<tr>
<td>Range of number of transmembrane domains</td>
<td>9-10</td>
<td>9-10</td>
</tr>
<tr>
<td>Range of molecular weights (kDa)</td>
<td>41.58-72.34</td>
<td>41.83-69.50</td>
</tr>
<tr>
<td>Range of pI values (isoelectric point)</td>
<td>8.53-9.36</td>
<td>7.69-9.74</td>
</tr>
</tbody>
</table>

Phylogeny and structure analyses of plant SLAC genes

To further understand the evolution of the SLAC family in plants, a phylogenetic analysis was performed. According to our phylogenetic reconstruction including \( M. \) truncatula and previously studied proteins of this family (Figure 1), SLAC channels can be grouped into three major clades. Clade I contains all the plant members plus four proteobacterial TehA proteins. The Haemophilus influenzae TehA is an anion channel involved in tellurite resistance. It shows chloride permeability when its gating phenylalanine is removed (Chen et al., 2010b). The plant SLAC channels can be further classified into three sub-groups: SLAH1/4, SLAH2/3, and SLAC1 branches. The guard cell-specific AtSLAC1 with selectivity for nitrate and chlorite clustered together with Medtr4g101880 from M. truncatula. Indeed, our promoter-GUS analysis revealed that Medtr4g101880 is not only expressed in leaves, but also that it specific to guard cells (Supplemental Figure S3). Based on phylogeny, amino acid conservation, and expression pattern, Medtr4g101880 seems to play the exact same physiological role as described for SLAC1 in Arabidopsis.

In the SLAH1/4 branch, SLAH1 regulates chloride contents in roots while its homologue SLAH3 is induced by salinity stress (Cubero-Font et al., 2016; Qiu et al., 2016). SLAH4 has not been biochemically characterized yet. Medtr4g049640, Medtr3g112530 and Medtr4g132400 are grouped with SLAH1/4. On the other hand, SLAH2 and SLAH3 are considered nitrate channels and are grouped together with three genes of \( M. \) truncatula (Medtr0549s0020, Medtr6g045200 and Medtr6g034030).
Clade II includes proteins from fungi and archaea: Ssu1 and Ssu2 from *Pichia angusta* and Mae1 from *Schizosaccharomyces pombe*. Clade III has three uncharacterized TDT homologs from prokaryotes. As shown in Figure 1, SLAC members in both Rosid species grouped with each other perfectly, suggesting that the SLAC subfamilies existed before their lineages split. On the other hand, the absence of plant SLAC proteins in Clade II, while this clade contains only fungal and archaeal proteins is very puzzling regarding the origins and evolution of this family in prokaryotes and the early eukaryote lineage.

In order to increase the resolution of our analysis and shed some light on evolution of the SLAC family in the plant lineage, a more comprehensive phylogenetic analysis was conducted by using SLAC proteins identified from green algae (*Chlamydomonas reinhardtii*), a bryophyte (*Physcomitrella patens*), a monocot (*Oryza sativa*), and two additional dicots (*Solanum lycopersicum* and *Populus trichocarpa*) (Supplemental Figure S2). We found only one SLAC channel in algae, three in *P. patens*, five in the *P. trichocarpa*, seven in tomato, nine in rice. However, four of SLAC channels in *P. trichocarpa* have only 2-5 TMDs, and the other one has 8 TMDs without the conserved gating phenylalanine. Thus, SLAC orthologs remain undefined in *P. trichocarpa*. One SLAC channels in rice, one in *P. patens*, and two in tomato possess less than nine transmembrane domains and were excluded in the analysis. Similar to the previous phylogeny, this tree revealed the same three clades: plant, TehA, and Mae1/Ssu1 clades. The plant clade confirmed the three subgroups: SLAH1/4, SLAC1 and SLAH2/3. SLAC1/4 is the oldest subgroup and may have existed before the appearance of bryophytes, since there is only one gene in the *P. patens* genome (Pp3c14_15640), located as an outgroup of this clade. This gene may represent the founding member of the plant clade. Interestingly, SLAC1 evolution follows the plant speciation history. It has been proposed that SLAC1 channels appear in gymnosperms and then separated in the angiosperm lineages (Hedrich and Geiger, 2017). As mentioned above, we found one SLAC gene (Cre13.9584200) in green algae, which was grouped with bacterial TehA. Indeed, by mining plant genome databases, Dreyer found that TDT members in algae were not orthologs to clade I (SLAC-like), but rather showed closer similarity to the fungal malate permease Mae1 (Dreyer et al., 2012). Thus, SLAC channels may have originated from a single gene (unknown), and further undergone gene duplication and neofunctionalization (Dreyer et al., 2012).
Figure 1. Phylogenetic analysis of SLAC family in the Medicago truncatula and Arabidopsis thaliana genomes. Five SLAC channels in A. thaliana and seven in M. truncatula plus all members of TDT family from the TCDB database were employed to construct multiple alignments with Clustal Omega (www.ebi.ac.uk/Tools/msa/clustalo), which was then used to construct a phylogenetic tree in MEGA6 with Maximum likelihood method (ML) with 1000 bootstraps based on the JTT matrix-based model. The tree was drawn to scale with branch lengths measured in substitutions per site. At: Arabidopsis thaliana; Mt: Medicago truncatula; Hi: Haemophilus influenza; E.coli: Escherichia coli; Sp: Schizosaccharomyces pombe; Sc: Saccharomyces cerevisiae; Pa: Pichia angusta; An: Aspergillus niger; Ab: Arthrodema benhamia; Af: Aspergillus fumigatus; Mj: Methanocaldococcus jannaschii; Sa: Sulfolobus acidocaldarius; Sc: Streptomyces coelicolor; Sp: Streptococcus pyogenes;Cb: Clostridium butyricum.

We further analyzed the amino acid identifies of SLAC channels since the evolutionary of divergence of the plant clade using dispersion of proteins as a proxy (Dreyer et al., 2012). As shown in Figure 2, amino acid identities are highly conserved within subgroup, at about 57% (48-84%) within SLAC1/4, 69%, within SLAC1 and 70%
(58-99%) within SLAH2/3 subgroups. In contrast, inter-group identities were much lower, ranging 32% to a maximum of 52%. Consistent with previous studies (Dreyer et al., 2012), SLAH1/4 is obviously the oldest SLAC subfamily whereas members in the SLAC1 and SLAH2/3 are derivatives. Indeed, the SLAH1/4 separated from the other branches before the appearance of bryophytes, some 470 Mya (Gensel, 2008; Shaw et al., 2011; Dreyer et al., 2012). Furthermore, a detailed phylogenetic analysis with 135 SLAC-like channels from plant species confirmed that the separation of the SLAH2/3 group from SLAC1 occurred in a subsequent event after the appearance of bryophytes, but before the occurrence of lycophytes (Dreyer et al., 2012). This evolutionary history has been preserved in the genome of *M. truncatula*, as seen in **Supplemental Figure S 1**.
Figure 2. Percent identity matrix of all SLAC members of two species. Matrix was created by Clustal Omega

Structural analysis of SLAC genes in the Rosids species under analysis

Gene structures can provide a basic evolutionary perspective of members of a gene family. To better understand the evolution of SLAC genes since speciation between *M. truncatula* and *A. thaliana*, their gene (exon-intron) structures were plotted along with their phylogenetic relationships (Figure 3). This analysis provides further details into the evolution of the SLAC family between both species under investigation. The similar exon number and sizes are clear among SLAC channels within the same subfamily, with a few exceptions. For both species, most genes in the SLAH2/3 clade have four exons, while one has five (SLAH3) in *A. thaliana* and one (Medtr6g045200) has six exons in *M. truncatula*. Similarly, each SLAC1 in both species shows three exons. In the SLAH1/4
clade, most genes contained two exons, with a single exception (Medtr3g112530, with three exons). For completeness, two proteobacterial genes coding for TehA anion channels were included and, as expected, lacked introns. The striking gene structure similarities between both plant species suggest SLAC channels are very well conserved, indicating a possible conservation of critical physiological functions. To date, all characterized SLAC channels in *A. thaliana* and a few other plant species revealed that they are permeable to nitrate and/or chloride with minimum functional overlap and under the control of different activation mechanisms (Jaborsky et al., 2016a; Sun et al., 2016; Muller et al., 2017). For instance, when SLAH1 and SLAH3 are down regulated by soil salinity, and SLAH2 mediates nitrate to possibly ensure nitrate loading of the xylem (Maierhofer et al., 2014b; Hedrich and Geiger, 2017). Also, SLAC1 and SLAH3 are both expressed in guard cells and involved in stomatal movement induced by ABA pathways by different kinase activation mechanisms under different circumstance (Zhang et al., 2016).

![Figure 3. Predicted SLAC gene structures of the SLAC members.](image)

We further analyzed channel properties of SLAC family. The most prominent structural difference between channels of the SLAH1/4 with the other two groups is their shorter proteins sequences (365-385) with much shorter cytosolic C- and N-termini (Figure 4). Indeed, the C-terminus of SLAH1/4 channels have only ~50 amino acid residues, while in the other two groups they are ~150 residues. In Arabidopsis, the N-
terminus was shown to contain important target residues of kinases that regulate SLAC1 and SLAH3 channel activities. Dreyer et al. (2012) used Netphos2 (Blom et al., 1999; Blom et al., 2004) to predict the conserved phosphorylation sites among the subgroup and found nine serine/threonine residues in channels of the SLAC1 group, six in the SLAH2/3 group and only one in the SLAH1/4 group (Dreyer et al., 2012). These residues are shown in red font in Supplemental Figure S1. Some amino acid residues on N-termini and the first transmembrane domain (TMD) were experimentally confirmed to be phosphorylated by the protein kinases OST1 (AtSLAC1: S59, S86, S120), CPK21 (AtSLAH3: T187) and CPK6 (AtSLAC1: S59) (Vahisalu et al., 2010; Brandt et al., 2012). Furthermore, Dreyer et al. (2012) also discovered one site on the C-terminus right after the last transmembrane domain of SLAC1 and SLAH2/3 (SLAC1: T513; and SLAH3: S580) (Dreyer et al., 2012). T513 (SLAC1) was identified as one of the targets of kinase OST1 (Vahisalu et al., 2010). Apparently, SLAH1/4 channels have too short termini and lack phosphorylation sites, which suggests they might have different activation mechanisms or internal phosphorylation sites within TMD linkers. Two phosphorylation sites were also predicted in the cytosolic linkers between the fourth and fifth TMD of SLAH1/4 channels and between the eighth and ninth TMD of SLAH2/3 (Dreyer et al., 2012). We found all phosphorylation sites conserved in M. truncatula SLAC channels among the subgroups, except for the SLAH1/4-like Medtr3g112530 gene product, which has a threonine (T187) instead of serine residue within the fourth TMD, which may be a site for PKC, based on its phosphorylation score (0.832) by using NetPhos 3.1 (www.cbs.dtu.dk/services/NetPhos). Additionally, our analysis revealed 19 other phosphorylation sites for Medtr3g112530 with score >0.6 (Dreyer et al., 2012), although they still are to be confirmed experimentally with phosphorylation site-specific antibodies.

Our study provides a valuable resource for future studies to understand the role and activation mechanisms of the different SLAC channels in M. truncatula. The conserved phosphorylation sites among subfamilies between these two Rosid model species indicate SLAC channels remained mostly unchanged in the last 100 Mya. The distinct phosphorylation sites among the channel subfamilies indicate they may require different kinases for activation and underlie possible functional variations between the subgroups. The fact that different phosphorylation sites existing in the same channel for the same kinase suggest each site might act distinctly based on their steric arrangements.
within the protein and with the membrane, which directly affect their accessibility to contribute to the regulation of the channel activities.
Figure 4. Predicted membrane topology of plant SLAC channels by TMHMM online program. Predicated transmembrane domains of SLAC channels in *M. truncatula* and *A. thaliana* sorted by subfamilies. Red lines on top of each prediction represent putative transmembrane domains. The blue line represents protein terminal tails on the cytoplasmic side of the membrane, while magenta represents tails outside of the membrane. Topology prediction was generated by the TMHMM program (Krogh et al., 2001).

**Duplication patterns of SLAC genes in plant genomes**

Arabidopsis and Medicago shared the whole-genome hexaploidy (pre-rosid WGH) event before the rodis-asterid split (Young et al., 2011). They split up into the Malvid and Fabid lineages, respectively, around 92 Mya (Blanc et al., 2003; Magallón and Castillo, 2009; Cannon, 2013), and underwent different duplication events. *M. truncatula* went through an additional genome duplication round circa 58-59 Mya (the so-called early legume-specific genome duplication), establishing the origins of the papilionoid lineage (Young et al., 2011). This last event had great influence on many genes families in agronomic papilionoid legumes generating paralogous duplicates, which since may have acquired distinct functions (Cannon, 2013). The Papilionoid (Faboideae) subfamily radiated into several clades, the largest of which split quickly at about 54 Mya into two subclades, the Hologalegina (or galegoids, including *M. truncatula* and *L. japonicus*) and the Milletioids (including *Glycine max* and other phaseloids). Furthermore, *M. truncatula* has undergone high rates of local gene duplication (*Medicago*-specific duplication) (Wojciechowski, 2003; Pfeil et al., 2005; Sprent and James, 2007; Young et al., 2011). Meanwhile, the *A. thaliana* lineage went through another round of genome duplication that occurred before the Arabidopsis/Brassica split (20-24 Mya) (Blanc et al., 2003; Ziolkowski et al., 2006). Thus, we focused on segmental and tandem duplication of SLAC genes in the genomes of both species to provide a more detailed understanding of the evolution of this gene family.

After the pre-rosid WGH event, most genes became substantially redundant in the genome of the immediate Rosid ancestor. Our phylogeny reconstruction (Figure 1) shows that all SLAC channels in both species perfectly group into three subfamilies, and that channels from the same species cluster together, indicating the SLAC gene family had already been defined into the three clades before the pre-rosid WGH event. Thus, we further investigated microsynteny relationships within the *M. truncatula* genome (Figure 5) as well as the shared synteny relationship between *A. thaliana* and *M. truncatula* (Figure 6). Our results revealed that many syntenic blocks were disrupted by chromosomal rearrangements. We found only one pair of a syntenic paralog of SLAC channel between both species (At1G12480 and Medtr4g101880 - Figure 6). Interestingly,
the only SLAC in the algae genome showed to be too remotely related to land plant SLAC channels (Supplemental Figure S2). On the other hand, the moss genome contains a SLAC gene at the root of the plant clade. The fact that tomato (Asterid) and the Rosids species poplar, Arabidopsis, and Medicago have SLAC members distributed in all three clades means that the functional diversification of SLAC channels in higher plants might have expanded before the Rosid-Asterid split event.

Microsynteny analysis helps us understand the evolution of a genome in a species. Our study identified one pair of tandem duplications in each species under study (Medtr6g034030 and Medtr6g045200; AT1G62280 and AT1G62262). Since these genes do not show conserved synteny between the genomes, we infer that they duplicated independently after the Malvid-Fabid split. Interesting, the duplicated pairs in both species belong to the same subfamily (SLAH2/3), which indicates that the duplication of the original gene contributed with better fitness in each species.
Figure 5. Microsynteny analysis within the *Medicago truncatula* genome. The analysis was carried out with SynMap software using coding sequence (JCVI genome sequence v.4.0) as input data and default parameters. Syntenic regions in the genome are presented as dots. The only one pair (Medtr6g034030.1 and Medtr6g045200.1) located in syntenic regions are highlighted with a red dot.

Figure 6. Synteny analysis between the *Medicago truncatula* and *Arabidopsis thaliana* genomes. The analysis was carried out with SynMap software using coding sequence (JCVI genome sequence v.4.0 and TAIR10) as input data and default parameters. Syntenic regions in the genome are presented as dots. The only one pair (Medtr4g101880.1 and AT1G12480.1) located in syntenic region is highlighted with a red dot.
Expression patterns of the SLAC genes in A. thaliana and M. truncatula

We carried out a comparative analysis of gene expression patterns within each the SLAC subfamilies to better understand how conserved their roles may be between both species, or the potential functional redundancies within a species genome, along with independent RNA-Seq data for SLAC genes of M. truncatula. The expression profiles (Figure 7A) showed that the SLAC orthologs are consistently expressed in both species, with a few exceptions. The SLAC1 genes in both species compose true paralogs with highest transcription activity in leaves, more specifically in guard cells (Supplemental Figure S3), indicating both genes most likely play the same function. In the SLAH1/4 subfamily, SLAH1, SLAH4, and Medtr4g132400 are exclusively expressed in roots, while Medtr3g112530 is expressed in roots and nodules, and Medtr4g049640 mostly in nodules. The three M. truncatula SLAH1/4 genes show an interesting evolutionary history. The Medtr3g112530, which is closest to the A. thaliana SLAH1/4 genes, has a conserved expression pattern, while the most distantly related gene (Medtr4g049640) has possibly gained a novel function in the nodule, an organ that does not exist in Arabidopsis. Interestingly, in the hierarchical clusterization of the samples analyzed, the aboveground tissues (leaf, vegetable bud, seedpod, flower) clustered together, while the belowground-expressed genes Medtr4g13240, which is exclusively expressed in roots, and Medtr4g049640, which is highly expressed in nodules, were placed at the edge of the tree (Figure 7B). Therefore, after gene duplication, Medtr4g049640 potentially underwent neofunctionalization to meet the distinctive symbiotic needs of legumes.

SLAH2/3 genes exhibit diverse expression patterns. SLAH3 is constitutively expressed in all tissues and especially in roots, whereas SLAH2 is exclusively expressed in roots. On the other hand, Medtr6g034030 showed a high level of transcription in several tissues of the mature plant, such as flowers, roots, and nodules and weak expression in seedpods and leaf (Figure 7B). Likewise, Medtr6g045200 displayed weak expression in leaf and vegetative bud. Both of these M. truncatula genes are duplicated in-tandem and code for proteins with 90% similarity (Figure 2), possibly being functionally redundant (Figure 5). The unmapped Medtr05490s0020 was expressed at low level in roots. It is perhaps undergoing pseudogenization or neofunctionalization with an expression pattern under special conditions. Apparently, some of SLAC channels have retained their function in both species even in the long evolution history, such as Medtr4g101880 expressed in guard cells that may be involved in stomatal movement, while others underwent
neofunctionalization to develop new functions for the special legume organ, nodules. SLAC channels are believed to conduct nitrate and/or chloride in plant cells, and interestingly all channels expressed in nodules were clustered in the SLAH1/SLAH4 subfamily and might efflux nitrate and/or chloride into nodules. Since nitrate is a strong negative regulator of symbiotic nitrogen fixation in legumes, SLAC channels might be involved in the autoregulation of nodulation (AON), thus a key component of mechanisms controlling nodule number and senescence. Therefore, drawing from the phylogeny reconstruction, synteny analysis, and gene expression profiles, while some SLAC channels might have retained their function after the Arabidopsis-Medicago split event, some channels evolved in distinct directions to meet specific demands in the *M. truncatula* lineage.

Figure 7. Expression profile of SLAC genes in the Rosids model species, *Arabidopsis thaliana* and *Medicago truncatula*. A. Gene expression represented in each organ of the plants growing under optimal conditions. Tissues with high expression above than average of *M. truncatula* genes are represented by red boxes, while expression of *A. thaliana* genes is indicated in purple. Faded color in boxes represent the weak expression, white boxes indicate no expression. Data of *M. truncatula* is from the Medicago Gene Atlas (Affymetrix microarray; (Benedito et al., 2010), and expression data for *A. thaliana* was retrieved from the Phytozome platform (phytozome.jgi.doe.gov/pz/portal.html). B. Heatmap of gene expression for *Medicago truncatula*. The graph was constructed using the gplots program in R with RNA-Seq data (Young et al., 2011). Lack of expression is indicated by white tiles; yellow indicates below-median values, orange indicates the median expression value, and above-median values are shown in gradations of red. Except for Medtr0549s0020, for which no expression value was detected, all other SLAC genes were expressed, covering every organ of the *M. truncatula* plant.

Regulatory gene networks centered on the SLAC1 channel and transcriptional master kinase

The structure analysis of HiTehA revealed that the internal five-helix transmembrane domains form a pore with a central, extremely conserved phenylalanine residue (F450 in
AtSLAC1) as the anion gate (Supplemental Figure S1). This conserved residue is regulated by the various kinases via phosphorylation of the N-terminal tail. The phosphorylation events that activate the gating phenylalanine in AtSLAC1 have been well studied. We could identify this phosphorylation site in MtSLAC1 (Supplemental Figure S1). So here we constructed a co-expression network of MtSLAC1. We used relevant publicly available global gene expression data for Medicago truncatula and correlated each annotated gene with MtSLAC1 expression values using non-parametric Spearman's rank-order correlation. All correlation values for expression values between each pair of other proteins and MtSLAC1 from a set of 99 biological samples were listed at Table 2.

Only 57 genes correlated MtSLAC1 expression profile at the conservative threshold of $p \geq 0.85$. No surprisingly, this set included proteins required for the basic cellular process, like DNA repair family protein, Calvin cycle protein, chlorophyllase enzyme, phototropin-2 protein). More interestingly that the expression of five ion transporters and channels (Medtr2g005410, Medtr1g009230, Medtr4g098570, Medtr8g464570, Medtr5g014310) are correlated with that of MtSLAC1. AtSLAC1 is functionally associated with regulators of stomatal closure under the different external stimuli by homeostatic regulation of ionic gradients such as nitrate, chloride, and malate. Since MtSLAC1 seems to be a true homolog of AtSLAC1, we sought to identify potential. Among the 57 genes in the MtSLAC1 network, nine kinases were strongly and positively corrected with MtSLAC1: Medtr8g098845 (MAP kinase), Medtr2g100030 (Ser/Thr kinase), Medtr1g067210 (2c2c domain Ser/Thr phosphatase), Medtr2g103810 (G-type lectin S-receptor-like Ser/Thr kinase SD2-5), Medtr4g063945 (Kinase 1B), Medtr5g025950 (LRR kinase) and Medtr4g091840 (G-type lectin S-receptor-like Ser/Thr kinase). Interestingly, the phosphatase encoded in the Medtr1g067210 locus belongs to PP2C family and its involvement in drought and cold stress response has been reported in Medicago truncatula (Yang et al., 2018).

Table 2 List of all correlation values for expression values between each pair of other proteins and MtSLAC1.

<table>
<thead>
<tr>
<th>Target gene</th>
<th>Gene ID</th>
<th>Annotation</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MtSLAC1</td>
<td>Medtr8g098845</td>
<td>MAP kinase</td>
<td>0.952582187</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr8g072950</td>
<td>Plant intracellular ras group-related LRR protein</td>
<td>0.95161334</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr2g100030</td>
<td>Serine/Threonine kinase family protein</td>
<td>0.950777919</td>
</tr>
<tr>
<td>---------</td>
<td>---------------</td>
<td>----------------------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr8g013990</td>
<td>DUF1677 family protein</td>
<td>0.940636872</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr8g068210</td>
<td>Dof domain zinc finger protein</td>
<td>0.934698431</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr1g067210</td>
<td>Serine/threonine phosphatase family%2C 2C domain protein</td>
<td>0.933096601</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr2g437590</td>
<td>Triose-phosphate transporter family protein</td>
<td>0.927900466</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr2g101140</td>
<td>DUF1677 family protein</td>
<td>0.926821039</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr5g009490</td>
<td>Plant/K16L22-9 protein%2C putative</td>
<td>0.921040694</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr1g108390</td>
<td>Auxin efflux carrier family protein</td>
<td>0.919909174</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr0194s0050</td>
<td>Transcription factor</td>
<td>0.919435377</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr5g095880</td>
<td>HhH-GPD base excision DNA repair family protein</td>
<td>0.916065699</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr7g087130</td>
<td>myb transcription factor</td>
<td>0.908269379</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr2g005410</td>
<td>Chloride channel CIC1 protein</td>
<td>0.907372539</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr1g009230</td>
<td>Extracellular ligand-gated ion channel protein</td>
<td>0.9063999431</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr7g056090</td>
<td>Xyloglucan endotransglucosylase/hydrolase family protein</td>
<td>0.900890358</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr7g108790</td>
<td>Plant invertase/pectin methylesterase inhibitor</td>
<td>0.900012976</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr6g027560</td>
<td>Plasma-membrane choline transporter family protein</td>
<td>0.896264081</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr4g068240</td>
<td>orn/lys/arg decarboxylase major region protein</td>
<td>0.895215437</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr1g064240</td>
<td>Cyclic nucleotide-gated ion channel-like protein</td>
<td>0.88932338</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr4g057830</td>
<td>Hypothetical protein</td>
<td>0.889099861</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr4g039540</td>
<td>Hypothetical protein</td>
<td>0.887068468</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr4g098570</td>
<td>Aluminum activated malate transporter family protein</td>
<td>0.884343383</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr6g051780</td>
<td>Transmembrane protein%2C putative</td>
<td>0.883147713</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr4g095660</td>
<td>Acetamidase/formamidase%2C putative</td>
<td>0.880388764</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr1g112240</td>
<td>PPR containing plant-like protein</td>
<td>0.876903246</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr6g022160</td>
<td>Transmembrane protein%2C putative</td>
<td>0.8743242</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr4g050000</td>
<td>Cellulose synthase H1-like protein</td>
<td>0.873875714</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr1g046250</td>
<td>Coenzyme F420-reducing hydrogenase%2C beta subunit</td>
<td>0.870591256</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr3g466200</td>
<td>Anionic peroxidase swpb3 protein</td>
<td>0.870550643</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr8g088560</td>
<td>UDP-glucosyltransferase family protein</td>
<td>0.869868288</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr3g463080</td>
<td>Sodium-dependent phosphate transporter%2C putative</td>
<td>0.869153947</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr8g461030</td>
<td>Methionine sulfoxide reductase B 2</td>
<td>0.869019703</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr4g014670</td>
<td>DUF688 family protein</td>
<td>0.86629506</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr6g042030</td>
<td>Transmembrane protein%2C putative</td>
<td>0.865344571</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr2g086590</td>
<td>DUF3049 family protein</td>
<td>0.864922828</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr8g464570</td>
<td>Cyclic nucleotide-gated ion channel-like protein</td>
<td>0.86356256</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr3g084450</td>
<td>Long-chain-fatty-acid CoA ligase (AMP-forming)</td>
<td>0.863315439</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr7g444860</td>
<td>Squamosa promoter-binding-like protein</td>
<td>0.863076498</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr5g014310</td>
<td>Aluminum activated malate transporter family protein</td>
<td>0.862677061</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr1g015410</td>
<td>Cytokinin oxidase/dehydrogenase-like protein</td>
<td>0.862149355</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr8g096320</td>
<td>Bidirectional sugar transporter</td>
<td>0.861351884</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr1g087030</td>
<td>Calvin cycle protein CP12-1</td>
<td>0.859076594</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr2g103810</td>
<td>G-type lectin S-receptor-like Serine/Threonine-kinase SD2-5</td>
<td>0.858710492</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr1g079870</td>
<td>Hypothetical protein</td>
<td>0.857276494</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr1g041390</td>
<td>Chlorophyllase enzyme</td>
<td>0.856166958</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr4g063945</td>
<td>Kinase 1B</td>
<td>0.855616487</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr2g097230</td>
<td>PGR5-like protein 1A%2C putative</td>
<td>0.855104606</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr4g088515</td>
<td>DnaK family protein</td>
<td>0.854552754</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr5g025950</td>
<td>LRR receptor-like kinase</td>
<td>0.854261078</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr1g088540</td>
<td>Plant invertase/pectin methylesterase inhibitor</td>
<td>0.854015015</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr2g095980</td>
<td>Phototropin-2 protein</td>
<td>0.85382559</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr8g077390</td>
<td>MYB family transcription factor</td>
<td>0.853411181</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr4g091840</td>
<td>G-type lectin S-receptor-like Serine/Threonine-kinase</td>
<td>0.852607033</td>
</tr>
<tr>
<td>MtSLAC1</td>
<td>Medtr4g063400</td>
<td>Hypothetical protein</td>
<td>0.852560129</td>
</tr>
</tbody>
</table>
Conclusion

Previous studies have shown the importance and versatility of SLAC channels to the plant physiology. They are not restricted to efflux of anions in guard cells and carbon metabolism, but are also important in root cells where they may function in anion homeostasis, membrane stabilization, osmoregulation, boron tolerance, and regulation of passive salt loading into the xylem vessels (Negi et al., 2008; Vahisalu et al., 2008; Ward et al., 2009; Chen et al., 2010b; Vahisalu et al., 2010; Brandt et al., 2012; Hedrich, 2012; Maierhofer et al., 2014a; Zheng et al., 2014; Jaborsky et al., 2016a; Qiu et al., 2016). A comprehensive genome-wide analysis of SLAC family was performed in an important legume species, *M. truncatula*. Here, we identified seven SLAC channels in the *M. truncatula* gene, each with a specific expression pattern. By comparing them to the five SLAC channels of *A. thaliana*, they could be classified into three subfamilies: SLAC1, SLAH1/4, and SLAH2/3. The channels displayed considerable variation in gene structure, and tissue expression patterns among the different subfamilies, which suggests functional diversification, while great conservation within the subfamilies. Segmental and whole genome duplication contributed to the expansion of SLAC family in the Rosid lineage. Our genome-wide identification and further functional characterization of SLAC family in *M. truncatula* will facilitate a systematic functional analysis of SLAC channels in legumes, and especially the unique roles they play in symbiotic nitrogen fixation. We showed here that three SLAC genes from SLAH1/4 subfamily are expressed in nodules, with Medtr4g049640 showing the highest transcription activity. Therefore, we propose that Medtr4g049640 may be playing a role in the nodule symbiosis. Since SLAC channels were implicated in the transport of various anions, including nitrate, and the fact that, in legumes, nitrate is a strong negative regulator of nodulation via largely unknown autoregulation mechanism, perhaps Medtr4g049640 plays a part in this inhibition. Altogether, this study provides a foundation to further investigate the functions of SLAC channels in legumes.
Material and Methods

Identification of SLAC channel genes in M. truncatula and other plant genomes

The whole protein dataset of M. truncatula (www.medicagogenome.org) was submitted to the TransportTP transporter prediction tool (http://bioinfo3.noble.org/transporter/) (Li et al., 2009) for identification and classification of membrane transporters into families based on the Transporter Classification system (TCDB, www.tcdb.org) (Saier et al., 2016; Saier et al., 2006). Briefly, the criteria for the SLAC family was previously described (Benedito et al., 2010) and the transporters were assessed with the typical features of the SLAC transporters in plants (e.g., 9-10 transmembrane domains, >300 amino acid residues, similarity to plant SLAC channels in curated as well as comprehensive databases), and assigned a functionality confidence level >=2 (1 for highest confidence the gene product is functional; 2 for a likely functional gene product, with a few features off or missing; and 3 for possibly pseudogenes and non-functional transporters). SLAC members from other plant species were searched in the Phytozome database containing the PF03595 conserved domain and then were used to predict transmembrane domains. Members with less than 9 TMDs were excluded from our analysis.

Phylogenetic analyses

Two phylogenetic trees were generated using protein sequences by the Maximum Likelihood (ML) algorithm with 1,000 bootstraps based on a JTT matrix-based model in MEGA v.6.0 (Tamura et al., 2013), after protein sequences alignment performed with ClustalW (www.ebi.ac.uk/clustalw). The first phylogeny was generated with 25 full-length SLAC channels from M. truncatula (7 channels), A. thaliana (5), and all listed SLAC proteins (13) from TCDB (family 2.A.16.5). The second analysis included four plant species, totaling 43 sequences. For the initial tree to be obtained, the Neighbor-Joining method was applied to a matrix of pairwise distances estimated using a JTT model. The trees were drawn to scale, with branch lengths measured in number of substitutions per site.

Synteny analysis
Synteny analysis was carried out using the SynMap tool (http://genomevolution.com/CoGe/SynMap.pl) based on coding sequences (JCVI *Medicago truncatula* genome sequence v.4.0: www.medicagogenome.org) as input data and default parameters (Zheng et al., 2015). The in-tandem gene duplicates in a genome (defined here as similar genes at most ten genes apart with a threshold e-value ≤10e-4) were determined using the Blast2raw script in the CoGe toolkit (Haug-Baltzell et al., 2017). Syntenic blocks (microsynteny within a genome or shared synteny between two genomes) were identified with SynMap via the DAGchainer algorithm using the relative gene order option with a minimum of 5 gene pairs aligned and a maximum distance of 20 genes between two matches. QuotaAlign was used to merge adjacent syntenic blocks (Zheng et al., 2015).

**Promoter analysis**

A 1771 bp fragment upstream the start codon of Medtr4g101880 (the SLAC1 ortholog) was cloned to drive GUS expression (proMtSLAC1::GUS) in the pKGWFS7 binary vector (**Supplemental Figure S4**, **Supplemental Table S1**). Transformation was performed using *Agrobacterium tumefaciens* strain EHA105 containing the binary vector. Young, well-developed leaflets of in vitro of *A. thaliana* ecotype Col-0 were used as explants for *Agrobacterium*-mediated transformation as described previously (Beaujean et al., 1998). Transgenic plants were selected and grown on kanamycin MS medium in magenta boxes. Leaves from 30-day old plants were collected and stained with X-Gluc (5-bromo-4chloro-3-indolyl-β-D-glucuronic acid, cyclohexyl ammonium salt) according to standard methods (Boivin et al., 1990; Veereshlingam et al., 2004). Briefly, leaves were vacuum infiltrated for 30 seconds with the fixing solution (0.3% p-formaldehyde, ten mM MES (2-(N-morpholino)-ethanesulfonic acid) pH 5.6; 0.3 M mannitol). Samples were left at room temperature for 1 hour followed by washing steps. The fixing solution was subsequently replaced by a washing solution (0.5 M MES and 0.3 M mannitol). Samples were rinsed three times for 20 min with moderate rotation movement, then incubated in pre-warmed (37°C) staining solution (0.08% X-Gluc, 50 mM potassium ferricyanide, 50 mM potassium ferrocyanide, and 0.1 M PIPES (pH 7.2)) for 16h at 37 C. Samples were rinsed with 0.1 M PIPES (pH 7.2), cleared with 70% ethanol for 30 mins to remove chlorophyll. The images were taken under MIF Olympus MVX10, a Dual upright microscope, and stereoscope (Olympus, Melville, NY).
Gene expression association analyses
Raw RNA-seq data (96 experiments) was retrieved from the SRA database at NCBI (https://www.ncbi.nlm.nih.gov/sra) with the following filters: Platform: Illumina; Assay type: RNA-seq; Organism: Medicago truncatula; Tissues: flower, leaf, leaf and root, leaves, leaves and stem, mature leaves, seeding, the whole plant, or young leaves. The expression analysis used the following programs in the Galaxy platform (https://usegalaxy.org): reads went through FastQC (FastQC v0.10.1 )for the quality control, and then trimmed with Trimmmomatic (v. 0.36.5), and the alignment and transcript quantification were carried out with Salmon (Galaxy version 0.9.1) using the Medicago truncatula genome version 4.0 as reference. Expression correlations between MtSLAC1 and each gene annotated in the genome were calculated using Spearman’s rank correlation coefficient in Microsoft Excel for Mac as previously described (Santos et al., 2017) (Santos et al., 2017).

References

Anjam MS, Ludwig Y, Hochholdinger F, Miyaura C, Inada M, Siddique S, Grundler
An improved procedure for isolation of high-quality RNA from nematode-infected Arabidopsis roots through laser capture microdissection. Plant Methods 12: 25


Anion Channel Subunit SLAH1 Gates SLAH3 Open for Chloride Root-to-Shoot Translocation. Curr Biol 26: 2213–2220


FAO (2016) Food and Agriculture Organization of the United Nations: Current world
fertilizer trends and outlook to 2016.


R973–R975


**Hedrich R** (2012) Ion Channels in Plants. Physiol Rev **92**: 1777–1811


enhanced transcriptome in soybean. Mol Cells 18: 53–62


Schmidt C, Schroeder JL (1994) Anion Selectivity of Slow Anion Channels in the


Sun SJ, Qi GN, Gao QF, Wang HQ, Yao FY, Hussain J, Wang YF (2016) Protein kinase OsSAPK8 functions as an essential activator of S-type anion channel OsSLAC1, which is nitrate-selective in rice. Planta 243: 489–500

Sun X, Gilroy EM, Chini A, Nurmberg PL, Hein I, Lacomme C, Birch PRJ, Hussain


Veereshlingam H, Haynes JG, Penmetsa RV, Cook DR, Sherrier DJ, Dickstein R (2004) nip, a symbiotic Medicago truncatula mutant that forms root nodules with


**Wojciechowski MF** (2003) Reconstructing the phylogeny of legumes (Leguminosae):


Young ND, Bharti AK (2012) Genome-enabled insights into legume biology. Annu Rev
Plant Biol 63: 283–305


Zhao M, Chen Y, Qu D, Qu H (2011b) TSdb: a database of transporter substrates linking metabolic pathways and transporter systems on a genome scale via their shared substrates. Sci China Life Sci 54: 60–4


Supplementary Figure S1. Alignment of SLAC channels in *A. thaliana* and *M. truncatula*. Potential conserved phosphorylation sites within subgroups are showed as red font. Transmembrane domains (TM) are indicated (in reference to the Arabidopsis SLAC1 sequence), as well as phenylalanine residue with the highlight box.
Supplementary Figure S2. Phylogenetic analysis of SLAC family in the *Medicago truncatula*, *Arabidopsis thaliana* plus other four plant species. The phylogenetic tree was generated with protein sequences by Maximum Likelihood (ML) algorithm with 1,000 bootstraps based on JTT matrix-based model in MEGA v.6.0. The tree was drawn to scale with branch lengths measured in some substitutions per site. At: *Arabidopsis thaliana*; Mt: *Medicago truncatula*; Hi: *Haemophilus influenzae*; Sl: *Solanum lycopersicum*; Os: *Oryza sativa*; Pp: *Physcomitrella patens*; Cre: *Chlamydomonas reinhardtii*; E.coli: *Escherichia coli*; Sp: *Schizosaccharomyces pombe*; Sc: *Saccharomyces cerevisiae*; Pa: *Pichia angusta*; An: *Aspergillus niger*; Ab: *Arthrodema benhamiae*; Af: *Aspergillus fumigatus*; Mj: *Methanocaldococcus jannaschii*; Sa: *Sulfolobus acidocaldarius*; Sc: *Streptomyces coelicolor*; Sp: *Streptococcus pyogenes*; Cb: *Clostridium butyricum*. 
Supplementary Figure S3. Promoter activity of SLAC1 ortholog in *Medicago truncatula*. GUS staining of a leaf sample from an *Arabidopsis thaliana* T0 plant transformed with the proMedtr4g101880::GUS construct. (A) The whole leaf under a stereomicroscope. (B-C) Gene expression is strong and specific to guard cells. Bar in A = 200 µm; Bar in B and C = 20 µm.

Supplementary Figure S4. Binary vector map.

Supplemental Table S1 List of primers used in these studies.

<table>
<thead>
<tr>
<th>Promoter cloning</th>
<th>Primer Sequence</th>
<th>Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>pSLAC1-Gtwy-fw</td>
<td>CACCTAGTATGAGAGAGATAAATCATATTGTTACGC</td>
<td>Mtr4g101880.</td>
</tr>
<tr>
<td>pSLAC1-Gtwy-rv</td>
<td>TACTTTGCTGTCTATAATAGATGCTTAGCAAAAC</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 5

Characterization of the nodule-specific SLAH1 membrane transporter in *Medicago truncatula*

Summary

*Medicago truncatula* is a well-established model that allows the study of intrinsic features of the Fabaceae family. Remarkably, legumes are capable of carrying out nitrogen fixation through association with rhizobial bacteria. Nodules develop in infected roots that contain nitrogen-fixing bacteroids enclosed in cellular structures called symbiosomes. Symbiotic nitrogen fixation (SNF) is established through a nutritional exchange, including reduced carbon (dicarboxylates) from the plant for reduced nitrogen (ammonium) from endosymbiotic bacteroids. We identified a nodule-specific transporter, MtSLAH1 (*Medtr4g049640*) belonging to the TDT/SLAC transporter family (TCDB # 2.A.16.5) related to the Arabidopsis SLAC1 channel expressed in guard cells. MtSLAH1 is highly expressed in nodules and induced 10-fold by low nitrogen. Nitrate, a known SNF repressor, is capable of reducing the expression MtSLAH1 5-fold in 2 days. The exact expression of MtSLAH1 in nodule tissues was determined by promoter::GUS analysis and *in situ* hybridization. The MtSLAH1 protein has 369 amino acid residues spanning nine transmembrane domains and lacks the phosphorylation site characteristic of SLAC1. MtSLAH1 knockdown via RNAi nodules did not exhibit any observable phenotype. The MtSLAH1 channel permeability to dicarboxylates was tested by patch-clamp of *Xenopus* oocytes, but no current was detected. This result indicates that MtSLAH1 either requires a cofactor for activation or effluxes different anions than carboxylates. Overall, we identified a nodule-specific SLAC channel, and the physiological significance of this transporter to nodule development/nitrogen fixation is under scrutiny.

Keywords: Auto regulation of nodulation (AON), Diazotroph, Mutualism, Rhizobia
Introduction

Plant anion channels allow the efflux of anions from cells driven by the electrochemical motive force without incurring binding affinity between protein and substrate. They serve wide physiological functions including turgor pressure control (Negi et al., 2008; Vahisalu et al., 2008), change in membrane potential (Vincill et al., 2005; Negi et al., 2008; Vahisalu et al., 2008), organic acid excretion (Furukawa et al., 2007; Liu et al., 2009; Wu et al., 2014; Wang et al., 2017), tolerance to salinity (Zheng et al., 2014; Qiu et al., 2016; Zhang et al., 2016) and inorganic anion nutrition (Cubero-Font et al., 2016; Jaborsky et al., 2016). The recent molecular identification of AtSLAC1 in guard cells and its homologs (SLAH1-3) in roots allows a better understanding of their function and of mechanism that controls their activation.

SLAC1, the first S-type anion channel in the guard cells of A. thaliana, is a key component of stomatal closure under the control of various biotic and abiotic signals through ABA signaling pathway (Geiger et al., 2011). It is expressed in guard cells and primarily involved in the release of NO₃⁻ and Cl⁻ (the ratio of 10) from guard cells during stomatal closure (Negi et al., 2008; Vahisalu et al., 2008). It shows relatively low homology to the E. coli tellurite resistance multidrug transporter TehA, a member of the Tellurite-resistance/C₄-Dicarboxylate Transporter family (TDT) (Negi et al., 2008; Vahisalu et al., 2008; Vahisalu et al., 2010). SLAH3 (SLAC1 homologue 1) is also expressed in guard cells and has been identified as a voltage-dependent, nitrate-activated S-type anion channel with a NO₃⁻/Cl⁻ ratio of 20 (Geiger et al., 2011). Although SLAC1 and SLAH3 showed some expression level overlap, they are involved in different functions through different activation mechanisms. SLAC1 is primarily involved stomatal closure under various signals (light intensity, CO₂ concentration, ROS and ABA), whereas SLAH3 is only induced by drought stress with nitrate metabolism and nitrate signaling (Negi et al., 2008; Vahisalu et al., 2008; Geiger et al., 2011).

Additionally to guard cells, SLAH3 also presents on the plasma membrane of cells in the root stele and pericycle of roots, and is believed to in regulate the shoot NO₃⁻/Cl⁻ ratio by favoring nitrate transport, and may play a role in NO₃⁻ alleviation of ammonium toxicity (Cubero-Font et al., 2016; Qiu et al., 2016). The S-type anion channel-associated homologs 1 to 3 (SLAH1-3) are expressed in roots, and have been identified as anion channels involved in the control of NO₃⁻ and Cl⁻ loading of the root xylem (Barbier-Brygoo
et al., 2011; Hedrich and Geiger, 2017).

SLAH1 is expressed in the xylem-pole pericycle of the root, where SLAH3 located in roots, and has a permeability to NO$_3^-$ and Cl$^-$ (the ratio of NO$_3^-$/Cl$^-$ is 5) (Qiu et al., 2016). It may regulate Cl$^-$ efflux from root to shoot, since SLAH1 mutant showed a lower accumulation of Cl$^-$ without significantly affecting the permeability of NO$_3^-$ (Maierhofer et al., 2014b; Qiu et al., 2016). SLAH1 and SLAH3 act in tandem to regulate the ratio of NO$_3^-$/Cl$^-$ for root xylem loading under high salinity exposure. The expression of SLAH3 was not affected by salt stress. However, the ability to transport Cl$^-$ is compromised (Cubero-Font et al., 2016), while the expression of SLAH1 was remarkably decreased. Interestingly, a subunit of SLAH1 activates SLAH3 by heteromerization, which facilitates SLAH3-mediated chloride efflux from pericycle cells into the root xylem vessels under salt stress (Qiu et al., 2016). SLAH2, the closest homolog of SLAH3, has been identified as a nitrate-specific channel because of its predominant permeability to NO$_3^-$ (Maierhofer et al., 2014b). The physiological role of SLAH2 has not been identified yet. It may be involved in nitrate specific loading of xylem vessels. Thus, plants regulate NO$_3^-$/Cl$^-$ ratio between root and shoot under salinity via differential expression SLAH1-3 as well as reassembly of SLAH1/SLAH3 subunits (Cubero-Font et al., 2016; Qiu et al., 2016).

Despite the importance of SLAC/SLAH family, most studies of members of this family are still limited to A. thaliana. Further studies of anion channels and their regulation to other plant species are needed since they could have acquired distinct functions. So far, no SLAC channel has been characterized in Medicago truncatula, ever though that some homologs are highly expressed in nodules indicating a role in nodules symbiosis, as described in Chapter 4.

_M. truncatula_ is a model legume species for studying symbiotic nitrogen fixation with well-established genomic data and molecular and genetic resources available (Lozano-Baena et al., 2007; Colditz and Braun, 2010; Young et al., 2011) (http://www.genome.ou.edu/medicago.html). Legume-rhizobia symbiosis contributes about 40 million tons of reduced nitrogen into agricultural fields each year and is essential to sustainable agricultural ecosystems. Thus, understanding the whole process is necessary for optimization of symbiosis nitrogen fixation (SNF) (Zhao and Dixon, 2009; Young et al., 2011).
Establishment of SNF requires enormous and comprehensive chemical signals and nutrients exchanges. Nitrogen-starving legumes secreted flavonoids into the rhizosphere, and bacteria in the soil receiving signals activated the nodulation factors (Nods). Perception of these Nod factors by the hosts triggered the reprogramming of root cells to allow bacteria invasion via tube-like infection threads and eventually released bacteria into the nodule primordium via endocytosis, forming an organelle-like structure called the symbiosomes. Within symbiosome, the bacteria differentiate into bacteroids that fix atmospheric nitrogen into ammonium for the plants in exchange of carbohydrate produced by the host. Reviews of legume-rhizobia symbiosis and legume root nodule development are available (Brewin, 1991; Kinkema et al., 2006; Desbrosses and Stougaard, 2011; Sulieman and Tran, 2014).

Several potential roles of SLAC/SLAH channels would involve in symbioses nitrogen fixation. They would control the transmembrane potential and play a role in osmoregulation and cytosolic ion homeostasis. The calcium spiking may trigger the activation of an anion channel that allows Cl\(^-\) efflux and K\(^+\), which might serve as a charge balance, which eventually initiates depolarization and induces downstream signaling pathways (Udvardi and Day, 1989). Alternatively, they could serve as nitrate transporters and response to external nitrate. All SLAC channels identified so far have great permeability of nitrate than chloride. Interesting, nitrogen-fixing symbioses are suppressed by nitrate, by inhibiting the existing nodules and promoting the senescence process (Streeter, 1985; Luciński et al., 2002; Van Noorden et al., 2016). The effects of nitrate on the symbiosis are complex and remain unclear. Several mechanisms have been proposed. First, increasing the concentration of NO\(_3^–\) in infected nodule cells could reduce membrane potential across the symbiosomes membrane, which would result in inhibiting any transport process dependent upon this energization by host cells (Udvardi et al., 1991). When the food supply has been cut off, the existing nodules may be senesced, involving restriction of oxygen diffusion into the infected zone and production of reactive oxygen species. Alternatively, nitrate uptake or Increased nitrate concentration may reduce the auxin level of autoregulation of nodulation, and so inhibit nodulation (Ferguson and Mathesius, 2014). The lastly, the presence of nitrate may cause changes in the resistance to O\(_2\) diffusion, which might directly affect on nitrogenase activity, even on bacteroid metabolism (Luciński et al., 2002).
Our study identified seven SLAC channels in the *M. truncatula* genome. By exploring their chromosomal distribution, structures of gene, expression patterns in chapter 4, we found out that a SLAC channels (Medtr4g048640) have the highest expression value in nodules among seven channels (Chapter 4) named as MtSLAH1 (previously MtSLAC3) as an anion channel based on its phylogenetic location. We determined tissue expression patterns via *in situ* hybridization, performed transcription analysis with treatments with different *sinorhizobium meliloti* mutants, and transport assays using a heterologous expression system. We also discuss the possible physiological function of MtSLAH1 in symbiosis nitrogen fixation. To our knowledge, this study would be a first SLAC/SLAH channels in legume nodules, which provides new insights into understanding SLAC channels in plants and an overall picture of the functional roles that SLAC transporters may play in symbiosis nitrogen fixation.

**Results and Discussion**

**MtSLAH1 is a nodule-specific TDT/SLAC transporter**

Using the *M. truncatula* gene expression Atlas (Benedito et al., 2010) and RNAseq data (Young et al., 2011), we identified a putative slow anion channel (*Medtr4g049640*) with the highest expression in nodules among seven SLAC genes of the TDT/SLAC family in this species (Benedito et al., 2010). Following a thorough analysis of the gene family in *Medicago truncatula* conducted on Chapter 4, this gene has been named *MtSLAH1* (first channel identified of a SLAH1/SLAH4 subgroup). MtSLAH1 is a nodule-specific transporter, spans a genomic region of 2,384 bp and consists of two exons (Figure 1A). It codes a putative protein with 368 amino acid residues containing nine transmembrane domains (Figure 1B). All members of this family have an absolutely conserved phenylalanine (F) as a gating amino acid residue on the eighth transmembrane domain (Chen et al., 2010), including MtSLAH1 (F315; Figure S1, Chapter 4). Compared to AtSLAC1 and AtSLAH3, the two well-studied channels, MtSLAH1 has a short N-terminal tail of 88 amino acid residues and very short C-terminal with only six amino acids (Figure 1B). MtSLAH1 also lacks the phosphorylation sites experimentally demonstrated in AtSLAC1 and AtSLAH3 (Supplemental Figure S1, Chapter 4). Dreyer et al. predicted
there is only one (Serine) phosphorylation site in the cytosolic linkers in a SLAH1/4 subgroup (Dreyer et al., 2012), which MtSLAH1 possesses the conserved phosphorylation site (S187) within the 4th TMD (Figure S1, Chapter 4). To date, all characterized SLAC channels have been activated by multiple kinases with or without the presence of nitrate. It is possible that MtSLAH1 has more than one phosphorylation sites. Protein alignment shows that MtSLAH1 lacks the first transmembrane domain as found in AtSLAC1, the 3D structure being well-studied, probability because the AtSLAC1 has a continuous string of hydrophobic amino acid residues in the middle portion of this domain, whereas hydrophobic residues are more evenly spread in MtSLAH1 (Figure 1C). This explains why the TMHMM program did not recognize this tract as a bona fide transmembrane domain. Thus, despite the overall high evolutionary conservation in the primary sequence between the MtSLAH1 and AtSLAC1 channels (51% similarity), many structural features diverged, suggesting MtSLAH1 evolved to play a distinct function in nodules and have a different activation mechanism.

As shown in Figure 1 of Chapter 4, MtSLAH1 clustered in the SLAH1/4 subgroup. AtSLAH1 is a slow anion channel with five-fold preferential NO\textsuperscript{3−} permeability than Cl\textsuperscript{−}. It is fully functional and capable of complementing the slac1 mutant (Negi et al., 2008; Vahisalu et al., 2008). AtSLAH1 is expressed in the xylem-pole pericycle and thought to
mediate Cl\(^-\) transport from root to shoot by changing the net loading of xylem vessels in roots based on the fact that the Arabidopsis \textit{slah1} loss-of-function mutant showed lower accumulation of Cl\(^-\) but not NO\(_3^-\). Reducing the net xylem loading of Cl\(^-\) is key for salt tolerance in some species (Munns and Tester, 2008; Teakle and Tyerman, 2010). Indeed, under salt stress, the ability of \textit{AtSLAH1} to transport Cl\(^-\) was remarkably reduced, while NO\(_3^-\) remained unaffected due to the endogenous down-regulation of \textit{AtSLAH1} directly caused by high salinity (Qiu et al., 2016). Meanwhile, instead of kinase and nitrate, a subunit of \textit{AtSLAH1} activates \textit{AtSLAH3}, which sets \textit{AtSLAH3} free from its original activation mechanism. In other words, they work in collaboration to maintain the NO\(_3^-/Cl^-\) ratio high, which might be beneficial for improving salt tolerance (Cubero-Font et al., 2016; Qiu et al., 2016).

Analysis of gene expression in different organs (leaf blade, petiole, vegetative bud, flower, root, nodule seed) and nodule development stages (0, 4, 6, 10, 14 and 28 days post-inoculation, dpi) was performed and revealed that \textit{MtSLAH1} is highly expressed in nodules with some instances of leaky expression in other tissues (Figure 2A). This indicates this channel might play a role in symbiotic nitrogen fixation. During nodule development, \textit{MtSLAH1} expression starts early at four dpi (small root bumps) and increases until peaking in the young mature nodule (10 dpi) (Figure 2B). Our time-course analysis via RT-qPCR on nodulating roots further confirmed this pattern (Figure 3A). \textit{MtSLAH1} expression decreased dramatically two days after application of high nitrate, a known repressor of SNF (Streeter, 1985; Van Noorden et al., 2016) (Figure 2B). This result indicates a potential symbiotic role of \textit{MtSLAH1} during SNF, especially in supporting/maintaining the nodule function or nodule formation.

To improve the resolution on which tissues of the nodule \textit{MtSLAH1} is expressed, a promoter-GUS analysis was carried out. GUS expression revealed that \textit{MtSLAH1} is expressed throughout the nodule, except the nodule tip where the meristem is located (Figure 2C).
Figure 2. Temporal and spatial expression profiles of MtSLAH1. (A-B) Expression of MtSLAH1 (Affymetrix probeset Mtr.49905.1.S1_at) in the *Medicago truncatula* Gene Atlas (mtgea.noble.org/v3; Benedito et al., 2008) in different tissues and during nodule development. (C) Whole-mount GUS staining of nodules derived from hairy roots transformed with promoter-GUS vector driven by 2011bp promoter region of the MtSLAH1 gene. Scale: 1mm.
We then examined the expression of MtSLAH1 in *M. truncatula* inoculated with *S. melliloti* mutant strains (*exoA*, *bacA* and *fixJ*), which are impaired in specific milestone stages of the establishment of the symbiosis (*Figure 3B*). Roots infected with *exoA* completely lacked nodule development at 28 dpi. Roots inoculated with this strain showed no *MtSLAH1* expression, indicating the requirement of infection for its expression. *MtSLAH1* expression elicited with *bacA* was significantly lower than the control sample inoculated with the wild-type bacterial strain (*Sm1021_lacZ*). *MtSLAH1* was highly expressed in roots inoculated with the *fixJ* strain at levels similar to the control nodulating roots thus demonstrating that *MtSLAH1* expression requires rhizobial infection, perhaps the start of nodule development, but it is independent of nitrogen fixation *per se*. Moreover, if *MtSLAH1* were involved in nodule senescence, we would expect higher expression in the *fixJ* nodules, since these nodules undergo early senescence that wild-type nodules, which did not occur. Overall, *MtSLAH1* is expressed in nodules, with an early onset, and might play a role in nodule formation or function.

*MtSLAH1* is expressed in all of the zones of the indeterminate nodule

The promoter-GUS analysis showed that *MtSLAH1* is expressed throughout the mature nodule except the nodule tip (*Figure 2C*). We further used *in situ* hybridization to directly virtualize *MtSLAH1* expression in nodule tissues. This technique revealed that *MtSLAH1* is expressed throughout all zones but at much higher intensity in infected cells, and clearly
not in non-infected cells of the nitrogen fixation zone (Figure 4). MtSLAH1 highly expresses in fixation zone as shown as the dark staining (Figure 4A), and it had a much higher expression in infected cells than in the uninfected cells (Figure 4C). Those results were consist to the publicly available laser capture microdissection (LCM) RNA-Seq data (Roux et al., 2014). MtSLAH1 was started to be detected in the meristem and the distal infection zone, gradually increasing its expression in the proximal infection zone, Interzone and finally peaking in the fixation zone (Supplementary Figure S1). Additional LCM data based on Affymetrix microarrays on different nodule cell types showed the expression of MtSLAH1 is slightly higher in infected cells than in uninfected cell (Limpens et al., 2013). These results consist us in situ hybridization figures. Altogether, MtSLAH1 expressed all of the zones and may be responsible for effluxing nitrate and chloride into cells for homeostasis.

![Figure 4. MtSLAH1 in situ hybridization in nodules of Medicago truncatula.](image)

MtSLAH1 knockdown nodules

Assessment of the MtSLAH1 loss-of-function nodule phenotype could help demonstrate the potential importance of the MtSLAH1 channel for rhizobial symbiosis. We compared
the MtSLAH1-RNAi composite plants (in which the whole root system was transgenic) with wild-type non-transgenic rooted cuttings (as controls) and found that the RNAi grew less vigorously than the control plants under same growing conditions due to less nodules produced in RNAi (Figure 5A-B). However, the nodules showed no apparent phenotype differences between mutant and the wild-type roots (Figure 5). qRT-PCR results confirmed that MtSLAH1 expression in the RNAi nodules was significantly lower than in control (Figure 5E). MtSLAH1 RNAi nodules were pink, meaning they were capable of normally producing leghemoglobin, the expression of which was confirmed by qRT-PCR (Figure 5C).

Figure 5. Phenotype of MtSLAH1 mutant plants. (A) The growth of whole plants after root transformation with the MtSLAH1 RNAi vector along with non-transgenic cutting as a control at 28 dpi with rhizobia. (1) a non-transgenic cutting plant, (2) a composite plant with whole transgenic roots and (3) a negative RNAi-transformation plant. (B) Root systems of the three treatments mentioned in (A). Wild-type (left) and transgenic nodule (right) under a light stereomicroscope (C) or fluorescent illumination (D). Transgenic nodules and roots appear red under red fluorescence filter due to the constitutive expression of DsRed. (E) Normalized qPCR ratios of MtSLAH1-RNAi and control (empty RNAi vector) nodulating root systems along with the expression of leghemoglobin (Metr5g066070) as a proxy for potential symbiotic nitrogen fixation. Bars= 1 mm.
MtSLAH1 likely requires an interacting protein to function

MtSLAH1 cRNA was injected into *Xenopus* oocyte system by voltage-clamp to examine whether the protein could transport malate and succinate, the presumed substrates for AtSLAC1, although at much lower permeability than nitrate or chloride (Negi et al., 2008; Vahisalu et al., 2008). No anion-mediated current was detected, even though the oocytes were confirmed to express the MtSLAH1 protein (data not shown). This result suggests that MtSLAH1 may require a co-factor, possibly a protein kinase to activate, or that it may appear impermeable to malate. Indeed, as mentioned above, all characterized SLAC channels in *A. thaliana* require protein kinases to be activated through phosphorylation (Negi et al., 2008; Vahisalu et al., 2008; Maierhofer et al., 2014b; Maierhofer et al., 2014a; Zheng et al., 2014; Qiu et al., 2016). We used the NetPhos 3.1 tool (www.cbs.dtu.dk/services/NetPhos/) to predict the potential phosphorylation sites and kinases of MtSLAH1. We found 15 potential sites (>=0.6) (Figure 6) that could be activated by PKC, UNSP, PKA or PKG. Seven sites were found in the MtSLAH1 N-terminus, while the others were located either on the seventh TMD or the helix linker (Y274).

![Figure 6: Prediction of phosphorylation sites in MtSLAH1. The Netphos 3.0 score for S, T, and Y residues was determined separately for each. The positional information is displayed explicitly.](image-url)
Values the prediction score (0.000-1.000). The displayed information is color-coded based on amino acid polarity.

Conclusion

The MtSLAH1 channel is specifically expressed throughout the indeterminate nodule of *Medicago truncatula*. It is induced after rhizobial infection, but it does not necessitate nitrogen fixation *per se*. MtSLAH1 may not be an essential gene for nodule functioning, as there might be some redundancy for nodule homeostasis, either through a channel of the same family (e.g., Medtr3g112530) or a different one. Our revised hypothesis is that MtSLAH1 mainly delivers NO$_3^-$ into nodule cells and might be a key element of nitrate sensing that controls the important mechanism of autoregulation of nodulation and inhibits the numbers of nodules in high-nitrate environments.

Perspectives

As mentioned earlier, the lack of evident phenotype of MtSLAH1 RNAi nodules might be due to several reasons. First, as mentioned in Chapter 3 for MtMATE30, RNAi often allows for some residual expression, which might be sufficient for fully accomplish the physiological function. In order to rule this out, we will apply CRISPR-Cas9 gene editing technology to knock the MtSLAH1 gene out. Two independent CRISPR/Cas9 vectors have been constructed, and production of hairy roots is underway. The nodulation number and nodule size in knockout roots would assess the potential involvement of MtSLAH1 on these processes. On the other hand, MtSLAH1 may be redundant with other members of the SLAC family, or different channels. As shown in Chapter 4, besides MtSLAH1, there are other two SLAC channels (Medtr6g034030 and Medtr3g112530) expressed in nodules. As mentioned earlier, all SLAC channels in *A. thaliana* were permeable to NO$_3^-$ and Cl$^-$.

Between these two additional channels, Medtr3g112530 belongs to the SLAH1/4 subfamily and might be more easily substituted the MtSLAH1 function and regulate the NO$_3^-$/Cl$^-$ ratio inside the cell for normal nodule function. Lastly, MtSLAH1 might just not be essential for nodule function under normal conditions, and only required under salt stress, for example. Indeed, additionally to SLAC channels, other anion channels were found in legume nodules, which might act redundantly to MtSLAH1. GmN70 and LjN70, inorganic anion transporters are expressed in the symbiosome membrane and have enhanced preference for nitrate, which might aid in the regulation of ion and membrane potential.
homeostasis, possibly in response to external nitrate concentrations for symbiosis (Vincill et al., 2005). The Medicago NIP/LATD (NUMEROUS INFECTIONS AND POLYPHENOLICS or LATERAL ROOT ORGAN-DEFECTIVE) transporter, a member of the PTR family, was demonstrated as a high-affinity nitrate transporter (Katta et al., 1999; Veereshlingam et al., 2004; Salehin et al., 2013). These transport activities might help regulate membrane potential homeostasis together with MtSLAH1. As mentioned earlier, RNAi mutant has less numbers of nodules than in wild-type nodules, which indicates MtSLAH1 might be involved in the controlling of the number of nodules via autoregulation of nodulation (AON) (Kinkema et al., 2006). MtSLAH1 could be a response to extracellular NO3-, which trigger AON signaling loop resulting reducing the number of nodules.

Material and Methods

Plant growth and biological material

Medicago truncatula (ecotype A17) was used for all the experiments. Plant growth conditions were as described in Chapter 3.

RNA extraction, cDNA synthesis and qPCR analysis

As described in Chapter 3.

Cloning the coding and promoter sequence of MtSLAH1

The MtSLAH1 promoter region (2011 bp upstream of the start codon) was amplified by using M. truncatula ecotype A17 genomic DNA as a template. The PCR product was cloned into pENTR/D and recombined using LR reaction (Gateway LR clonase enzyme mix) into the destination vector pKGWFS7 (Karimi et al., 2002) with dual reporter genes (GFP and GUS) to create proMtSLAH1::GFP-GUS fusion expression vector. The maps representing each of the final constructs used are included as Chapter 3 as Supplemental Figure S1 (All binary vector maps used in this chapter were shared same backbone with the promoter or the CDs of MtSLAH1).
To examine the spatial distribution of MtSLAH1 at the subcellular level, MtSLAH1 transcript without the stop codon was produced with cDNA from the total RNA of 28-dpi nodules of *M. truncatula* ecotype A17 as a template for amplification using KOD Hot Start DNA Polymerase (Millipore). The PCR product was directly cloned into pENTR/D Gateway entry vector (ThermoFisher) and recombined into the destination vector pGWB554 (Chapter 3 Supplemental S1).

**Root transformation, histochemical staining, light microscopy, and *in situ* hybridization**

The procedures of root transformation, histochemical staining and light microscopy of MtSLAH1 was identical to that described for MtMATE30 in chapter 3.

**MtSLAH1 RNAi nodule analysis**

The target region of the MtSLAH1 transcript (635-996 nt) was amplified, cloned into the pENTR/D entry vector, and transferred into inverted-repeat binary destination vector pK7GW1WG2_II-RedRoot with DsRed report gene (Jach et al., 2001; Limpens et al., 2004). MtSLAH1-pK7GW1WG2_II-RedRoot was transformed into *Agrobacterium rhizogenes* *Arqua1*, which was used for hairy root transformation. Transgenic nodules produced in composite plants were harvested and analyzed after 28 days of inoculation with *Sm1021-lacZ* rhizobia. Transgenic roots and nodules were screened for the presence of constitutive red fluorescence (DsRed) under MIF Olympus MVX10.

**Statistical analysis**

Statistical analysis was performed ANOVA single factor following by Student T-test to compare the treatments with the control group in Excel.

**References**


Benedito VA, Torres-Jerez I, Murray JD, Andriankaja A, Allen S, Kakar K, Wandrey


Broughton WJ, Dilworth MJ (1971) Control of leghaemoglobin synthesis in snake


RCAR1. Sci Signal 4: ra32


throughput quantitative RT-PCR analysis of transcription factor gene expression in Medicago truncatula. Plant Methods 4: 18


Transcriptome-Switches. PLoS One 5: e9519


is not essential for aerobic nitrogen fixation, but confers significant protection to oxygen-mediated inactivation of nitrogenase in vitro and in vivo. Mol Microbiol 14: 101–14


Roberts SK (2006) Plasma membrane anion channels in higher plants and their


Changes of Outer Cortical Root Cells in Early Determinate Nodulation. 14: 839–847


Sun SJ, Qi GN, Gao QF, Wang HQ, Yao FY, Hussain J, Wang YF (2016) Protein kinase OsSAPK8 functions as an essential activator of S-type anion channel OsSLAC1, which is nitrate-selective in rice. Planta 243: 489–500


Zhao M, Chen Y, Qu D, Qu H (2011b) TSdb: a database of transporter substrates linking metabolic pathways and transporter systems on a genome scale via their shared substrates. Sci China Life Sci 54: 60–4


Supplementary table

Supplemental Table 1 List of primers used in these studies

<table>
<thead>
<tr>
<th>Quantitative RT-PCR</th>
<th>Sequence</th>
<th>Gene ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT-MtSLAH1-fw</td>
<td>ACAAATGGGTTGGAAAGAGAGTG</td>
<td>Medtr4g049640</td>
</tr>
<tr>
<td>RT-MtSLAH1-rv</td>
<td>GCAGCAAAAGAAACAAGAAGAACAC</td>
<td></td>
</tr>
<tr>
<td>MtEF1a(RT)-fw</td>
<td>TGACAGGCAGCTGTAAGG</td>
<td>Medtr4g008860 (Zhang et al., 2010)</td>
</tr>
<tr>
<td>MtEF1a(RT)-fw</td>
<td>TCAGCGAAGGTCTCAACCAC</td>
<td></td>
</tr>
<tr>
<td>MtUbi(RT)-fw</td>
<td>GCAGATAGACACGCTGGGA</td>
<td>Medtr3g091400 (Kakar et al., 2008)</td>
</tr>
<tr>
<td>MtUbi(RT)-rv</td>
<td>AACTCTTGGGAGAGCAATGA</td>
<td></td>
</tr>
<tr>
<td>MtMSC27-fw</td>
<td>GTTGAAGTAGACATGGTGCTAACG</td>
<td>Medtr2g436620</td>
</tr>
<tr>
<td>Reaction/Method</td>
<td>Forward Primer</td>
<td>Reverse Primer</td>
</tr>
<tr>
<td>----------------</td>
<td>----------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>MtMSC27-rv</td>
<td>AAAGCAGGTTGTTCCTGCTAATGCT</td>
<td></td>
</tr>
<tr>
<td>RT-LEG-fw</td>
<td>GCGCCTACAGAGGCTTGAAGAAG</td>
<td>Medtr5g081000</td>
</tr>
<tr>
<td>RT-LEG-rv</td>
<td>CAGTGCTCAATTCTTGCTCC</td>
<td></td>
</tr>
<tr>
<td>RNAi and <em>in situ</em> hybridization</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MtSLAH1-RNAI-fw</td>
<td>CACCCGGTTGAAAGAGATTGCAATATG</td>
<td></td>
</tr>
<tr>
<td>MtSLAH1-RNAI-rv</td>
<td>TGCCAAAGCTTTAGAGGAAATG</td>
<td></td>
</tr>
<tr>
<td>Promoter cloning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pMtSLAH1-GTWY-FW</td>
<td>CACTACAACACACTTTCTATTATCTCCT</td>
<td></td>
</tr>
<tr>
<td>pMtSLAH1GTWY-RV</td>
<td>TTCAAAAGATGCTATAAAGGAAATGACGTGCGATGTTGCGACC</td>
<td></td>
</tr>
<tr>
<td>CRISPR/CAS9 CLONING</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MtSLAH1CAS-G1-FW</td>
<td>ATTGGATTCCGGCGATGTTGCGAC</td>
<td></td>
</tr>
<tr>
<td>MtSLAH1CAS-G1-RV</td>
<td>AAACGTCGCAACATCGCCGGAATC</td>
<td></td>
</tr>
<tr>
<td>MtSLAH1CAS-G2-FW</td>
<td>ATGGCAGCAAAGAAACAAGAAGA</td>
<td></td>
</tr>
<tr>
<td>MtSLAH1CAS-G2-RV</td>
<td>AAACTTCTTCTGCTTTGCTGAC</td>
<td></td>
</tr>
<tr>
<td>SUBCELLULAR LOCALIZATION OR PATCH-CLAMP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MtSLAH1-GTWY-FW1</td>
<td>CACCATGGGTGACATTGAAAAACAACC</td>
<td></td>
</tr>
<tr>
<td>MtSLAH1-GTWY-RV</td>
<td>ACCTTTATTAATAACATTTTCTTCTTCTTTATCAATTCC</td>
<td></td>
</tr>
<tr>
<td>MtSLAH1-GTW-RV-NO</td>
<td>CATTTAGGCATCCTATGTTGAGCAG</td>
<td></td>
</tr>
</tbody>
</table>
Conclusion and perspective

The symbiotic nitrogen fixation (SNF) of legumes plays an essential role in agricultural sustainability and ecological systems. SNF is a sophisticated process, which requires both partners reprogramming their transcription networks to accommodate each other. It requires massive signal and nutrient exchanges, most of which rely on transporters in both partners’ cell membranes. With the advent of deep sequencing technologies along with bioinformatics tools, many nodule-specific transporter genes were identified genome-wide, most of which await full functional characterization despite the potential importance of their biochemical and physiological roles for the process. Our lab focuses on the functional characterization of nodule-specific transporters and transcription factors in the model legume, *Medicago truncatula*. This dissertation reports the efforts to characterize two such membrane transporters. Moving forward it will be important to better understand the exact transcription mechanisms of membrane transporter genes via regulatory gene networks within nodule zones and cell types, and how certain processes (e.g., rhizobial infection, bacteroid differentiation) define the expression patterns of so many genes in the nodule.

From our results, the SLAC anion channel MtSLAH1 hardly plays a role in nodule organogenesis *per se*, but it may rather be involved in maintaining or supporting nodule function. So next question is what the physiologic function might MtSLAH1 play in the nodule? SLAC family has been well characterized in *A. thaliana*. Our preliminary patch-clamp analysis in *Xenopus* oocytes showed that MtSLAH1 is not, just by itself, a malate transporter. This discarding did not support our first hypothesis that MtSLAH1 could be the key component mediating dicarboxylate efflux to the symbiosome in support of energy requirements for nitrogen fixation by bacteroids. However, MtSLAH1 could still be a malate transporter, but requires a co-factor to activate the activity in oocytes. Alternatively, it might efflux other anions, such as nitrate or chloride, two common substrates of SLAC transporters in *A. thaliana*. To test the first possibility, we are testing a MtSLAH1 permanently open version of the channel by mutating the gating amino acid. If MtSLAH1 has favorable selectivity for dicarboxylates, we will be able to detect activity with patch-clamp assays. As for the second possibility, if MtSLAH1 transports nitrate and chloride, it may be regulating the membrane potential and may even be coupled with other ion channels, just like reported for AtSLAC1 and AtSLAH3. Ca$^+$ influx that occurs during intercellular invasion may activate an anion channel that effluxes Cl$^-$, and K$^+$, balancing out the charge and allowing Ca$^{2+}$ to get into the cell. Since MtSLAH1 was also expressed...
in the infection zone, it could serve this purpose. To test this hypothesis, we need to add Ca\(^{2+}\) in the buffer solution and verify whether it increases the current. If the current increases, MtSLAH1 can indeed efflux chloride and/or nitrate and is induced by Ca\(^{2+}\) in order to cause membrane depolarization during Ca\(^{2+}\) spiking. If MtSLAH1 only effluxes nitrate, it could be involved in autoregulation of nodule (AON). The AON mechanism remains largely unclear at the genetic and physiological levels. MtSLAH1 could efflux nitrate into the symbiosome and trigger mature nodules to senesce. Beyond its ion selectivity, the MtSLAH1 activation mechanism is another interesting question. Firstly, with a co-expression map available for MtSLAH1, one could assess which kinases share a similar expression pattern. Following, proteins should be further tested through several protein-protein interaction techniques (e.g., yeast two-hybrid, bimolecular fluorescence complementation, co-immunoprecipitation) to find bona fide interactors. Overall, the physiological function of MtSLAH1 will mainly depend upon the selectivity and membrane localization.

Regarding our nodule-specific MATE transporter, MtMATE30 might be involved in supporting or maintaining nodule function, rather than nodule formation. MtMATE30 might localize to either the plasma membrane of infected root cells, or the symbiosome membrane facilitating the efflux the substrates out of the cytoplasm. Indeed, complementation experiments using prokaryotic heterologous system showed that MtMATE30 works as an effluxer. Since, the tobacco nicotine transporter NtJAT1 is the closest homologue to MtMATE30, we compared nicotine structure with other potential alkaloids previously described in the nodule and, despite the paucity of work in the scientific literature, we hypothesized trigonelline as a potential substrate for MtMATE30. Trigonelline, a byproduct of pyridine nucleotide cycle, is produced by the legume and effluxed into the symbiosome potentially to regulate nodule function. Trigonelline may be used as a NAD\(^+\) cache and salvaged back into nicotinic acid to maintain the high respiratory needs of the bacteroids. To further test this hypothesis, we will need to feed bacteroids with radiolabeled trigonelline and detect radioactivity in compounds of the pyridine nucleotide cycle. We also need to detect the activity of several enzymes in nodule cells (quinolinate phosphoribosyltransferase, nicotinate, phosphoribosyltransferase, nicotinamide phosphoribosyltransferase, nicotinamidase, and trigonelline synthase) to guarantee that trigonelline conversions are carried out directly in the nodule, and are not just translocated to this organ. In addition to trigonelline, we realize the urgent need to
characterize the full metabolomics profile in legume nodules, and ideally compare wild-type nodules with knockout mutants (for key legume genes as well as defective rhizobial strains) to complete the picture of potential substrates of transporters and enzymes.

Overall, we identified two nodule transporters, which are two small pieces to better understand the physiology of the legume nodule. These efforts may contribute to enhancing SNF in legume crops, and perhaps introduce this trait in non-legume crops, thus avoid the burden of high carbon footprint in nitrogen fertilizer usage in crop fields, and significantly reducing the cost of crop production.