UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

ANALYSIS OF THE FUNCTIONAL CONSERVATION OF GENES IN A PLANT SINGLE CELL TYPE: THE ROOT HAIR CELL

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

ZHENZHEN QIAO
Norman, Oklahoma
2017

ANALYSIS OF THE FUNCTIONAL CONSERVATION OF GENES IN A PLANT SINGLE CELL TYPE: THE ROOT HAIR CELL

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF MICROBIOLOGY AND PLANT BIOLOGY

BY

_____
Dr. Marc Libault, Chair

_____
Dr. John P. Masly

_____
Dr. Ben F. Holt III

_____
Dr. Laura E. Bartley

_____
Dr. Scott D. Russell

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my advisor Dr. Marc Libault for continuous and tremendous supporting me through the years of my Ph.D. I truly appreciate his patience, encouragement and closest guidance, and I wouldn't have finished my project without his generous support.

I also would like to thank my advisory committee members. I thank Dr. Bartley for her generous help since the time I applied for OU graduate program. Thanks to the Plant anatomy and Plant physiology classes offered by Dr. Russell and Dr. Holt, I can have a smooth transition from my master major cell biology to plant biology. I appreciate Dr. Masly for teaching me the concept of coding and the R programing knowledge, which helps a lot in my research. I also thank my advisory committee members for helping me to polish this dissertation.

It's a great pleasure to work with all the past and present lab members in Libault lab. Especially I thank Dr. Prince Zogli and Dr. Lise Pingault for advice and helps on the project. I also thank my fellow graduate students and friends, Neng Cheng, Mehrnoush Rey, Chengcheng Zhang, Fan Lin, Kangmei Zhao, Hengping Xu, Jing Yuan, Daniel Jones, and Zack Meyers for the support and encouragement.

Last but not least, my deepest thanks go to my parents, Jiuju Qiao and Xiuhua Qiao, for their unconditional and endless love and support. Also, I am so grateful to my beloved husband Huang Chen, for always being there for me, and I look forward to our adventures to come.

Table of Contents

Acknowledgements iv

List of Tables x

List of Figures xii

Abstract xvii

Chapter 1: Introduction 1

Overview of the molecular evolution of plant genes following genome duplications 2

Plant gene evolution at a single plant cell type-root hair cell level 3

Single plant cell type – the root hair cell 3

Molecular mechanisms controlling plant gene expression and its evolution 5

Molecular evolution of nodulation genes among legume species 6

Plasma membrane microdomain-associated proteins participate in legume nodulation 9

Figures and Tables 11

Chapter 2: Unleashing the potential of the root hair cell as a single plant cell type model in root systems biology 13

Abstract 14

Experimental objectives 15

Limitations of current techniques 17

Detailed protocol of the optimized method 19

Use of an ultrasound aeroponic system to enhance root hair density and treatment 19

Root hair isolation procedure 21

Molecular quantification of the level of purity of the root hair cell preparations 22
Applications of the ultrasound aeroponic system to investigate biological questions 23
Conclusion 24
Material and Methods 25
Plant growth 25
RNA extraction and cDNA synthesis 26
Quantitative real-time PCR and data analysis 26
Cloning and Soybean hairy root transformation 27
Figures and Tables 29
Chapter 3: A Comparative Genomic and Transcriptomic Analysis at the Level of Isolated Root Hair Cells Reveals New Conserved Root Hair Regulatory Elements 35
Abstract 36
Introduction 37
Results and Discussion 40
Re-annotation of the soybean genes preferentially expressed in root hairs. 40
Gene loss and gene retention after soybean WGDs depend on the transcriptional patterns of paralogous genes. 41
Transcriptional evolution between soybean paralogous genes at the level of the root hair cell. 45

Comparative genomic and transcriptomic analysis between soybean and Arabidopsis genes preferentially expressed in root hairs. 46
Conservation of the root hair activity of soybean promoters between plant species. 49
Comparative analysis of the promoter sequence of soybean and Arabidopsis genes preferentially expressed in root hairs revealed conserved DNA motifs. 51
Conclusion 54
Experimental Procedures: 55
Bacterial cultures 55
Cloning 55
Plant transformation and microscopy 56
RNA extraction, DNase treatment, reverse transcription, quantitative PCR primer design, quantitative PCR reaction conditions and data analysis. 57
Bioinformatic analyses 58
Figures and Tables 60
Chapter 4: Comprehensive comparative genomic and transcriptomic analyses of the legume genes controlling the nodulation process 73
Abstract 74
Introduction 75
Results 78
Syntenic relationships reveal the orthologous relationships between legume nodulation-related protein-coding genes 78

Comparative transcriptomic analysis reveals conservation and neo-/sub-functionalization between M. truncatula, G. max, L. japonicus and P. vulgaris nodulation genes 80
Legume genes specialized in the nodulation processes are forming gene modules on legume chromosomes 83
Discussion 85
Applying comparative genomic and transcriptomic analyses to reveal the conservation and divergence between nodulation orthologous/paralogous genes 85
Material and Methods 87
Identification of the nodulation genes across legume species 87
Syntenic analysis between nodulation genes 88
Transcriptomic analysis of soybean and M. truncatula nodulation genes 88
Gene density analysis 89
Figures and Tables 90
Chapter 5: The Gm*FWL1* (*FW2-2-like*) nodulation gene encodes a plasma membrane microdomain-associated protein 96
Abstract 97
Introduction 98
Results 101
GmFWL1 interacts with membrane microdomain-associated proteins 101
GmFWL1 directly interact with prohibitin proteins, markers of plasma membrane microdomains 103

Microscopic evidence for membrane microdomain localization of GmFWL1 and its role during the early stages of the nodulation process 104
GmFWL1 punctate localization is conserved in nodule plasma membranes ... 106
Legume comparative genomics reveal the direct interaction between GmFWL1 and GmFLOT2/4, the soybean ortholog to MtFLOT2 and MtFLOT4 107
GmFWL1 and GmFLOT2/4 accumulate at the tip of soybean root hair cells in response to B. japonicum inoculation 109
Discussion 111
GmFWL1 encodes a membrane microdomain-associated protein 111
Function of plasma membrane microdomains during legume nodulation 113
Role of plant FWL proteins 115
Material and Methods 117
Bacterial culture 117
Clonings 117
Co-IP assay and mass-spectrometry analysis 120
Plant transformation, treatment, and observation 123
Figures and Tables 127
Chapter 6 Conclusion and Perspective 153
Root hair cell – A single plant cell type 154
Functional conservation and divergence of duplicated genes 155
Function of plasma membrane microdomain-associated proteins during legume nodulation 157
References 163

List of Tables

Table 2- 1 qRT-PCR primers 34

Supplemental Table 3- 1 Annotation of the soybean root hair genes and loci according to the Wm82.a2.v1 version of the soybean genome sequence. 69

Supplemental Table 3- 2 Expression levels of soybean genes preferentially expressed in the root hair cells and their paralogous genes. 69

Supplemental Table 3- 3 Characterization of Arabidopsis genes preferentially expressed in root hairs upon mining and integration of six different root hair transcriptomic data sets 69

Supplemental Table 3- 4 List of the 6 soybean genes preferentially expressed in root hair cells and selected for the analysis of the activity of their promoter sequence. 70

Supplemental Table 3- 5 Primers designed to quantify gene transcription ("qRT-PCR"). 71

Supplemental Table 3- 6 Primers designed to clone selected soybean root hair promoter sequences ("Promoter cloning") 72

Supplemental Table 4- 1 Legume orthologous genes controlling nodulation. 95

Supplemental Table 4- 2 Expression levels of *G. max*, M. *truncatula*, *P. vulgaris* and *L. japonicus* genes controlling nodulation or orthologous to legume genes controlling nodulation. 95

Supplemental Table 4- 3 *G. max*, M. *truncatula*, *P. vulgaris* and *L. japonicus* paralogous and orthologous genes induced in root hairs upon rhizobia inoculation and/or specifically in nodules. 95

Supplemental Table 4- 4 Reference genomes, their annotation and the annotation of legume nodulation genes were used to calculate gene density for each legume chromosome. 95

Supplemental Table 5- 1 List of the designed primers. 146

Supplemental Table 5- 2 Characterization of 178 soybean proteins co-immunoprecipitated with GmFWL1 148

List of Figures

Figure 1- 1 Developmental stages of legume nodule organogenesis. 11

Figure 1- 2 Nodulation signal recognition pathway 12

Figure 2- 1 Root hair cells (black arrow pointing at one of the root hair cells) are single tubular root cells. 29

Figure 2- 2 Soybean seedlings grown in the ultrasound aeroponic system..... 30

Figure 2- 3 Isolated root hairs in the light microscope..... 31

Figure 2- 4 Expression analyses of soybean root hair specific genes.. 32

Figure 2- 5 Soybean transgenic roots and root hairs generated in the ultrasound areoponic system 33

Figure 3- 1Transcriptional evolution of soybean paralogs in root hair cells..... 60

Figure 3- 2 Validation of the preferential and non-preferential expression of 50 soybean genes in root hair cells. 61

Figure 3- 3 Evolution of the expression of soybean and Arabidopsis genes preferentially expressed in root hairs 62

Figure 3- 4 Transcriptional activity of four soybean promoter sequences in soybean hair cells..... 63

Figure 3- 5 Transcriptional activity of four soybean promoter sequences in Arabidopsis root hair cells. 64

Figure 3- 6 Transcriptional activity of two RHEs cloned upstream to the minimal 35S in soybean (B, D, and F) and A. thaliana (C, E, and G) root hair cells 65

Supplemental Figure 3- 1 Synteny relationships existing between the 20 soybean chromosomes (x-axis) and the 5 Arabidopsis chromosomes (y-axis). 66

Supplemental Figure 3- 2 Venn diagram of the Arabidopsis genes preferentially expressed in root hairs and characterized from six independent studies. 67
Supplemental Figure 3- 3 The relative level of expression of the 399 soybean genes preferentially expressed in the root hair cells 68
Supplemental Figure 3- 4 Promoter sequences of 6 soybean genes preferentially expressed in root hairs and cloned upstream the *UidA* reporter gene. 69
Figure 4- 1 Syntenic relationships between Mt*NSP1* (A) and Mt*FLOT2/4* (B) loci and chromosome regions from *Glycine max*, *Medicago truncatula*, *Lotus japonicus* and *Phaseolus vulgaris* 90
Figure 4- 2 Relative expression levels of *NF-YA1/2* (A, B), *NIN* (C, D), *NSP1* and *FLOT2/4* (E, F) *M. truncatula* genes (B, D, F) and their *G. max* orthologs (A, C, E) in inoculated and mock-inoculated root hair cells 91
Figure 4- 3 Relative expression levels of *NF-YA1/2* (A, B, C, D), *NIN* (E, F, G, H), *NSP1* and *FLOT2/4* (I, J, K, L) of *M. truncatula* genes (B, F, J) and their *G. max* orthologs (A, E, I), *P. vulgaris* orthologs (C, G, K) and *L. japonicus* (D, H, L) in various plant organs. 92
Figure 4- 4 Comparison of the expression profiles between *G. max*, *M. truncatula*, *P. vulgaris* and *L. japonicus* orthologous genes and groups during the early (i.e., root hair response to rhizobia inoculation; A, C, E) and late stages of the nodulation process (i.e., preferential expression in mature nodules; B, D, F). 93
Figure 4- 5 Representation of the density of the legume nodulation genes along each legume chromosome 94

Supplemental Figure 4- 1 Syntenic relationships between *Glycine max*,*Medicago truncatula*, *Lotus japonicus* and *Phaseolus vulgaris* to reveal orthology and paralogy between nodulation genes. 95

Figure 5- 1 Distribution of soybean proteins interacting with GmFWL1 in nodules according to their gene ontology. 127

Figure 5- 2 Subcellular localization of GmFWL1 in tobacco leaf protoplasts (A-D) and in tobacco leaf cells (E-F)..... 128

Figure 5- 3 Subcellular localization of the GFP-GmFWL1 chimeric protein in soybean root hair cell in response to *B. japonicum* inoculation. 129

Figure 5- 4 The punctuate plasma membrane localization of GmFWL1 was revealed in nodules using transmission electron micrographs after immunogold labeling. 130

Figure 5- 5 Subcellular localization of mCherry-GmFLOT2/4 fusion protein in soybean root hair cell in response to *B. japonicum* inoculation 131

Supplemental Figure 5- 1 Construction of the CGT3304 (A.) and CGT3305 (B.) vectors..... 133

Supplemental Figure 5- 2 Western blot analysis of the GmFWL1 co-immunoprecipitations.. 134

Supplemental Figure 5- 3 A. Schematic representation of the organization of the GmFWL1 protein across the plasma membrane predicted by Protter (Omasits et al., 2014; http://wlab.ethz.ch/protter/start/).. 135

Supplemental Figure 5- 4 Normalized level of expression of the 21 soybean prohibitins genes across 9 different soybean cell type and organs. 136

Supplemental Figure 5- 5 Split luciferase assays reveal the homo- and heterodimerization of GmFWL1 (F1) and prohibitins 137
Supplemental Figure 5- 6 Quantification of the intensity of the fluorescence of the GFP-GmFWL1 and mCherry-GmFLOT2/4 proteins (y-axis) across the soybean root hair cells (x-axis, μm) in mock-inoculated (blue) and *B. japonicum* inoculated (orange) conditions. 138
Supplemental Figure 5- 7 Subcellular localization of GFP-FWL1 (A-D, I-L) and mCherry-GmFLOT2/4 (E-H, M-P) fusion proteins driven by the Gm*FWL1* promoter in soybean root hair cell in response to *B. japonicum* inoculation. 139
Supplemental Figure 5- 8 Subcellular localization of pCvMV-GFP (A and C) and pCvMV-GmFWL1-GFP (B and D) in soybean root hair cell in response to *B. japonicum* inoculation. 140
Supplemental Figure 5- 9 Macro (A.) and microsyntenic (B.) relationships between MtSYMREM1 (i.e., Mt8g097320) and 4 soybean orthologous genes located on the chromosome 01, 05, 08 and 11 141
Supplemental Figure 5- 10 Expression patterns of 4 soybean genes encoding remorin proteins orthologous to Mt*SYMREM1* 142
Supplemental Figure 5- 11 Characterization of Gm*FLOT2/4* genes based on syntenic relationships with MtFLOT4 (i.e., Mt3g106430) and their nodule-specific expression patterns. 143
Supplemental Figure 5- 12 Macro (A.) and microsyntenic (B.) relationships between Medtr8g104870 and Gm*FWL3* (i.e., Glyma.08G043500). 144

Supplemental Figure 5- 13 Expression levels of soybean genes encoding GmFWL1 protein partners and preferentially (≥3- and <10-fold change between the expression levels in nodules compared to the second most highly expressed genes 145

Abstract

The evolution of plant species is tightly associated with major changes in their genome such as their allopolyploidy and autopolyploidy. These paleopolyploid events and recent genome duplications have contributed to the large sizes of the plant genomes and to the abundance of duplicated genes. This increases genomic content and, as a consequence, provides material for genetic mutations, drift, and selection. Therefore, genome duplication creates new possibilities of molecular evolution. Several studies focusing on the evolution of duplicated genes during plant development and in response to environmental stresses has been conducted. However, the cellular complexity of the plant organs used in these studies represents a difficulty to precisely characterize the molecular and functional conservation and divergence of duplicated genes.

In this dissertation, taking advantage of publicly available genomic and transcriptomic sequences and functional genomic datasets, we aimed to precisely delineate the conservation and divergence of plant gene transcription and protein function. This analysis has been conducted with an unprecedented level of resolution using a single plant cell type, the root hair cell which emerged around 400 million years ago (mya). The rationale of the selection of this single plant cell type to conduct the projects is the following: the molecular response of a plant tissue or organ, which is often selected when conducting plant molecular studies, is a reflection of the average molecular reponses of the different cell types composing the tissue/organ. This cellular complexity is a limitation in our understanding of the molecular evolution of plant genes. This concept will be largely discussed in the introduction of this dissertation (Chapter 1).

In chapter two, we described the development of an innovative plant culture system, the ultrasound aeroponic system, to access the root hair cells. This innovative plant culture system not only provides easy access to isolated root hair cells but also facilitates root hair observation and isolation, as well as the generation of transgenic root hair cells to enhance functional genomic studies. Finally, the ultrasound aeroponic system is compatible with the application of biotic and abiotic treatments on the plant root system. This is important because it opens avenues to precisely understand the adaptation of different plant species to environmental stresses and the evolution of these responses.

In chapter 3, mining the *Arabidopsis* (*Arabidopsis thaliana*) and soybean (*Glycine max*) genome sequence and root hair transcriptomes, we performed a comparative analysis to reveal the molecular evolution of plant genes at the single cell type level. Our analysis revealed that the transcriptional activity of plant genes and the mechanisms controlling their expression in root hair cells are highly conserved between plant species.

Focusing on nodulation, a biological process initiated by the symbiotic interaction between nitrogen-fixing bacteria (*Rhizobia*) and the legume root hair cell, we also performed a comparative genomic and transcriptomic analysis of the major regulators of the nodulation process and their homologs. This analysis is reported in Chapter 4. This study revealed the level of conservation and divergence of the nodulation-related genes across various legume species.

In chapter 5, we further examined the molecular and functional conservation of genes and proteins by performing a comparative functional analysis of two regulators of

the nodulation process: GmFWL1 and its interaction partner, a flotillin protein. Both proteins are microdomain-associated proteins. In response to rhizobial infection, these proteins translocate to the root hair cell tip where rhizobial recognition and invasion occur prior to the initiation of the infection process. Similar localization patterns of the *Medicago* (*Medicago truncatula*) flotillin ortholog in response to rhizobial infection (i.e., translocation at the tip of the root hair cell) suggests the conservation of the soybean and *Medicago* flotillin cellular functions and, more broadly, the role of plasma membrane microdomains during the nodulation process.

This dissertation describes the use of a plant single cell type, the root hair cell, to study the conservation of plant genes expression and function. This work will expand our knowledge in plant evolutionary biology.

Chapter 1: Introduction

Portions of this Introduction Chapter are from an invited book chapter: Genomics of Assimilation: Nitrogen, Genetics, and Genomics of Legumes, which is in preparation for submission.

Oswaldo Valdes-Lopez**,** Prince Zogli, **Zhenzhen Qiao** and Marc Libault. (2017). Invited book chapter: Genomics of Assimilation: Nitrogen, Genetics, and Genomics of Legumes.

Author contributions:

Dr. Marc Libault conceived the book chapter. I wrote two subchapters of this manuscript “Biological fixation and assimilation of the atmosphere nitrogen”.

Overview of the molecular evolution of plant genes following genome duplications

Polyploidy, which refers to organisms containing more than two complete sets of chromosomes, is recognized as a widespread biological phenomenon in the plant kingdom (Wendel, 2000). It is estimated that 50%-80% of the angiosperms have polyploidy history (Blanc and Wolfe, 2004b), including most of major crop species (potato, wheat, rice, corn, oat, soybean, cotton, *Arabidopsis thaliana*, *Medicago truncatula*, alfalfa, coffee, etc.) (Blanc et al., 2003; Blanc and Wolfe, 2004b; Lynch and Conery, 2000; Wendel, 2000). Polyploidy events mostly occurred early during plant evolution, around 340 million years ago (mya) for seed plants and 170 mya for angiosperms (Jiao et al., 2011). In addition, multiple rounds of whole genome duplication (WGD) also occurred over the past 170 million years of angiosperm evolution (Lee et al., 2013; Lyons et al., 2008a; Renny-Byfield and Wendel, 2014; Soltis et al., 2009; Soltis et al., 2014) including recent WGD events (D'Hont et al., 2012; Lu et al., 2013; Myburg et al., 2014; Velasco et al., 2010; Wang et al., 2012). For instance, in *Arabidopsis* lineage, 3 rounds of WGD, named as α, β, and γ happened between 50 and 140 mya (Beilstein et al., 2010; Bowers et al., 2003; Moore et al., 2007). In legumes (i.e., Fabales clade), upon their divergence from *Arabidopsis* 90 mya (Yang et al. 1999), soybean (*Glycine max*) and the model legume *Medicago truncatula* experienced one round of WGD 56.5 mya. Then, after the divergence between these two legume species, the soybean genome experienced one more round of WGD, 13 mya (Lavin et al., 2005; Schmutz et al., 2014). Besides WGD events, gene duplication is also the product of tandem duplication, transposon-mediated duplication, segmental duplication as well as retro-duplication (Panchy et al., 2016).

Upon duplication, most duplicated genes are lost or pseudogenized due to mutations or genetic drift (Panchy et al., 2016). In *Arabidopsis*, approximately 60% of the genes have at least one duplicated copy, of which, most are derived from WGD events (Bowers et al., 2003). From a functional point of view, Blanc and Wolfe (2004a) pointed out that genes involved in signal transduction pathway and transcriptional regulation network were preferentially retained, while those involved in DNA repair were degenerated. In addition, in this same study, the authors mentioned that duplicated genes diverged concertedly to form new parallel pathway with respective partners, which also might have co-evolved.

Plant gene evolution at a single plant cell type-root hair cell level

Single plant cell type – the root hair cell

The number of publicly available genome sequences and functional genomic datasets from various plant species and plant organs recently expanded allowing comparative analyses of plant genes and transcripts (Blanc and Wolfe, 2004a; Chaudhary et al., 2009; Galbraith and Birnbaum, 2006; Higgins et al., 2012; Roulin et al., 2013). However, the multicellular complexity of the plant organs remains a challenge when conducting comparative analyses because the data collected reflects the average response of all the cell types composing the organ, which differ between plant species. Therefore, the multicellular complexity of plant organs and the differential representation of the different cell types composing the same organ isolated from different plant species affect comparative transcriptional analyses. Hence, there's a

need to develop a single plant cell type approach to enhance our understanding of the molecular mechanisms controlling gene activity and their evolution.

Technological advances now allow the separation then isolation of specific single plant cell types. For instance, GFP can be used as a marker of specific cell types if specifically expressed under cell type-specific promoters. The labeled cells can then be isolated using fluorescence activated cell sorting (FACS) technology (Birnbaum et al., 2003; Brady et al., 2007a). Other methods exist to isolate single plant cell types such as laser capture micro dissection (Ithal et al., 2007; Klink et al., 2005; Santi and Schmidt, 2008; Takehisa et al., 2012) or isolation of nuclei tagged in specific cell type (INTACT) methods (Deal and Henikoff, 2010a; Deal and Henikoff, 2011). Nevertheless, these approaches remain challenging and often lead to a limited number of isolated cells. To overcome these limitations, several plant single cell types emerged, such as trichomes, pollen tubes, cotton fiber cells and gametophytes (Arpat et al., 2004a; Becker et al., 2003; Brechenmacher et al., 2009a; Brechenmacher et al., 2012a; Hossain et al., 2015; Libault et al., 2010d; Libault et al., 2010h; Nguyen et al., 2012b; Qiao and Libault, 2013b; Schmutz et al., 2010b; Wang et al., 2010a). These cell types are characterized by their ease to be isolated from the rest of the plant. Among them, the root hair cell differentiated 400 mya (Kim et al., 2006a) and is distributed in the entire plant lineage. The major biological function of the root hair cell is the uptake of water and nutrients from the rhizosphere. Its polar elongation helps to fulfill this function by increasing the surface of interaction between the root system and the soil. In addition, in legumes, the root hair cell is the primary site of infection by nitrogen-fixing bacteria named rhizobia and the establishment of their symbiotic relationship. Hence, based on

its biological function and its presence in numerous plant species, the root hair cell is a suitable to study the molecular evolution of plant gene activity and the molecular evolution of plant genes in response to a large number of biotic and abiotic stresses.

Molecular mechanisms controlling plant gene expression and its evolution

The epigenome and the binding of transcription factors to *cis*-regulatory elements located in the enhancer sequences of genes are two major regulatory mechanisms controlling gene expression. Several regulatory elements have been characterized notably those controlling plant response to abiotic (Guevara-Garcia et al., 1998; Inaba et al., 2000; Inaba et al., 1999; Matsuura et al., 2013; Nagano et al., 2001; Petit et al., 2001; Safrany et al., 2008) and biotic stresses (Lota et al., 2013; Pontier et al., 2001), and plant response to hormonal treatments (Hoth et al., 2010; Pla et al., 1993; Takeda et al., 1999). More recently, the emergence of genomic and transcriptomic resources led to the predictions of new plant regulatory elements (Choudhury and Lahiri, 2008; Ding et al., 2012; Gao et al., 2013; Haberer et al., 2004; Li et al., 2008; Satheesh et al., 2014; Vandepoele et al., 2009; Yilmaz et al., 2009; Zhang et al., 2005b). From an evolutionary point of view, the transcriptional regulatory mechanisms regulating the same biological process across different plant species are well conserved. For instance, Becker et al. (2014b) deciphered one regulatory element (i.e., AAAACAAA) present in the promoters of genes highly expressed in *Arabidopsis* cells characterized by their apical growth (e.g., root hair cell and pollen tube). Another comparative analysis performed by Kim et al. (2006a) revealed a set of root hair-specific regulatory elements upon comparative analysis between *Arabidopsis EXPANSIN7* (AtEXPA7) and its

homologs in different angiosperms, including barley (*Hordeum vulgare*), wheat (*Triticum aestivum*), maize (*Zea mays*), radish (*Raphanus sativus*), celandine (*Chelidonium majus*), and balsam (*Impatiens balsamina*). These results suggest a strong conservation of the molecular mechanisms controlling plant gene transcriptional activity during evolution.

Molecular evolution of nodulation genes among legume species

Nodulation is a mutualistic symbiotic relationship between legume plants and rhizobia, nitrogen-fixing bacteria. The nodulation process leads to the formation of a novel lateral root organ named the nodule where rhizobia fix atmospheric nitrogen to supply nitrite/ammonia or urea to the plants. In exchange, the plants provide a steady source of photosynthetic products to the bacteria. From an ecological point of view, symbiotic nitrogen fixation has a significant impact on the entry of nitrogen into ecosystems (Wagner, 2012).

Phaseoloid leguminous plants such as *Lotus joponicus*, *Medicago truncatula*, *Glycine max*, and *Phaseolus vulgaris* nodulate. This biological process is initiated by the biosynthesis of legume flavonoids to attract compatible rhizobia. After sensing this plant molecule signal, rhizobia release in the rhizosphere lipochito-oligosaccharides named Nod Factors (NFs), which are recognized by legume NF receptor, such as the Lotus NF receptor 1 (LjNFR1) (Radutoiu et al., 2003), the *Medicago* NF receptor like kinase MtLYK3/HCL (Limpens et al., 2003; Smit et al., 2007a) and the soybean GmNFR5 (Indrasumunar et al., 2011). The recognition of these molecular signals between the plant and the microbe triggers several molecular and cellular responses. For

instance, upon recognition of the bacterial NF, the plant root hair cell initiates calcium spiking leading, ultimately, to a transcriptional response and to the deformation of the root hair cell to maximize rhizobial infection. Then rhizobia infect the plant root hair cell through the formation of a tubular structure named the infection thread (Figure 1-1). Concomitantly to this infection, the plant cortex cells enter in a series of cell divisions. The primordia will lead to the formation of a new root organ, the nodule. Upon progression of the infection thread in the root, the bacteria are released into the plant cortex cells surrounded by plant cell membrane. This unique organelle is named "symbiosome" (Figure 1-1). In the symbiosome, rhizobia continue to proliferate and differentiate into bacteroids, which are capable of fixing atmospheric nitrogen (Gage, 2004; Jones et al., 2007; Oldroyd et al., 2011b).

Taking advantage of the availability of the genome sequences of several legume plants [e.g., soybean (*Glycine max*) (Schmutz et al., 2010a), *Medicago* (*Medicago truncatula*) (Young et al., 2011a), Lotus (*Lotus japonicus*) (Sato et al., 2008a), and common bean (*Phaseolus vulgaris*) (O'Rourke et al., 2014)], and the development of forward and reverse genetic tools, researchers have characterized the role of many legume genes involved in the nodulation process (Oldroyd, 2013) (Figure 1-2). For instance, mutants defective in NF perception are characterized by the absence of root hair deformation and by the interruption of IT formation and elongation (Arrighi et al., 2006). For example, the mutation of *LysM* family member, Mt*LYK3/HCL* or its lotus ortholog Lj*NFR1*, leads to the development of aberrant ITs (Limpens et al., 2003; Radutoiu et al., 2003). Downstream components such as Mt*DMI1*, *2* and *3* (*Does not make infection 1, 2 and 3*) are involved in signal transduction. Lj*POLLUX* and its

Medicago ortholog Mt*DMI1* encode ion channels. Mutation in Mt*DMI1* inhibits calcium spiking in the root hair cell in response to rhizobia inoculation (Ané et al., 2004). Proteins LjSYMRK and its *Medicago* ortholog MtDMI2 are not only mediating signal transduction during initial rhizobial infection but also have roles during the later stages of nodule development: the *dmi2* mutant is defective in the formation of symbiosomes (Kosuta et al., 2011). Mt*DMI3* encodes calcium and calmodulin dependent protein kinase (CCaMK) located in the nuclear membrane and is involved in the perception of calcium oscillations (Lévy et al., 2004). Specifically, MtDMI3 ortholog in lotus, LjCCaMK, phosphorylates LjCYCLOPS, a nuclear protein responsible for inducing Lj*NIN* (*NODULE INCEPTION*) gene expression (Messinese et al., 2007; Singh et al., 2014).

Obviously, the molecular mechanisms controlling the early stages of the nodulation process (i.e., rhizobial infection) across different legumes are conserved. In fact, the mechanism of signal recognition/transduction in rhizobia infection was probably recruited from the most ancient symbiosis program between plant and arbuscular mycorrhizae (AM). This hypothesis is supported by the conserved role of leucine rich repeat receptor like kinases in both nodulation and mycorrhization symbioses (Kistner and Parniske, 2002).

While the initial molecular response to rhizobia is conserved between legume species, later, during the later stages of the nodulation process, legume plants initiate distinctive nodule developmental strategies. For instance, *Medicago* develops indeterminate nodules originated from outer cortex root cells, with longitudinal shape and permanent apical meristem. Oppositely, soybean nodules are determinate nodules

characterized by their spherical shape and a non-permanent meristem, and originated from inner cortex root cells. While there are conserved nodulation factors such as NIN and NSP1 (Smit et al., 2005) controlling the later stages of nodule development across different legumes, the molecular function of the proteins encoded by these genes and leading to the differential development of the nodules remains unknown.

Plasma membrane microdomain-associated proteins participate in legume nodulation

In infected root hair cells, concomitantly to the progression of the IT, rhizobia proliferate before to infect the cortex cells. Several legume mutants are defective in IT formation and elongation leading to *nod-* (no nodules) or uninfected nodule phenotype. For example, *cyclops* and *ern1* mutants are characterized by a lower number of ITs (Yano et al., 2006; Yano et al., 2008) while the *nap1*, *pir1*, *sysrem1* and *flot4* mutants showed abnormal or arrested ITs (Haney and Long, 2010b; Lefebvre et al., 2010a; Miyahara et al., 2010; Yokota et al., 2009). The symbiotic *FLOTILLIN 2* and 4 (*FLOT2* and 4) and *SYMBIOTIC REMORIN 1* (*SYMREM1*) which encode plasma membrane microdomain-associated proteins, are also important regulators of *M. truncatula* rhizobial infection (Haney and Long, 2010b; Lefebvre et al., 2010a). Microdomains are highly dynamic sub-compartments of the plasma membrane enriched in sterols and sphingolipids. They are involved in signal recognition and transduction (Simons, 2000) as well as endocytosis and exocytosis processes (Parton and Richards, 2003). Lefebvre et al. (2010a) demonstrated that MtSYMREM1 is located in the plasma membrane and interacts with the symbiotic receptors NFP, LYK3 and DMI2 (Amor et al., 2003; Ane et

al., 2002; Smit et al., 2007b). MtFLOT4 also co-localize with MtLYK3 in the root hair cell plasma membrane upon rhizobial inoculation (Haney et al., 2011). In addition, MtFLOT4 was also detected in the IT membrane. These results strongly suggest that, during the early stages of the nodulation process and, more precisely, during root hair cell infection by Rhizobia, membrane microdomain-associated proteins act in the NF signaling recognition and signaling cascade, and are possibly involved in IT progression.

In this thesis, to accurately characterize the molecular evolution of plant genes, we first developed technologies and methods to fully take advantage of the single plant cell type-the root hair cell to perform accurate functional and comparative studies. In chapter 3, applying comparative genomic and transcriptomic analysis between *Arabidopsis* and soybean root hair cell genes, we characterized the transcriptional evolution of plant root hair genes and functionally characterized two conserved *cis*-regulatory elements controlling gene transcription in root hair cells. In chapter 4, we analyzed the level of conservation and divergence of legume nodulation genes. In chapter 5, comparing the role of the *FWL1* and *FLOT2/4* plasma membrane microdomain-associated genes between soybean and *Medicago*, we identified that the proteins encoded by these two genes are microdomain-associated proteins possibly working together to control legume-rhizobia mutualistic symbiosis nodulation process. Taken together, this work expands our knowledge on the conservation and divergence of the transcriptional regulation of plant genes in root hair cell and during nodulation across multiple plant species, which may further promote our understanding of plant evolution.

Figures and Tables

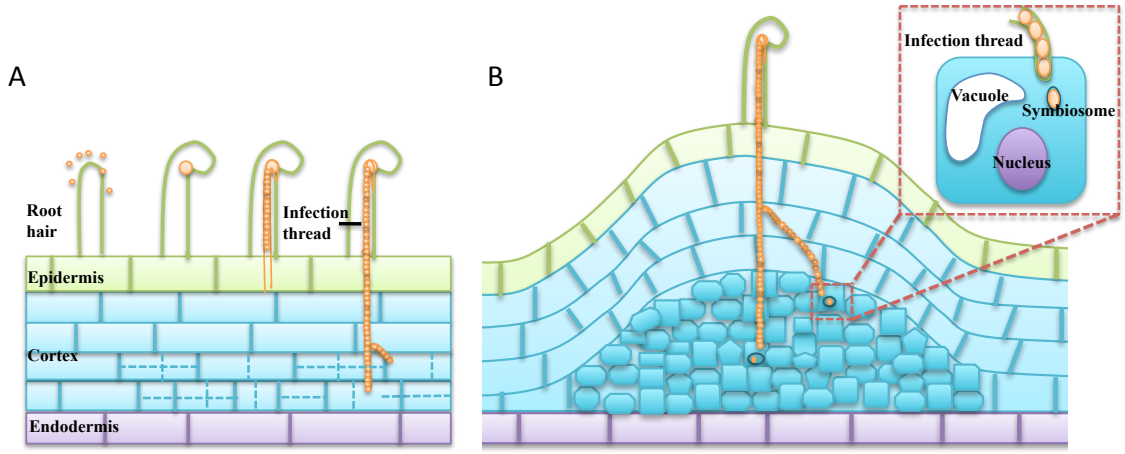


Figure 1- 1 Developmental stages of legume nodule organogenesis. A, early stage: rhizobia infection; B, late stage: nodule development and bacteroid formation.

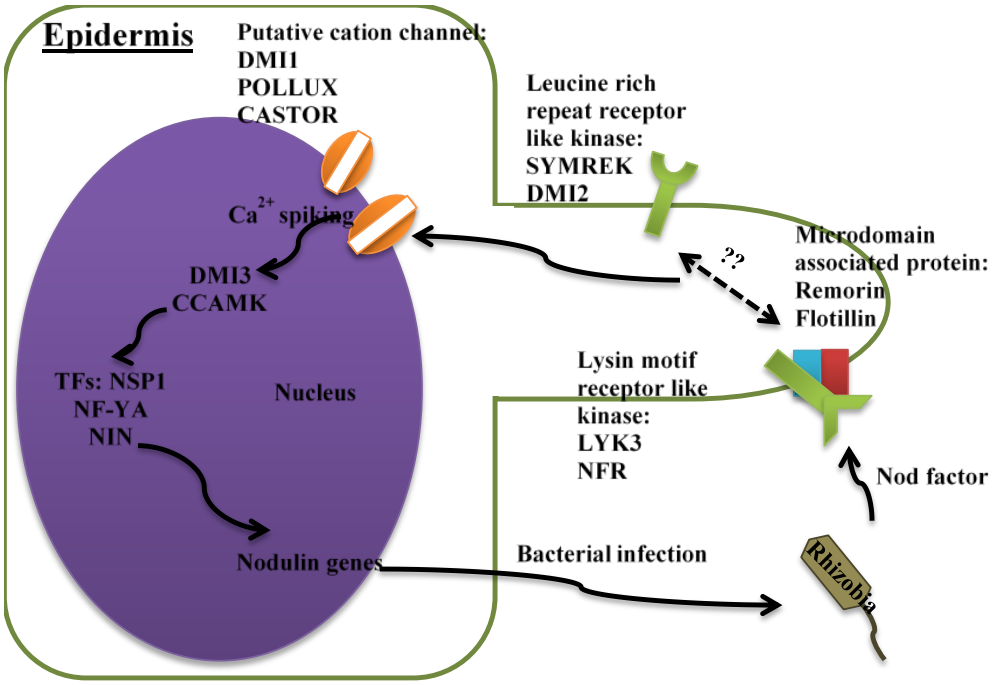


Figure 1- 2 Nodulation signal recognition pathway. See details about nodulation regulators in Supplemental table 4-1.

Chapter 2: Unleashing the potential of the root hair cell as a single plant cell type model in root systems biology

Authors: **Zhenzhen Qiao** and Marc Libault

Publication status: This chapter was published in Frontiers in Plant Science (DOI: 10.3389/fpls.2013.00484), which allowed incorporation into this dissertation.

Author contributions:

Dr. Marc Libault designed the study. I carried out the experiments, including root hair isolation, root hair RNA extraction, qPCR primers designation, qPCR analysis, hairy root transformation, GUS staining and microscopy. Micaela Langevin, the lab technician at that time, built up the ultrasound aeroponic plant culture system with my assistance. Dr. Marc Libault and I wrote the article published in Frontiers in Plant Science.

Abstract

Plant root is an organ composed of multiple cell types with different functions. This multicellular complexity limits our understanding of root biology because –omics studies performed at the level of the entire root reflect the average responses of all cells composing the organ. To overcome this difficulty and allow a more comprehensive understanding of root cell biology, an approach is needed that would focus on one single cell type in the plant root. Because of its biological functions (i.e. uptake of water and various nutrients; primary site of infection by nitrogen-fixing bacteria in legumes), the root hair cell is an attractive single cell model to study root cell response to various stresses and treatments. To fully study their biology, we have recently optimized procedures in obtaining root hair cell samples. We culture the plants using an ultrasound aeroponic system maximizing root hair cell density on the entire root systems and allowing the homogeneous treatment of the root system. We then isolate the root hair cells in liquid nitrogen. Isolated root hair yields could be up to 800 to 1000 mg of plant cells from 60 root systems. Using soybean as a model, the purity of the root hair was assessed by comparing the expression level of genes previously identified as soybean root hair specific between preparations of isolated root hair cells and stripped roots, roots devoid in root hairs. Enlarging our tests to include other plant species, our results support the isolation of large quantities of highly purified root hair cells, which are compatible with a systems biology approach.

Experimental objectives

Our understanding of root biology (i.e. root development, root cell differentiation and elongation, response to biotic and abiotic stresses) is based on -omic studies performed at the level of the entire root system or specific regions of the root as well as from the identification of mutants showing defects in root development. These mutants were characterized from the model plant *Arabidopsis thaliana* (Benfey et al., 1993; Mishra et al., 2009; Rogg et al., 2001) as well as other plants where genetic tools are well developed [e.g. *Medicago truncatula* (Catoira et al., 2000; Tadege et al., 2008a), *Oriza sativa* (Kurata and Yamazaki, 2006)*, Lotus japonicus* (Perry et al., 2003; Schauser et al., 1998). These valuable studies led to the identification of important genes and even gene networks controlling plant development and adaptation to stresses (Bruex et al., 2012; Schiefelbein et al., 2009).

To enhance our current understanding of root biology, a systems biology approach is needed to take advantage of the recent improvements in technologies such as mass spectrometry and high-throughput sequencing. One challenge when studying root biology is the multicellular complexity of plant roots. For example, -omic analysis at the level of a complex organ such as the root represents an average of the responses of the different cells composing the sample. Consequently, cell specific transcripts, proteins and metabolites as well as cell-specific epigenomic changes will not be revealed resulting in a partial understanding of the specific response of a cell or cell type to a stress and difficulties to fully integrate the various –omic data sets.

To demonstrate that a single cell type model represents an attractive alternative to overcome plant multicellular complexity and to better understand gene networks, we

compared the transcriptomes of the soybean root hair to that of the whole root (Libault et al., 2010d). Of the 5671 transcription factor (TF) genes known in soybean (Schmutz et al., 2010b; Wang et al., 2010b), we were able to detect transcripts for 3960 TF genes mining the whole root transcriptome. Out of the 1711 TFs undetected in the whole root transcriptome, 425 (25%) were only detected in the root hair cell transcriptome. This result is surprising since root hair cells were clearly one of the cell type represented in the root samples used for transcriptomic analysis. We are assuming that the low proportion of root hair cells in the root sample led to a dilution of root hair specific transcripts challenging their detection. This analysis strongly supports the need to work on a single cell type such as the root hair cell rather than an entire tissue to enable a more sensitive and accurate depiction of transcript abundance and, as a consequence, plant cellular responses to environmental perturbation. In addition, working at the single cell level will provide data more amenable to the development of computational models and the mapping of gene networks. Using a single cell type system as a model, the information obtained will be clearly unambiguous and would lead to a better characterization of gene networks.

The understanding of root hair cell biology requires the application of the full repertoire of functional genomic tools. However, major challenges in characterizing the biology of a single differentiated root cell type are the limited access to the root system and the isolation of the root cells of interest. In this manuscript, we describe a method to: 1- homogeneously treat the plant ‘s root hair cells; 2- easily access the root system and, *a fortiori*, the root hair cell; 3- isolate large quantities of this single cell type.

Limitations of current techniques

The isolation of single differentiated root cell types is limited by: 1- the accessibility to the root system; 2- the cell wall which confers the rigidity of the plant and its overall structure. Laser capture microdissection is a popular technique to isolate specific cells types but it is labor-intensive and cell yields are very limited. Nevertheless, it has been successfully applied to study root biology (Ithal et al., 2007; Klink et al., 2005; Santi and Schmidt, 2008; Takehisa et al., 2012). A second method based on the labeling of cell type by the GFP has been recently established to measure *Arabidopsis thaliana* single plant cell type transcriptomes and their regulation in response to environmental stresses (Petersson et al., 2009; Zhang et al., 2005a). Using a collection of transgenic plants expressing the GFP in different root cell lines, *A. thaliana* root cell types were isolated after digestion of the cell wall and isolation of the resulting GFP positive protoplasts using cell sorting technology. This strategy allowed the identification of root cell type-specific genes validating the concept of root cell-specific transcriptomes. However, as reported by the authors of these studies, the digestion of the cell wall also led to a few changes of the plant transcriptome independently of the cell line or treatment. In addition, several studies highlighted a massive restructuration of the chromatin and epigenetic marks in leaf protoplasts in comparison to differentiated leaves cells (Chupeau et al., 2013; Ondřej et al., 2009; Tessadori et al., 2007; Zhao et al., 2001). A third method, the INTACT method, was applied on *Arabidopsis thaliana* to isolate hair and non-hair cells and analyze their transcriptome and epigenome (Deal and Henikoff, 2010a; Deal and Henikoff, 2011). This method is based on the expression of biotinylated nuclear envelope protein under

the control of a cell type-specific promoter sequence and the isolation of labeled nuclei using streptavidin-coated magnetic beads. The characterization of a cell-specific promoter is a pre-requisite to the INTACT method. While RNA and chromatin structure can be accessed using the INTACT method, other aspects of the biology of the plant cell such as its entire proteome and metabolome cannot be reached with this method.

Another strategy to study plant single-cell biology is to massively isolate easily accessible cell types. Such method has been successfully applied on aerial parts of the plant. For example, cotton fiber and pollen cells were isolated to investigate plant cell elongation mechanisms (Arpat et al., 2004b; Franklin-Tong, 1999; Padmalatha et al., 2012; Ruan et al., 2001). More recently, the soybean root hair (Figure 2-1) has emerged as a new single cell type model (Libault et al., 2010a). Various studies validate the use of the root hair cell as a model in systems biology through the analysis of the infection of soybean root hair cells by mutualistic symbiotic bacteria [i.e. the soybean root hair cell is the first site of infection by *Bradyrhizobium japonicum*, the nitrogen-fixing symbiotic bacterium involved in soybean nodulation (Gage, 2004; Kathryn et al., 2007)]. In these studies, soybean seedlings were germinated on agar plate preliminary to the inoculation of the plants with *B. japonicum* followed with the isolation of the root hair cells. Various -omics approaches were successfully used to decipher root hair cell biology, including transcriptomic (Libault et al., 2010d), proteomic (Brechenmacher et al., 2009c; Brechenmacher et al., 2012b; Wan et al., 2005b), phosphoproteomic (Nguyen et al., 2012c) and metabolomic (Brechenmacher et al., 2010a) methods. In addition to being a model to investigate plant microbe interactions, the root hair cell is

also an excellent model to decipher plant cell regulatory networks in response to abiotic stresses. This is based on their primary role in water and nutrient uptake.

To utilize full potential of this attractive single cell type as a model in root systems biology, root hairs must be evenly treated preliminary to their isolation from the rest of the root system in quantities compatible with any -omic analysis, and a fortiori, transgenic root hair cells must be isolated to perform functional genomic studies at the level of a single cell type. To reach these two goals, we developed the method described below combining the use of an ultrasound aeroponic system to generate and evenly treat a large population of root hair cells and the purification of frozen root hair cells using a highly selective filtration system. This method overcomes the limitations related to the use of the agar media to germinate seedlings such as the heterogeneity of the root hair cell population produced (i.e. root hair cells interact with the agar or are expanding in the atmosphere impacting their physiology) and open new avenues to investigate root hair cell biology because enabling functional genomic studies (see below). To date, we focused on the isolation of soybean root hair cells but the method described below has been validated using other plant models such as maize, sorghum and rice.

Detailed protocol of the optimized method

Use of an ultrasound aeroponic system to enhance root hair density and treatment

The study of root hair cell response to stresses presupposes: 1) the even treatment of the root system under control and stressed conditions to minimize biological variations; 2) the optimization of the growth conditions of the root system and the enhancement of the

differentiation of root hair cells on the root system; 3) an easy access to the root hair cell compatible with their observation and isolation; 4) the development of methods to efficiently isolate them.

We recently developed a method which fulfills these different requirements. Five days-old soybean seedlings germinated on a mixture of vermiculite and perlite (3:1) were transferred to the ultrasound aeroponic system under controlled conditions (long day conditions, 25-27 °C, 80 % humidity; Figure 2-2). This system is composed of two units: the fogger system and the cloner unit (EZ-CLONE Enterprises Inc.). The fogger system relays on the production of a five micrometres (µm) droplets of nutritive solution by ultrasound misters (OCEAN MIST®, DK24) which atomize nutrient solution into a nutrient-rich mist by vibrating at an ultrasonic frequency [in the case of soybean, we are using the B&D nutritive solution (Broughton and Dilworth, 1971b)]. An air flow pushes the cool mist into the cloner unit where plants are growing. The quantity of mist produced by the fogger system is controlled by the number of mist makers used per fogger system as well as by a timer controlling the frequency and duration of the production of mist. Using a thin mist to feed the plant maximized the oxygenation of the root system, an important factor contributing to a higher density in root hair cells of the root system (Shiao and Doran, 2000); Figure 2-2C&2D). Altogether, this unique system optimizes root growth, enhances root hair cell density and offers an easy access to the root hair cell compatible with their observation and isolation (Figure 2-2E).

Root hair isolation procedure

Root hair cell isolation has been repetitively applied on soybean (Brechenmacher et al., 2010a; Libault et al., 2010d; Nguyen et al., 2012c; Perry et al., 2003; Wan et al., 2005b) Concomitantly to the development of the aeroponic system, the method used to isolate soybean root hair cell was updated to reach two objectives: 1- maintain or enhance the level of purity of the root hair cell preparation from the rest of the root system; 2- maximize root hair yields. Several methods exist to isolate root hairs including gentle brushing of the frozen root system into liquid nitrogen (Bisseling and Ramos Escribano, 2003) or stirring of the roots immersed in the liquid nitrogen with glass rod preliminary to their isolation (Bucher* et al., 1997; Roehm and Werner, 1987). The first method maximizes root hair purification but root hair yields are low and the method is labor intensive. The second method provides large quantities of plant material but the root hair cell preparation could be easily contaminated by non-root hair cells such as root fragments resulting from the stirring.

We optimized the latter method as described below. Briefly, the root systems of three weeks-old soybean plants are isolated, rapidly wiped off to remove extra moisture then immediately immersed into liquid nitrogen. This rapid freezing prevents the undesirable stress of the root and root hair cells due to their manipulation. All subsequent steps are performed in liquid nitrogen. Frozen roots are gently stirred into liquid nitrogen by a glass rod for 10 minutes. The flow of liquid nitrogen is sufficient to break root and root hairs. The liquid nitrogen containing the root hairs is filtered through a 90µm sieve into a beaker. Based on stereomicroscopic observations, this mesh offers the best compromise to maximize the level of purification of the root hair

cells without compromising the yield (Figure 2-3). The stripped roots are rinsed 5 to 7 times to collect the remaining root hair cells and increase the yield (i.e. as much as 1000 mg of isolated root hair cells were isolated from 60 three-week old soybean plants). The plant material harvested is usable the most up-to-date molecular approaches.

Molecular quantification of the level of purity of the root hair cell preparations

To evaluate the purity of the root hair cell preparations, we quantified the expression of several "root hair-specific" genes in both isolated root hair and stripped root samples. These genes were selected from the soybean transcriptome atlas (Libault et al., 2010c) based on their high or specific expression in root hair cells compared to stripped roots (Figure 2-4A). We are assuming that the low transcript abundance of these "root hair-specific" genes in stripped roots is the consequence of the presence of remaining root hair cells or root hair cell nuclei in the stripped root samples (i.e. the nucleus of mature root hairs are located at the base of the cell).

The fold change of gene expression level in root hair cell *versus* stripped root ranged from 11.9 (Glyma09g05340) to 44.1 (Glyma15g02380) based on RNA-seq data (Figure 2-4A). Applying qRT-PCR methods, we analyzed the quality of the plant material collected using our optimized method compared to a previous root hair cell isolation method (Wan et al., 2005) (Figure 2-4B). Independently of the root hair isolation method used, we observed a higher abundance of transcripts encoded by the nine candidate genes in isolated root hairs compared to stripped roots supporting high levels of purification of the root hair cells. Using root hair cells and stripped roots collected using the method described by Wan et al. (Wan et al., 2005b), the fold

changes of expression of root hair specific genes between root hairs and stripped roots ranged between 42.4±56.6 (Glyma02g01970) and 133.0±67.1 (Glyma08g12130). Our optimized root hair cell method repetitively led to fold changes of expression of root hair specific genes ranging between 124.4±117.8 (Glyma04g04800) and 385.8±164.0 (Glyma02g01970). This result supports a higher enrichment in root hair cells in root hair cell preparation using the updated method compared to Wan et al. (2005) method.

Applications of the ultrasound aeroponic system to investigate biological questions

The flexibility of the use of the ultrasound aeroponic system is fully compatible with the homogeneous treatment of the root system to analyze root hair response to biotic and abiotic stresses. The nature of the abiotic stresses allowed by the aeroponic system is diverse including: 1- changes of the chemical composition of the nutritive solution to analyze root hair response to nutrient deprivation, low and high pHs, salinity or heavy metal contaminations, etc.; 2- changes of the environmental conditions such as temperature, water potential, etc.; 3- inoculation of plants with pathogenic and symbiotic microorganisms. The latter was validated by inoculating two weeks-old root systems of the hypernodulation soybean mutants [i.e. NOD1-3, NOD2-4 and NOD3-7(Ito et al., 2007)] with a bacterial suspension of *Bradyrhizobium japonicum*, the soybean nitrogen-fixing symbiont. As soon as 10 days after inoculation, nodules emerged. Thirty days after inoculation, a large number of nodules were developing on the soybean roots [NOD1-3 (106.8 ± 27.7 nodules), NOD2-4 (159.7 ± 42.7 nodules) NOD3-7 (99.7 ±29 nodules.)]. As an element of comparison, NOD1-3, NOD2-4 and NOD3-7 mutants grown in vermiculite showed 441, 344 and 143 nodules, 17-18 days

after inoculation, respectively (Ito et al., 2007). While nodule number per root system is lower when using the aeroponic system, our data supports the use of the aeroponic system to study the interaction between the plant root and rhizobia and nodule development. The easy access to inoculated root hairs and root nodules compatible with their isolation and their observation upon rhizobia inoculation is a clear advantage to study the early and late stages of legume nodulation.

Another potential attractive application of the aeroponic system is the generation of composite plants (i.e. plants carrying a mixture of transgenic and non-transgenic roots growing from a wild type shoot) and, *a fortiori*, the easy access to a large mass of transgenic roots compatible with their observation and various molecular analyses. To test this potential utilization of the aeroponic system, we inoculated soybean shoots with *Agrobacterium rhizogenes* carrying our transgene of interest (in this case, a fusion between the cassava vein mosaic virus promoter and the *UidA* gene which encodes the beta-glucuronidase). Ten days after bacteria inoculation, a callus was formed and roots started to emerge (Figure 2-5A). Four weeks after inoculation, the emerged root system was stained using X-Gluc to reveal the β-glucuronisase activity (Ithal et al., 2007; Libault et al., 2010h) (Figure 2-5B). In average, we observed seven transgenic roots emerging from each composite plant. Stereomicroscopic observations revealed that these roots carry an impressive number of transgenic root hair cells (Figure 2-5C).

Conclusion

In this manuscript, we combined the use of an ultrasound aeroponic system with updated method to isolate root hair cells to maximize the potential of plant root hair cell

as a single cell type model for systems biology. This updated method has the following advantages: 1- enhance root hair cell density on the root system; 2- even and long-term treatment of the entire population of root hair cells to access the molecular response of the root hairs to various biotic and abiotic stresses.; 3- compatibility with the microscopic observation of the root hair cells; 4-leading to high yields of isolated root hair cells compatible with any –omic analyses.

In addition to be well-suited to perform –omics analyses at the level of one single cell type, the ultrasound aeroponic system has been validated to study plant-bacteria interactions and to produce large quantities of easy accessible plant material allowing functional genomic studies. Undoubtly, our updated method of generating a large amount of pure root hair cells will promote the progress of deciphering the regulatory mechanism of plant cell biology including plant cell response to environmental stresses.

Material and Methods

Plant growth

Soybean seeds (*Glycine max* [L.]) were surface-sterilized by three sequential treatments with 1.65% sodium hypochlorite (10 minutes each), rinsed three times with deionized water before a 10 minutes treatment with 10 mM hydroxychloride. Seed was finally washed three times with sterile water before sowing on sterilized mixture of vermiculite and perlite (3:1 ratio). Seeds were germinated under permanent light conditions at 25°C. One week later, the seedlings were transferred into fogger system and supplement with the mist of B&D plus 10 mM KNO_3. Cultured for another two

weeks in areoponic system, the seedlings were collected into liquid nitrogen for root hair isolation.

RNA extraction and cDNA synthesis

Total RNAs were extracted using Trizol Reagent (Invitrogen). Six to ten µg of total RNA were extracted from one preparation of isolated root hairs. Total RNA samples were treated with the TURBO DNase (Ambion) according to the protocol provided by the manufacturer before to reverse transcribe one µg of DNA-free RNA using oligodT and the Moloney murine leukemia virus reverse transcriptase as previously described (Libault et al., 2010d).

Quantitative real-time PCR and data analysis

Quantitative real-time PCR (qPCR) primers were designed using Primer3 software (http:// biotools.umassmed.edu/bioapps/primer3_www.cgi; Table 2-1). qPCR reactions were performed as described by (Libault et al., 2008) including an initial denaturation step of 3 min at 95 °C followed by 39 cycles of 10 seconds at 95 °C and 30 seconds at 55 °C. Dissociation curves were obtained using a thermal melting profile performed after the last PCR cycle: a constant increase in the temperature between 65°C and 95°C.

Cycle threshold (Soltis et al.) values were obtained based on amplicon fluorescence thresholds. According to Vandesompele et al. (Vandesompele et al., 2002), delta Ct was generated using the geometric mean of the cycle threshold of three reference genes [*Cons6*, *Cons7* and *Cons15* genes (Libault et al., 2008)]**.** PCR

efficiency (P_{eff}) for each sample was calculated using LinRegPCR ((Ramakers et al., 2003), and the expression level (E) were calculated using the equation $E = P_{eff}^{(-\Delta Ct)}$. The fold change of the gene expression levels between root hair *versus* stripped root was calculated for each root hair specific gene. Three independent biological replicates were generated for each condition and Student t-tests with two tails and two samples equal variance were applied to display the significant differences of gene expression between root hair and stripped root samples. *P* value <0.05 was regarded significant.

Cloning and Soybean hairy root transformation

As described by Libault et al. (Libault et al., 2010e), cloning of the cassava vein mosaic virus promoter upstream of the *UidA* gene was performed using the Gateway® system (Invitrogen, http://www.invitrogen.com/). The cassava vein mosaic virus promoter fragment was introduced first into the pDONR-Zeo vector (Invitrogen) using the Gateway® system BP Clonase®II enzyme mix, then into pYXT1 destination vectors carrying the *UidA* genes, using the Gateway® LR Clonase® II enzyme mix.

Two weeks-old soybean plants grown on pro-mix were used to generate composite plants. K599 *Agrobacterium rhizogenes* bacterial strain carrying the transgene of interest, a transcriptional fusion between the cassava vein mosaic virus promoter and the *UidA* gene, was grown at 30°C in LB medium supplemented with kanamycin. The bacteria were pelleted by centrifugation, and re-suspended in B&D medium supplemented with 10 mM potassium nitrate and acetosyringone (20 µM) to an optical density at 600 nm = 0.35.

Soybean shoots were cut between the first true leaves and the first trifoliate leaf

and placed into rock-wall cubes (Fibrgro). Each shoot was inoculated with 4 mL of *A. rhizogenes* suspension and then allowed to dry for approximately 3 days (23°C, 50% humidity, long-day conditions) before watering with deionized water. After 1 week, instead to transfer the composite plants into vermiculite-perlite as described by (Libault et al., 2009b), the transformed soybean shoot were transferred into the ultrasound aeroponic system supplemented with B&D medium plus 10 mM potassium nitrate. After two weeks, the β-glucuronisase activity of the soybean root system was revealed as described by (Libault et al., 2010h).

Figures and Tables

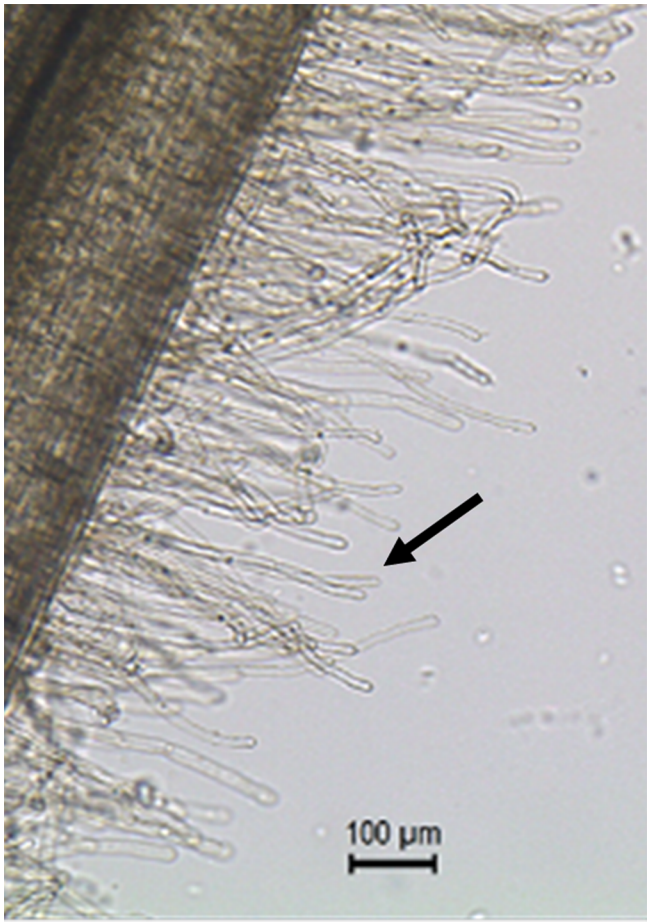


Figure 2- 1 Root hair cells (black arrow pointing at one of the root hair cells) are single tubular root cells. Their distinctive lateral elongation increases the surface of exchange between the plant's root system and the soil. The main function of root hairs is the uptake of water and nutrients from the rhizosphere.

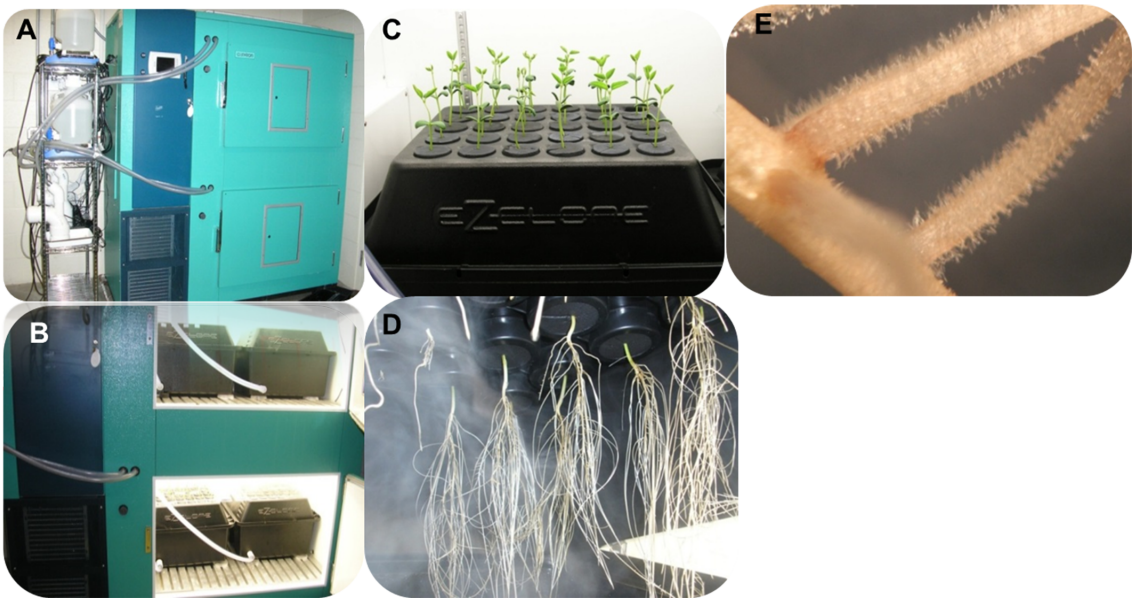


Figure 2- 2 Soybean seedlings grown in the ultrasound aeroponic system; (A) &(B) the whole system for plant culturing; (C)&(D) the plants in the EZ-cloner; (E) soybean root showing a high density in root hair cells.

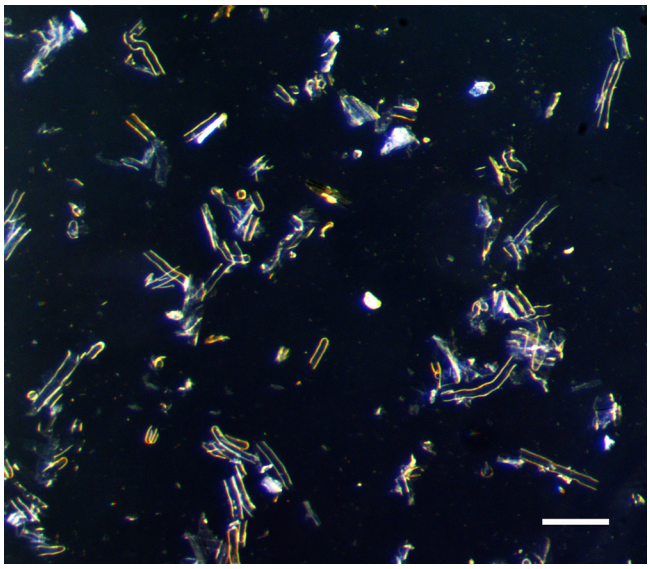


Figure 2- 3 Isolated root hairs in the light microscope. Bar = 100 µm.

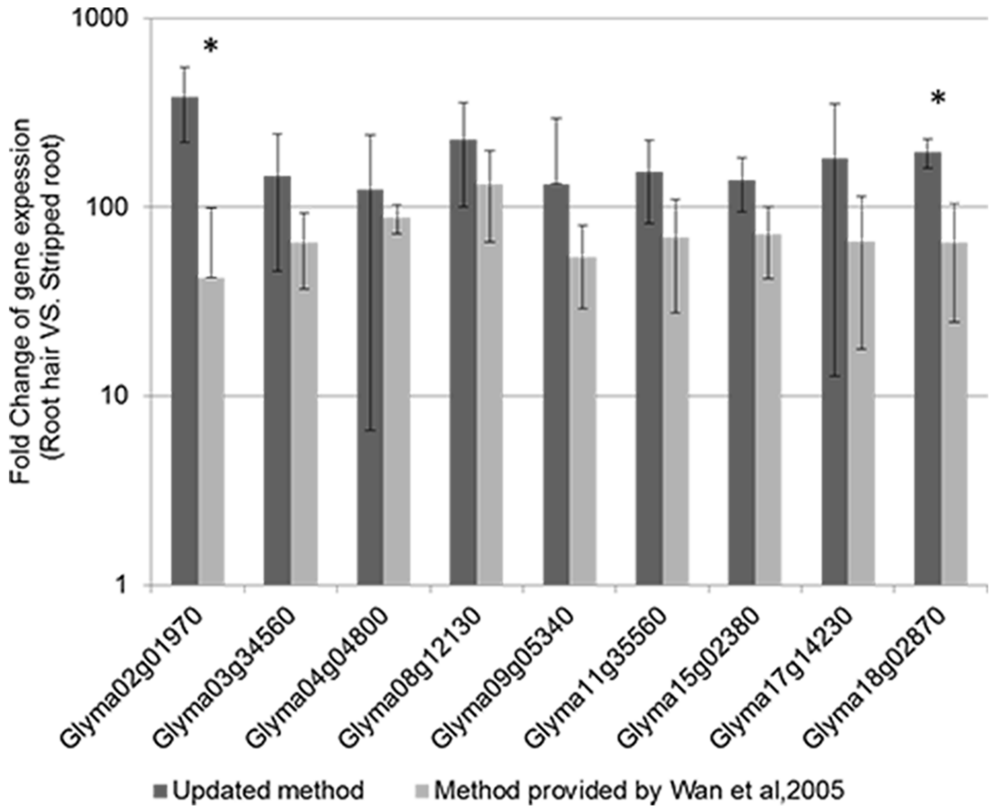


Figure 2- 4 Expression analyses of soybean root hair-specific genes. (A) Relative expression levels of nine soybean genes in root hair cells and stripped roots from the Illumina read data (Libault et al., 2010c); (B) The fold-change of the expression of nine root hair specific genes was quantified between isolated root hairs and stripped roots. The plant material was generated using our optimized protocol (dark bars) and the method provided by Wan et al. (2005) (light grey bars). For each experiment, a minimum of three biological replicates were performed and analyzed. The student t-test was applied to highlight significant differences between these two methods. The asterisk indicates significantly difference (*, $p < 0.05$).

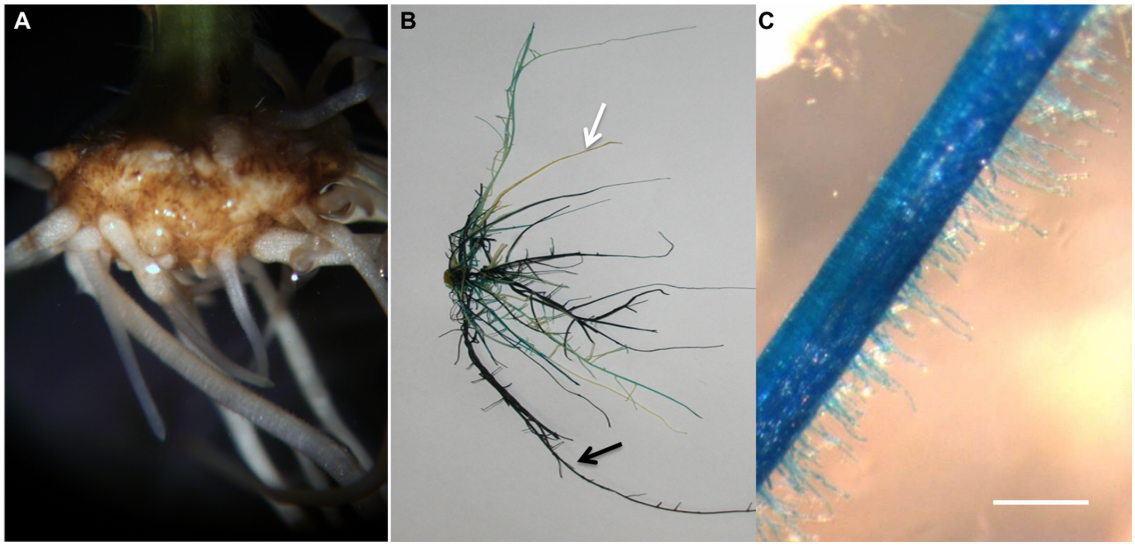


Figure 2- 5 Soybean transgenic roots and root hairs generated in the ultrasound areoponic system; (A) transgenic roots emerging from the callus 10 days after *A. rhizogenes* inoculation; (B) *GUS*-stained soybean root system, the black arrows point at the transgenic root and the white arrows point at the non-transgenic root; (C) *GUS*-stained transgenic root hair cells. Bar = 200 µm.

Table 2- 1 qRT-PCR primers

Soybean gene ID	Forward primer	Reverse primer
Glyma02g01970	tggctgcaaagtgaaaatga	tcaattcttcgtgccaatga
Glyma03g34560	atgagttggggcagtacgac	tagttgagcttgacgccaga
Glyma04g04800	ccaacggaacaaaggttgat	tatcggagcgtacatccaca
Glyma08g12130	gcccaacaaaggattaacga	tatcctccacatggcactca
Glyma09g05340	ggcatgacaagggctcatac	gcctgttccgttgttgt
Glyma11g35560	tgctacgtgaagcctgtt	agtggagcaccattgaga
Glyma15g02380	caaggtgaacctggagctgt	tctcccaacctctcaacgat
Glyma17g14230	cgtgatgaatgttggaggtg	gttgcaaatgcctggtatga
Glyma18g02870	gacccttagctttccgtcct	tctcaatgcatggtcaaagg

Chapter 3: A Comparative Genomic and Transcriptomic Analysis at the Level of Isolated Root Hair Cells Reveals New Conserved Root Hair Regulatory Elements

Authors: **Zhenzhen Qiao***, Lise Pingault*, Prince Zogli, Micaela Langevin, Niccole Rech, Andrew Farmer and Marc Libault

*These authors equally contributed to this work.

Publication status: This chapter was published in Plant Molecular Biology (DOI: 10.1007/s11103-017-0630-8), which allowed incorporation into this dissertation.

Author contributions:

Dr. Marc Libault designed the research. Micaela Langevin and I cloned soybean root hair promoters. Micaela Langevin, myself, Niccole Rech and Dr. Prince Zogli tested the soybean promoter activity in soybean, *Medicago* and *Arabidopsis*. I performed the *Arabidopsis* root hair specific genes data mining, qPCR primers design and qPCR assay. Dr. Marc Libault generated the evolution pathway. Dr. Lise Pingault and Dr. Andrew Farmer performed bioinformatics analysis and identified the *cis-element*. I performed the cloning of *cis-element* and tested in soybean plant. Dr. Marc Libault, Dr. Lise Pingault, and I wrote the article published in Plant Molecular Biology.

Abstract

Our understanding of the conservation and divergence of the expression patterns of genes between plant species is limited by the quality of the genomic and transcriptomic resources available. Specifically, the transcriptomes generated from plant organs are the reflection of the contribution of the different cell types composing the samples weighted by their relative abundances in the sample. These contributions can vary between plant species leading to the generation of datasets which are difficult to compare. To gain a deeper understanding of the evolution of gene transcription in and between plant species, we performed a comparative transcriptomic and genomic analysis at the level of one single plant cell type, the root hair cell, and between two model plants: *Arabidopsis* (*Arabidopsis thaliana*) and soybean (*Glycine max*). These two species, which diverged 90 million years ago, were selected as models based on the large amount of genomic and root hair transcriptomic information currently available. Our analysis revealed in detail the transcriptional divergence and conservation between soybean paralogs (i.e., the soybean genome is the product of two successive whole genome duplications) and between *Arabidopsis* and soybean orthologs in this single plant cell type. Taking advantage of this evolutionary study, we combined bioinformatics, molecular, cellular and microscopic tools to characterize plant promoter sequences and the discovery of two root hair regulatory elements (RHE1 and RHE2) consistently and specifically active in plant root hair cells.

Introduction

Following speciation and whole genome duplications (WGDs), plant genes are potentially lost or subject to neo- and sub-functionalization (Adams, 2007; Wendel et al., 2016). Various studies performed at the level of plant organs and tissues have revealed the fate of plant genes and, notably, their transcriptional evolution (Blanc and Wolfe, 2004a; Chaudhary et al., 2009; Galbraith and Birnbaum, 2006; Higgins et al., 2012; Roulin et al., 2013). While comparative genomic studies were clearly enhanced after the release of plant genome sequences (https://phytozome.jgi.doe.gov/pz/portal.html), comparative transcriptomic analyses are highly dependent of the quality of the samples used when establishing plant transcriptomes. For instance, the cellular complexities of plant organs, which vary between plant species (e.g., the differing anatomy of their root systems), remains a difficulty faced by scientists interested in revealing the transcriptional evolution of plant genes (Libault and Chen, 2016).

At the molecular level, the transcriptional signature of plant cell types depends, in part, on the pool of transcription factors in the cell and their interaction with transcriptional regulatory elements (REs) of the promoters. In plants, many REs have been characterized based on their role in controlling the expression of genes involved in plant response to biotic (Lota et al., 2013; Pontier et al., 2001) and abiotic stresses (Guevara-Garcia et al., 1998; Inaba et al., 2000; Inaba et al., 1999; Matsuura et al., 2013; Nagano et al., 2001; Petit et al., 2001; Safrany et al., 2008), as well as genes responding to hormone treatments (Hoth et al., 2010; Pla et al., 1993; Takeda et al., 1999) and genes controlling metabolic pathways (Dare et al., 2008; Kim et al., 2012). In

addition, cell type-specific REs have also been characterized (Abe et al., 2001; Alvarado et al., 2011; Becker et al., 2014a; Tsuwamoto and Harada, 2010; Weterings et al., 1995) including root hair REs (RHEs) (Kim et al., 2006b; Won et al., 2009). More recently, the emergence of genomic and transcriptomic resources led to the predictions of new plant REs (Choudhury and Lahiri, 2008; Ding et al., 2012; Gao et al., 2013; Haberer et al., 2004; Li et al., 2008; Satheesh et al., 2014; Vandepoele et al., 2009; Yilmaz et al., 2009; Zhang et al., 2005b).

The characterization of cell type-specific REs and their evolution after plant speciation and WGDs requires access to both genomic and cell-type specific transcriptomic information. Molecular analyses at the level of single cell types remain a major challenge in plant science due to the fact that each cell is physically bound to its neighboring cells via the cell wall preventing their isolation. For this reason, only a limited number of plant cell types are currently considered as models, including, pollen, trichomes, cotton fiber, and root hair cells (Arpat et al., 2004a; Becker et al., 2003; Brechenmacher et al., 2009a; Brechenmacher et al., 2012a; Hossain et al., 2015; Libault et al., 2010d; Libault et al., 2010h; Nguyen et al., 2012b; Qiao and Libault, 2013b; Schmutz et al., 2010b; Wang et al., 2010a). Among them, the root hair cell, which evolved 400 million years ago (MYA) (Kim et al., 2006b), is characterized by its tubular extension. This polar elongation maximizes the surface of interaction between the root system and the rhizosphere to enhance water and nutrient absorption. In addition, the legume root hair cell is also the first site of infection by rhizobium, the nitrogen-fixing symbiont involved in nodulation. Based on these essential functions, the root hair cell is now considered as a model to study cell determination and

differentiation, plant cell elongation, plant-microbe interaction and plant response to various abiotic stresses (Hossain et al., 2015; Libault et al., 2010d).

Recently, various -omic studies, including transcriptomic analyses, were conducted on root hair cells isolated from various plant species including soybean (*Glycine max*) and *Arabidopsis* (*Arabidopsis thaliana*). The soybean transcriptome atlas revealed 451 genes and unannotated loci preferentially expressed in root hair cells compared to root tip, root, nodule, leave, shoot apical meristem, flower and green pod (≥3-fold change between the expression levels in root hair and the second most highly expressed tissue) (Libault et al., 2010d). In *Arabidopsis*, multiple transcriptomic studies have been conducted to characterize the genes preferentially expressed in root hairs making the *Arabidopsis* root hair cell the most transcriptionally well-described single plant cell type (Becker et al., 2014a; Brady et al., 2007b; Bruex et al., 2012; Deal and Henikoff, 2010b; Jones et al., 2006; Lan et al., 2013).

In this manuscript, taking advantage of these transcriptomic resources and the access to genomic sequences, we describe a combination of computational and experimental approaches to analyze the evolution of the expression of plant genes preferentially expressed in root hairs and the conservation and divergence of the structure of their promoter activity and sequences. Specifically, we took advantage of the release of *Arabidopsis* and soybean transcriptomic and genomic datasets to reveal the sets of regulatory elements conserved between these plant species and conferring a preferential expression of plant genes in root hairs. Hypothesizing that RHEs playing a critical role in regulating the preferential expression of plant genes in root hair cells should be conserved even between evolutionary distant plant species, these analyses

were conducted under the context of the divergences between *Arabidopsis* (27,416 protein-coding loci) and soybean (56,044 protein-coding loci) whose last common ancestor is estimated at 90 MYA. In addition, to better evaluate the influence of plant WGDs on plant gene neo- and sub-functionalization, we also took in consideration the successive WGDs of the soybean genome during our analyses [i.e., the first duplication occurred ~56.5-59 MYA followed by the most recent WGD ~13 MYA (Lavin et al., 2005; Schmutz et al., 2014). Using microscopic approaches, we functionally validated the transcriptional activity of selected soybean promoter sequences across plant species including *Arabidopsis*. These results suggest a strong conservation of root hair transcriptional activity of plant genes during evolution. Hypothesizing that this conservation is driven by the presence of conserved RHEs, we combined bioinformatics and functional genomic analyses to correlate the root hair transcriptional activity of genes with the presence of conserved RHEs. Among them, we confirmed the root hair transcriptional activity of 2 newly discovered *cis*-regulatory elements in soybean and *Arabidopsis* plants.

Results and Discussion

Re-annotation of the soybean genes preferentially expressed in root hairs.

In a previous study, taking advantage of the release of the soybean genome sequence [*Glycine max* v1.0; (Schmutz et al., 2010b)], 357 soybean genes and 94 unannotated genomic loci were defined as preferentially expressed in root hairs (Libault et al., 2010d). The recent update of the annotations of the soybean genome (i.e., Wm82.a2.v1; http://phytozome.jgi.doe.gov/soybean/) allows us to revise their

annotation (see "Material and Methods" section for details) leading to the characterization of 363 Wm82.a2.v1 genes preferentially expressed in root hair cells (Supplemental Table 3-1). Including the 36 v1.0 soybean root hair genes, which are currently unannotated in the Wm82.a2.v1 soybean genome (Libault et al., 2010d; Schmutz et al., 2010b), we established a list of 399 soybean genes, which are preferentially expressed in root hair cells (Supplemental Table 3-1). Supporting their preferential expression in soybean root hair cells, several of these genes encode enzymes essential to root hair biological functions such as cell wall biosynthesis (e.g., pectate lyase, pectinesterase, xyloglucan/xyloglucosyl transferase, polygalacturonase, cellulase), polar cell elongation, vesicle trafficking (e.g., Pollen proteins Ole e I like, annexin), and transcriptional regulation of root hair genes expression [e.g., RSL2 (Lin et al., 2015)] (Supplemental Table 3-2, groups 4 to 11) (Libault et al., 2010b; Vijayakumar et al., 2016). For instance, we characterized three genes encoding predicted Ole e 1-like pollen proteins. Interestingly, among their *Arabidopsis* orthologs, which are also preferentially expressed in root hairs (Supplemental Table 3-2), one was recently functionally characterized as a major regulator of root hair cell elongation (Boron et al., 2014).

Gene loss and gene retention after soybean WGDs depend on the transcriptional patterns of paralogous genes.

Considering that the soybean genome is the product of two successive WGDs (Lavin et al., 2005; Schmutz et al., 2014), we hypothesized that a significant number of the 399 soybean genes preferentially expressed in root hairs are paralogs. To better

understand the evolution of the expression of these genes upon WGD, we analyzed their paralogous relationships based on microsynteny analyses using the CoGe application and its associated bioinformatics tools such as SynFind (Lyons and Freeling, 2008; Lyons et al., 2008b) (Supplemental Table 3-2; default parameters were used including a gene window size of 40 genes, a minimum number of genes of 4 and a collinear arrangement of the syntenic genes). We further refined our analysis by including the degree of paralogy existing between these genes in regards to the recent and ancient WGDs of the soybean genome (Schmutz et al., 2014) (Supplemental Table 3-2). Together, we characterized 143 out of the 399 soybean genes preferentially expressed in root hairs as singletons. The remaining 256 genes preferentially expressed in root hairs are divided into 201 paralogous groups composed by a total of 557 soybean paralogs, products of the recent and ancient WGDs (Supplemental Table 3-2). Among these groups, 103, 41 and 57 are composed by 2, 3 or 4 soybean paralogs, respectively (Figure 3-1). Specifically focusing on the recent soybean WGD, the 399 soybean genes preferentially expressed in root hair cells are divided in 445 groups of recent paralogs.

The soybean genome arose from two successive duplication events. In soybean, the loss rates of paralogous genes after the 13 and 56.5 MYA WGDs are estimated at 56.6% and 74.1%, respectively (Schmutz et al., 2010b). However, as previously reported, gene loss depends on the expression profile of the genes. For instance, highly and ubiquitously expressed plant house-keeping genes are preferentially lost consecutively to WGDs while genes controlling highly specific biological processes are preferentially retained in plant genomes after WGD (De Smet et al., 2013). To support the idea that highly specialized genes such as genes regulating root hair cell biology

should be preferentially retained after WGDs, we calculated the rate of lost genes after the ancient and recent WGDs in paralogous groups composed by at least one gene preferentially expressed in root hairs (i.e., number of missing/unidentified genes in each paralogous group containing at least one soybean genes preferentially expressed in root hair cells; Supplemental Table 3-2). Following the most recent soybean WGD (i.e., 13 MYA), the gene loss rate of these paralogous groups is 42.7% (190 out of 445 recent paralogous groups). Compared to the previously published overall loss rate of soybean paralogous genes [i.e., 56.6% after the recent WGD (Schmutz et al., 2010b)], our analysis confirms the higher retention rate of preferentially expressed root hair genes in paralogous groups (Chi-square test, p-value<0.013). Performing a similar analysis on the genes resulting from the ancient WGD, we observed a 70.6% gene loss rate [243 out of 344 paralogous groups (i.e., 143 singletons plus 201 paralogous groups)]. This rate is not significantly different compared to the previously published overall loss rate of soybean paralogous genes upon the ancient WGD [i.e., 74.1%, (Schmutz et al., 2010b)] (Chi-square test, p-value=0.55). In soybean, recent studies revealed a higher level of gene retention compared to other plant species (Tiley et al., 2016) potentially due to the unbiased fractionation (loss of functioning DNA sequence) (Langham et al., 2004) of the soybean genome upon WGDs (Garsmeur et al., 2014). Our analysis focusing specifically on genes preferentially expressed in root hairs provides additional information about gene retention and its dependence on the transcriptional pattern of genes. Our results support that specialized genes are retained during a longer period of time after WGDs compared to genes encoding proteins with broader function. We are

hypothesizing that the limited functional redundancy between these specialized genes drives their higher retention rate after WGD.

Hypothesizing that a higher number of genes preferentially expressed in root hairs per paralogous group could also influence paralog retention rate, we analyzed the qualitative and quantitative composition of the genes composing our 201 soybean paralogous groups (Supplemental Table 3-2). Among them, 151 (75.1%) include only one single gene preferentially expressed in root hairs and are composed by an average of 2.69 genes per group. The remaining 50 groups (24.9%) include at least 2 paralogs preferentially expressed in root hairs (i.e., 46, 3 and 1 groups composed by 2, 3 or all 4 paralogs preferentially expressed in root hair cells, respectively) and are represented by an average of 3.02 genes per group (Supplemental Table 3-2; groups 4-11). Based on this result, we conclude that gene retention rate significantly increases in paralogous groups composed by multiple specialized genes (i.e., preferentially expressed in root hairs) compared to paralogous groups containing only one specialized genes (Student's *t* test, p-value<2.2E-16). Based on this observation, we hypothesize that the preferential retention of plant paralogs helps to insure the plasticity of plant genomes and their evolution. This hypothesis suggests that large paralogous groups (e.g., composed by 4 paralogs) should be enriched in genes characterized by their different transcriptional profiles. Oppositely, smaller paralogous groups (e.g., composed by only 2 paralogs) should be composed of genes sharing a higher level of transcriptional redundancy. To verify this hypothesis, we calculated the percentage of genes preferentially and non-preferentially expressed in root hairs in paralogous groups composed by 2 and 4 paralogs. The majority of the groups composed of 2 soybean

paralogs are preferentially expressed in root hairs (>60%; Figure 3-1). Oppositely, the percentage of genes preferentially expressed in root hairs strongly decreases in groups composed of 4 paralogs (<36%; Figure 3-1). Taken together, these results suggest that soybean WGDs promotes the transcriptional evolution of paralogous genes and that this evolution depends on the level of gene retention upon WGDs.

Transcriptional evolution between soybean paralogous genes at the level of the root hair cell.

Upon WGDs, retained paralogs can become pseudogenes or functionally diverged leading to their sub- or neo-functionalization through, among other evolutionary paths, a change of their expression patterns (Adams, 2007; Wendel et al., 2016). The rate of transcriptional divergence of the soybean genes after WGDs has been previously evaluated as 50% upon comparative transcriptomic analysis (Roulin et al., 2013). Here, we confirm this rate of divergence upon analysis of the transcriptional profile of our 557 paralogous genes [i.e., 256 (46%) and 301 (54%) are preferentially and non-preferentially expressed in soybean root hairs, respectively].

To further validate the evolutionary fate of the expression of soybean root hair paralogs following WGDs (Libault et al., 2010d), we quantified the expression of 50 soybean genes from 15 paralogous groups using quantitative reverse transcription-PCR (qRT-PCR) on isolated root hair cells and stripped roots (i.e., roots devoid in root hairs). Among these genes, 33 and 17 are preferentially and non-preferentially expressed in root hair cells according to previously published RNA-seq datasets (Libault et al., 2010d) (Figure 3-2). Across 3 independent biological replicates, our qRT-PCR assay

validates the preferential root hair expression for 28 out of the 33 root hair genes and the non-preferential root hair expression of 12 out of the 17 soybean genes (fold-change>3; p-value<0.05; Figure 3-2). Consistent with the concept that WGD triggers the neo- and sub-functionalization of paralogs (Adams, 2007; Wendel et al., 2016), our comparative genomic and transcriptomic analysis clearly highlights the evolutionary fate of soybean paralogous genes following WGDs at the level of the root hair cell including gene loss and the transcriptional divergence of the retained genes.

Comparative genomic and transcriptomic analysis between soybean and Arabidopsis genes preferentially expressed in root hairs.

To better understand the evolution of the expression of soybean paralogs in root hair cells, we analyzed the conservation and divergence of their expression pattern into a broader evolutionary context. To date, *Arabidopsis* offers the most complete transcriptomic information on plant root hair cells (Becker et al., 2014a; Brady et al., 2007b; Bruex et al., 2012; Deal and Henikoff, 2010b; Jones et al., 2006; Lan et al., 2013). In addition to these transcriptomic resources, the speciation between soybean and *Arabidopsis* which occurred 90 MYA (Yang et al., 1999) allows the characterization of a large number of microsyntenic blocks between these two plant species facilitating the identification of orthologs and the comparison of their transcriptional patterns (Supplemental Figure 3-1). At the level of the entire genome, 34.5% of the soybean genes have at least one *Arabidopsis* ortholog based on the CoGE database.

Among the 27416 predicted *Arabidopsis* protein-coding genes (https://phytozome.jgi.doe.gov/pz/portal.html; http://www.arabidopsis.org/; TAIR10), 2982 (10.9%) were characterized as preferentially expressed in root hair from six independent transcriptomic analyses including 766 (2.8%) repetitively identified in at least two independent studies (Supplemental Table 3-3; Supplemental Figure 3-2). These *Arabidopsis* root hair-preferential genes were characterized upon comparative transcriptional analyses between root hair and plant tissues and cell types [(Becker et al., 2014a), (Lan et al., 2013), (Brady et al., 2007b), (Deal and Henikoff, 2010b); lower bound of fold-change >1.2, 2, 1.2 and 1.3, respectively] or between wild-type and root hairless mutant plants [(Bruex et al., 2012), (Jones et al., 2006); lower bound of fold-change >2 in both studies].

Taking advantage of these transcriptomic and genomic resources, we first characterized the orthologous relationship existing between soybean genes preferentially expressed in root hairs and *Arabidopsis* genes. Based on microsyntenic analyses, we identified a total of 216 *Arabidopsis* genes orthologous with 174 soybean genes preferentially expressed in root hairs (43.6% of our initial pool of 399 soybean genes preferentially expressed in root hairs). Among them, 42 are orthologous to 26 soybean root hair singletons while the remaining *Arabidopsis* genes are orthologous to 336 soybean genes (148 preferentially and 188 non-preferentially expressed in root hairs) and belong to 112 root hair paralogous groups (Supplemental Table 3-2). We did not identify *Arabidopsis* orthologs for the remaining 225 soybean genes preferentially expressed in root hairs. To reveal the transcriptomic evolution between plant orthologous in root hair cells, we integrated comparative genomic study with the

published *Arabidopsis* and soybean root hair transcriptomes. Among the 216 *Arabidopsis* genes orthologous to soybean root hair genes, 62 (28.7%) are preferentially expressed in root hairs including 40 genes (18.5%) repeatedly characterized as preferentially expressed in root hairs in independent transcriptomic analyses (Supplemental Table 3-3). Compared to the 2982 *Arabidopsis* root hair genes (10.9% of the 27,416 predicted protein-coding loci), our results reveals a significant enrichment in *Arabidopsis* and soybean orthologous genes preferentially co-expressed in root hairs (Chi-square, p-value = 4.7E-12). This result suggests that, before speciation, a significant subset of the ancestral genes to the soybean and *Arabidopsis* root hair genes were also preferentially expressed in root hair cells.

To better evaluate the transcriptional evolution of the soybean paralogous genes in root hair cells after plant speciation and WGDs, we categorized the 557 soybean and 216 *Arabidopsis* orthologs and paralogs characterized in this study into 14 different evolutionary pathways retracing the gain or loss of their preferential expression in soybean root hairs (Figure 3-3). Taking into consideration only the unambiguous evolutionary fate of the genes, we calculated the percentage of divergence and conservation of the preferential expression of soybean paralogs in root hair cells. Following the most ancient soybean WGD, a larger fraction of the soybean paralogs transcriptionally diverged (evolutionary pathways #2, 6, 7 and 8; 62.21%; Figure 3-3, long blue arrows). More rarely, soybean paralogs shared similar expression profiles (evolutionary pathways #3, 4, 5, 9, 10 and 11; 37.79%; Figure 3-3, long grey arrows). Performing a similar analysis on the genes resulting from the second and more recent soybean WGD, their transcriptional profiles were preferentially conserved

(evolutionary pathways #1, 2, 3, 4, 5, 6, 7 and 8; 57.97%; Figure 3-3, short grey arrows). A smaller fraction of the two recent paralogs shared different expression profiles preferentially diverged (evolutionary pathways #1, 2, 3, 9, 10 and 11; 42.03%; Figure 3-3, short blue arrows). As expected, the transcriptional divergence between paralogs often affects only one of the two pairs of recently diverged paralogs (evolutionary pathways #1, 2 and 3; 94.4%) and rarely the two sets of paralogs (evolutionary pathways #9, 10 and 11; 5.6%) (Figure 3-3). Those results confirm that, upon successive WGDs, the increase of the number of paralogous genes triggers their transcriptional divergence.

Conservation of the root hair activity of soybean promoters between plant species.

Our comparative genomic and transcriptomic analyses revealed numerous soybean and *Arabidopsis* paralogous and orthologous genes characterized by their preferential expression in root hairs. We hypothesize that the conservation of these preferential expression patterns results from the conservation of plant RHEs after plant speciation and WGDs. Based on this hypothesis, we expect that the preferential root hair transcriptional activity of a soybean promoter should be conserved in other plant species such as the non-legume plant *Arabidopsis*. In order to validate this hypothesis, we analyzed the promoter activity of several soybean genes preferentially expressed in root hairs and tested the potential conservation of their transcriptional activity in *Arabidopsis*. Therefore, we cloned around 2,000 bp promoter sequences of 6 soybean genes preferentially expressed in root hair (i.e., Glyma.08G115000, Glyma.15G020700, Glyma.18G025200, Glyma.03G188300, Glyma.12G012500, and Glyma.19G226300)

upstream to the *UidA* reporter gene encoding the β-glucuronidase (GUS assay). Among them, 3 genes have low expression levels in root hairs (i.e., Glyma.08G115000, Glyma.12G012500 and Glyma.19G226300 which belong to the evolutionary pathways 14, 8 and 1) and 3 genes are characterized by their high transcriptional activities in root hairs (i.e., Glyma.15G020700, Glyma.18G025200, Glyma.03G188300 which belong to the evolutionary pathways 14, 6 and 8) (Supplemental Figure 3-3; Supplemental Table 3-4).

To validate the root hair activity of these promoters, we applied the soybean hairy root transformation method to generate soybean composite plants. Upon transformation, we detected β-glucuronidase activity for 4 out of the 6 promoter-*UidA* transcriptional fusions (Figure 3-4). While the promoter sequences of Glyma.08G115000, Glyma.15G020700, Glyma.18G025200 and Glyma.03G188300 are exclusively driving the expression of the *UidA* reporter in soybean root hairs (Figure 3-4), we did not detect any activity for the Glyma.12G012500 and Glyma.19G226300 promoter sequences. We hypothesize that the low transcriptional activity of these two genes limits the detection of the β-glucuronidase activity in our assay (Supplemental Figure 3-3) or that critical regulatory elements are missing in their cloned promoter sequences.

To test the potential conservation of the activity of the 4 soybean promoters in *Arabidopsis* root hair cells, we applied the floral dip transformation assay on *Arabidopsis*. Upon staining, we observed that the 3 out of 4 soybean root hair promoters (i.e., Glyma.18G025200, Glyma.15G020700 and Glyma.08G115000) tested also drive the expression of the *UidA* reporter gene in *Arabidopsis* root hair cells (Figure 3-5). The

conservation of the transcriptional activity of the Glyma.18G025200 promoter in *Arabidopsis* root hairs is reinforced by the characterization of *Arabidopsis* orthologs preferentially expressed in root hair cells (Figure 3-3, Supplemental Table 3-2, pathway 6). Oppositely, similar evolutionary evidences did not support the maintenance of the transcriptional activity of the soybean Glyma.15G020700 and Glyma.08G115000 promoters in *Arabidopsis* root hairs because *Arabidopsis* orthologs to the Glyma.15G020700 and Glyma.08G115000 genes were not identified in our study (Figure 3-3, Supplemental Table 3-2, pathway 14). These results support the strong conservation of the cellular specificity of the activity of plant promoter sequences between plant species and suggest that the molecular mechanisms (e.g., binding of transcription factors to regulatory elements) controlling gene transcriptional activity in root hairs are maintained upon plant speciation.

Comparative analysis of the promoter sequence of soybean and Arabidopsis genes preferentially expressed in root hairs revealed conserved DNA motifs.

The conservation of the transcriptional patterns between plant paralogs and orthologs (Supplemental Table 3-2) as well as the conservation of the root hair activity of soybean promoters in *Arabidopsis* (Figure 3-5) suggest that RHEs are highly conserved between plant genes preferentially expressed in root hairs. Hence, we hypothesize that the conservation, gain or loss of the root hair transcriptional activity of genes is directly related to the presence or the absence of these RHEs. To identify these RHEs, we performed a motif analysis on the promoter sequence of the 248 and 377 soybean paralogous and *Arabidopsis* orthologous genes preferentially and non-

preferentially expressed in root hair cells, respectively (Supplemental Table 3-2; evolutionary pathways 1 to 11). Using the MEME package (Bailey and Elkan, 1994), we asked for the identification of 20 conserved and enriched motifs in the promoter sequences of soybean and *Arabidopsis* genes preferentially expressed in root hairs (e-value < $1e10^{-1}$). Using the FIMO package (Grant, Bailey, and Noble 2011), we looked at their respective enrichment in the promoter sequences of genes preferentially *versus* non-preferentially expressed in root hairs (fold-change >3). This analysis allowed the identification of 2 conserved motifs named RHE1 and RHE2 (Figure 3-6A), which are both frequently represented (≥ 20%) in soybean and *Arabidopsis* genes preferentially expressed in root hairs.

Accordingly, we selected consensus sequences for the RHE1 and RHE2 motifs to analyze their transcriptional activity in plant root hair cells (Figure 3-6). To validate these activities, we synthesized gateway compatible versions of these *cis*-regulatory elements to clone them upstream to the minimal 35S promoter, a promoter sequence which has minimal activity unless *cis*-elements are nearby (Benfey et al., 1989), and upstream to the *UidA* reporter gene. The transgenes were used to transform soybean plants using the hairy root transformation method and *Arabidopsis* plants using the floral dip transformation method (Clough and Bent, 1998). In soybean, while the minimal 35S only show minimal activity in the vascular tissues (Figure 3-6f), we observed a specific transcriptional activity of the transgenes driven by RHE1 and 2 containing promoters in soybean root hair cells (Figure 3-6b and d, black arrows). In *Arabidopsis*, we did not detect any activity of the minimal 35S as previously reported (Figure 3-6g) (Benfey et al., 1989). However, upon transformation with the RHE1-

driven minimal 35S promoter, we observed a β-glucuronidase activity specifically in *Arabidopsis* root hairs (Figure 3-6c, black arrow). This result confirmed the conservation of the transcriptional activity of RHE1 in the root hair cells of evolutionary distant plant species. The cloning of the RHE2 motif upstream to the minimal 35S promoter led to the detection of a broader β-glucuronidase activity in *Arabidopsis* compared to soybean: the reporter was detected in both root hair and non-root hair cells (Figure 3-6e, black arrow). In addition, we also noticed that RHE2 was also capable to confer a more specific activity in emerging root hair cells (Figure 3-6e, grey arrow). The differential transcriptional activity of RHE2 in soybean and *Arabidopsis* roots might be the reflection of the differential differentiation process of root hair cells between the two plant species. The conservation of the root hair activity of RHE1 and 2 between soybean and *Arabidopsis* support the idea that, upon speciation, the transcriptional regulatory mechanisms are conserved between species. Our results also reveal the transcriptional evolution of the root hair specificity of RHE2 upon speciation between soybean and *Arabidopsis*. Taken together, we hypothesize that the level of specificity of the expression of plant genes in root hairs is gained through the combination of different RHEs including highly and less specific RHEs, such as RHE1 and RHE2, respectively. This hypothesis is supported by Haberer et al. (2004) who described that the neo- and sub-functionalization of *Arabidopsis* duplicated genes was limited to presence or absence of only several regulatory elements (Haberer et al., 2004). Together, these results also validate our subtractive analysis of promoter motifs between root hair and non-root hair specific promoter sequences to characterize new RHEs.

The previous studies reported one RHE has been functionally characterized for its role in driving the expression of multiple genes in root hair cells in different plant species (e.g., AtEXP7, AtEXP16 and their orthologs from different plant species, AtLRX1, AtLRX2 and AtPRP3) (Kim et al., 2006b; Won et al., 2009). The functional conservation of this RHE in the promoter sequence of several *Arabidopsis* genes suggests the co-option of this regulatory element to drive gene expression in root hair cells. In our study, we revealed the conservation of RHE1and 2 between soybean and *Arabidopsis*. The comparison of the nucleotidic sequences between the 2 RHEs identified in this study (RHE1 and 2) and the different variants of the first discovered RHE sequence (Kim et al., 2006b; Won et al., 2009) did not reveal any significant similarity (lowest p-value=0.053; score=18.3). Likewise, this first characterized RHE was not identified in the promoter sequence of the soybean genes preferentially expressed in root hair cells. Taken together, these results suggest that RHE1 and 2 are novel root hair *cis*-regulatory elements. Based on the broad conservation of their transcriptional activity in plant root hair cells, we are assuming that RHE1 and 2 co-evolved in and between plant species as previously mentioned by Kim et al. (2006). Our analysis also highlights the diversity of these RHEs to better regulate the transcriptional activity of genes in root hair cells.

Conclusion

The root hair cell has been used for over a decade as a plant single cell type model to characterize the role of transcription factors and the identification of regulatory elements (Kim et al., 2006b; Li and Lan, 2015; Won et al., 2009). In our

study, we hypothesized that several RHEs might be broadly conserved in and between plant species, upon speciation and WGDs. To verify this hypothesis, we took advantage of the release of the soybean and *Arabidopsis* genome sequences and root hair transcriptomes to provide a detailed picture of the evolution of the expression of paralogous and orthologous root hair genes. Performing a comparative analysis of the promoter sequences of these genes, we also clarified the evolutionary relationship existing between the transcriptional activity of genes in root hairs and the presence of RHEs. Specifically, upon integration of the soybean and *Arabidopsis* transcriptomic and genomic information, we revealed the conservation of RHE 1 and 2 between soybean and *Arabidopsis*. The development of sequencing technologies allowing an easier access to plant single transcriptome will likely enhance our understanding of the evolution of plant transcriptomes.

Experimental Procedures:

Bacterial cultures

Escherichia coli (DH5a), *Agrobacterium rhizogenes* (K599) and *Agrobacterium tumefaciens* (GV3101) were grown and washed as previously described (Libault et al., 2009a; Libault et al., 2010h).

Cloning

The cloning of six soybean promoter sequences (i.e. 2000 bp of nucleotide sequences located upstream to the first codon encoding soybean proteins; Supplemental Figure 3-4) upstream to the GUS cDNA was performed as previously described (Libault

et al., 2010h). Primers used to amplify these promoters with the Gateway compatible AttB boxes using soybean Williams 82 genomic DNA as template are listed in Supplemental Table 3-6.

Using the Gateway® BP and LR Clonase® II enzyme mixes (Invitrogen, Carlsbad, CA, USA), the soybean promoter fragments were introduced into the pDONR-Zeo vector (Invitrogen, Carlsbad, CA) then into the pYXT1 destination vectors carrying the GUS reporter gene (Xiao et al., 2005). The root hair promoter-pYXT1 vectors were used to transform *A. rhizogenes* (strain K599) as well as *A. tumefaciens* (strain *GV3101)*.

Plant transformation and microscopy

Soybean hairy root transformations was performed as described by Libault et al. (2009) using A. rhizogenes strains carrying the six root hair promoter-GUS binary vectors. Two week-old soybean shoots were cut then placed into rock-wool cubes (Fibrgro, Sarnia, Canada) and inoculated with five ml of *A. rhizogenes* carrying the construct of interest (OD_{600}=0.3). After 3 days (23°C, 50% humidity, long day conditions), the plants were watered with deionized water. One week later, the plants were transferred to vermiculite: perlite mix (3:1) wetted with nitrogen free-plant nutrient solution (Lullien et al., 1987). After growing ten days, the shoots were transferred to the ultrasound aeroponic system under short day conditions (Qiao and Libault, 2013). Four weeks later, the hairy roots were collected then stained and fixed to reveal the β-glucuronidase activity of the soybean root hair promoters as described by Libault et al. (2010e). To express our transgenes in *Arabidopsis*, *Agrobacterium*

tumefaciens stain GV3101 cultures harboring the various promoter:: GUS constructs were transformed into *Arabidopsis* by floral dip (Clough and Bent, 1998) to generate stably transformed plants. The GUS localization was observed with a MZ 10 F fluorescence research stereo microscope (Leica, Germany) coupled with a Retiga 4000DC FAST 1394 camera system.

RNA extraction, DNase treatment, reverse transcription, quantitative PCR primer design, quantitative PCR reaction conditions and data analysis

Total RNA isolations, DNase treatments, reverse transcription and quantitative PCR primer design were performed as described by Libault et al. (2010e). The primers used in this study were designed with primer3 software (http://biotools.umassmed.edu/bioapps/primer3_www.cgi) using the criteria described by Libault et al. (2010e). The primer list used for quantifying gene expression by qRT-PCR is available in the Supplemental Table 3-5.

The qRT-PCR reactions were performed as described by Libault et al. (2010e) using the geometrical average of *Cons4*, *6* and *7* expressions to normalize the expression levels of root hair genes preferentially expressed in root hairs. Fold-changes of gene expression between root hair and stripped root samples were calculated for the three different biological replicates. The average of these fold-changes is represented. To statistically validate the differences, a Student's t-test between the three root hair samples and the three stripped root samples expression levels was calculated.

Bioinformatic analyses

Updated annotation of the soybean genes preferentially expressed in root hair cells

To update the annotation of the soybean genes preferentially expressed in root hairs, we applied three different analyses: 1- comparison of the gene annotations between the v1.0 and Wm82.a2.v1 as reported on Phytozome v10.3 (279 updated annotations; Supplemental Table 3-1); 2- blast (E-value ≤ 10^{-10}) and parseblast (>99% of identity) analyses of the v1.0 root hair cDNA sequences against the Wm82.a2.v1 transcript sequences (42 updated annotation; Supplemental Table 3-1); 3-mapping against the Wm82.a2.v1 genome sequence of the RNA-seq reads matching the 94 v1.0 unannotated loci preferentially transcribed in root hairs. The latter led to the mappings of RNA-seq reads of 90 unannotated v1.0 loci against the Wm82.a2.v1 soybean genome (Supplemental Table 3-1). Among them, 74 genomic loci overlap 68 soybean Wm82.a2.v1 genes not previously annotated in the *Glycine max* v1.0 genome (Schmutz et al., 2014). The remaining 16 genomic Wm82.a2.v1 loci do not match predicted genes (Supplemental Table 3-1).

Bioinformatic characterization of regulatory elements of genes preferentially expressed in root hairs

To characterize conserved *cis*-regulatory elements between soybean and *Arabidopsis* orthologous and paralogous genes preferentially expressed in root hairs, 2 kb sequences upstream and downstream to the first and last codons, respectively, were isolated from the soybean and *Arabidopsis* genomes. The positions of these codons

were identified according to the current gene annotation files (https://phytozome.jgi.doe.gov/pz/portal.html; https://www.arabidopsis.org/).

Using MEME v4.10.2 (Timothy L. Bailey et al. 2009), we looked for conserved *cis*-regulatory elements present in root hair promoter sequences (Supplemental Table 3-2) with the following parameters: mod zoops, nmotifs 10, minw 6, maxw 50, revcomp, evalue 10^{-6}. MEME root hair motifs were scanned using FIMO (Grant, Bailey, and Noble 2011).

Figures and Tables

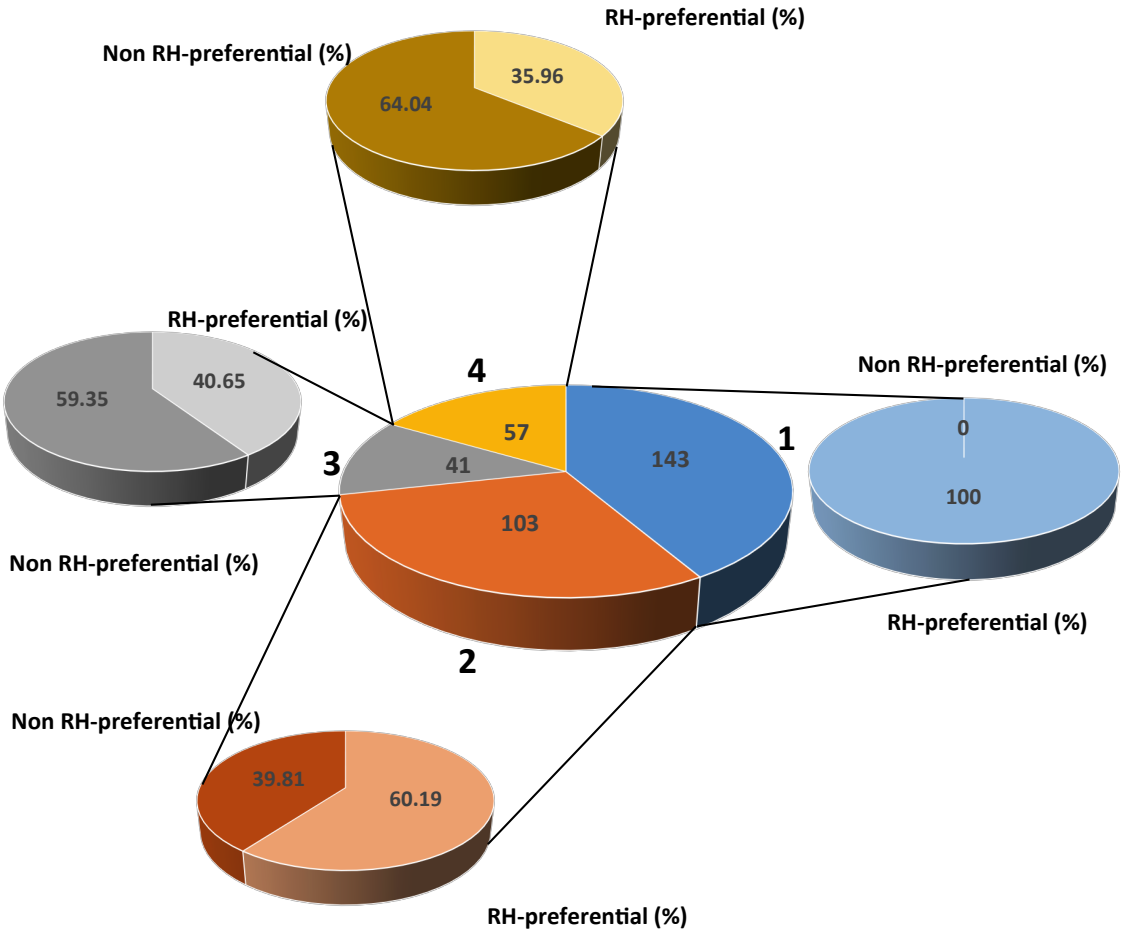


Figure 3- 1 Transcriptional evolution of soybean paralogs in root hair cells. The central pie chart highlights the number of soybean paralogous groups containing at least one gene preferentially expressed in root hairs (blue: single-gene groups; orange: 2 paralogs; grey: 3 paralogs; yellow: 4 paralogs). Peripheral pie charts describe the percentage of genes preferentially (light color) and non-preferentially expressed in root hairs (dark colors) for each category of the paralogous group.

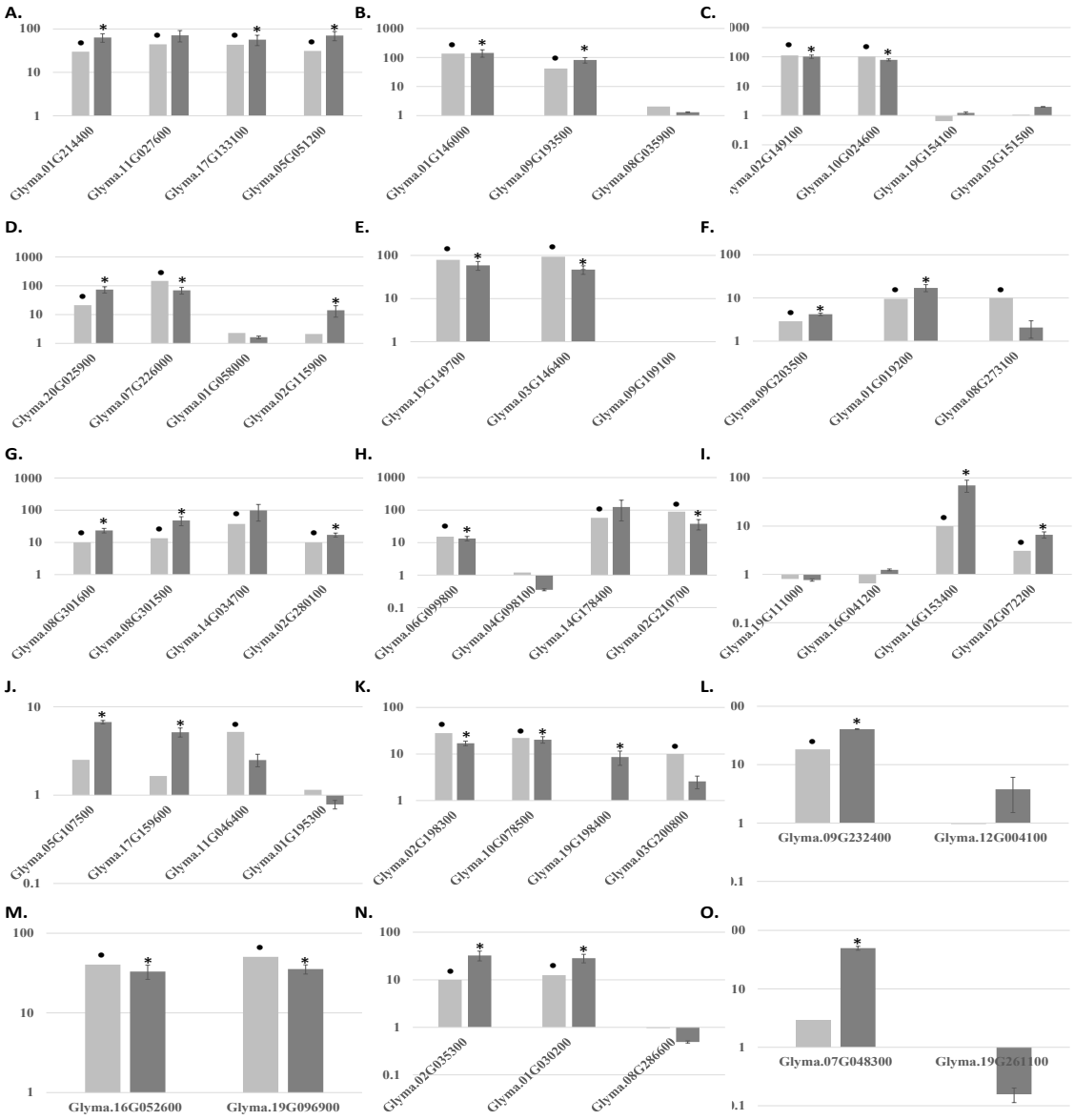


Figure 3- 2 Validation of the preferential and non-preferential expression of 50 soybean genes in root hair cells. These 50 genes are distributed in 15 paralogous groups composed by 2 to 4 paralogs (A-O). The fold-change of expression between root hair and stripped roots (i.e., root devoid in root hairs; y-axis; logarithmic scale) was calculated according to previously published RNA-seq datasets (light grey, Libault et al., 2010d) and by qRT-PCR across three independent biological replicates (dark grey). The soybean genes preferentially expressed in root hairs and revealed by RNA-seq technology according to Libault et al. (2010d) are highlighted with a black dot. Significant differential expression revealed by qRT-PCR across three independent replicates are highlighted with asterisks (i.e., fold-change root hair versus stripped root expression levels >3; Student's *t* test p-value < 0.05). Standard errors are indicated for the qRT-PCR assay.

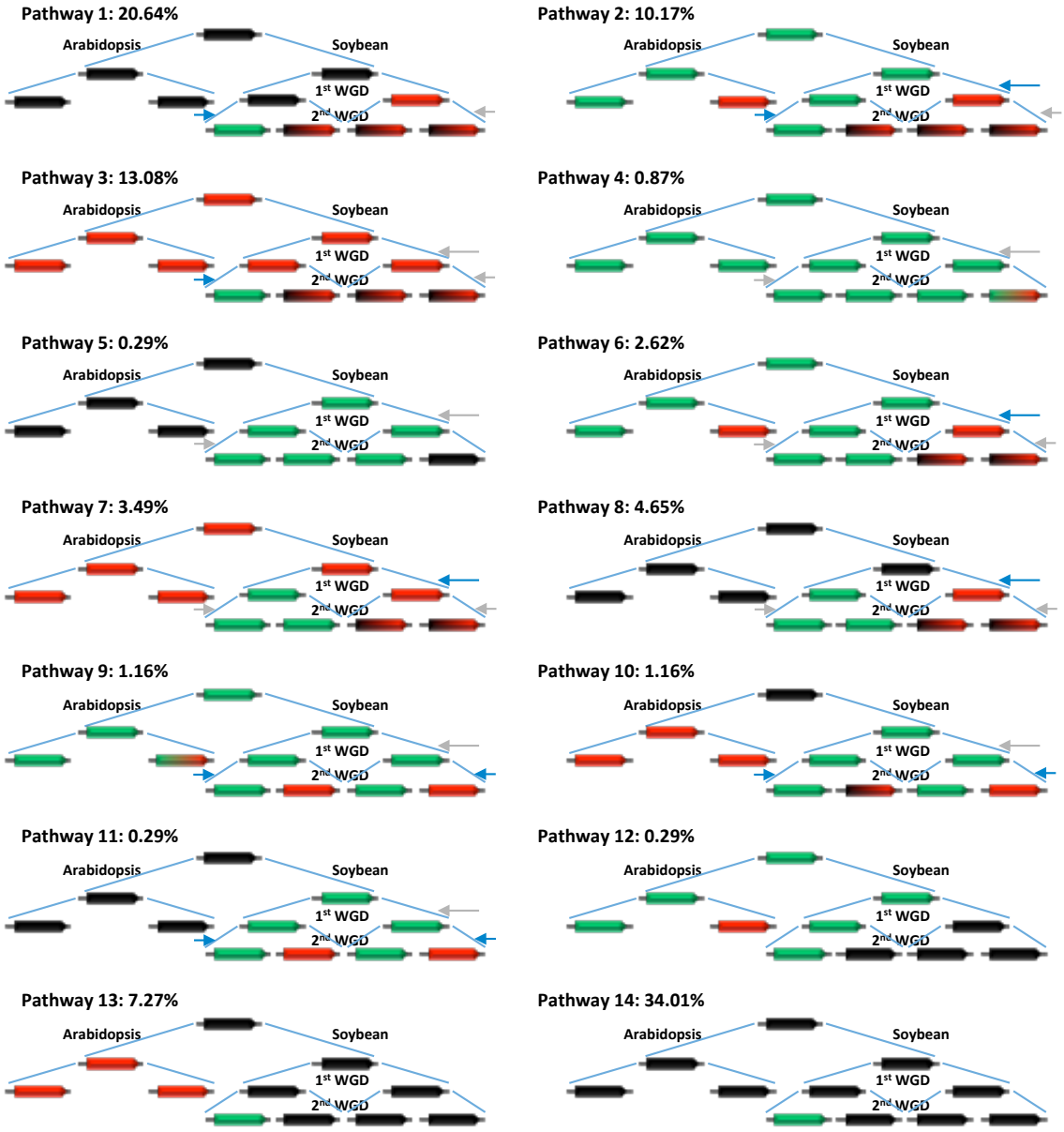


Figure 3- 3 Evolution of the expression of soybean and *Arabidopsis* genes preferentially expressed in root hairs. Fourteen different profiles of transcriptional evolution are proposed taking into consideration the ancient and recent soybean WGDs (1st WGD and 2nd WGD, respectively; see Supplemental Tables 3-1 and 2 for details). Genes preferentially and non-preferentially expressed in root hair cells are highlighted in green and red, respectively. Genes which have an undetermined root hair transcriptional activity are indicated in black. Gradients reflect various possibilities in the root hair expression profile of genes. Based on this study, a percentage of occurrence is included in each profile. Arrows highlight the transcription divergence (blue arrows) or conservation (grey arrows) between recent (short arrows) and ancient (long arrows) soybean paralogs.

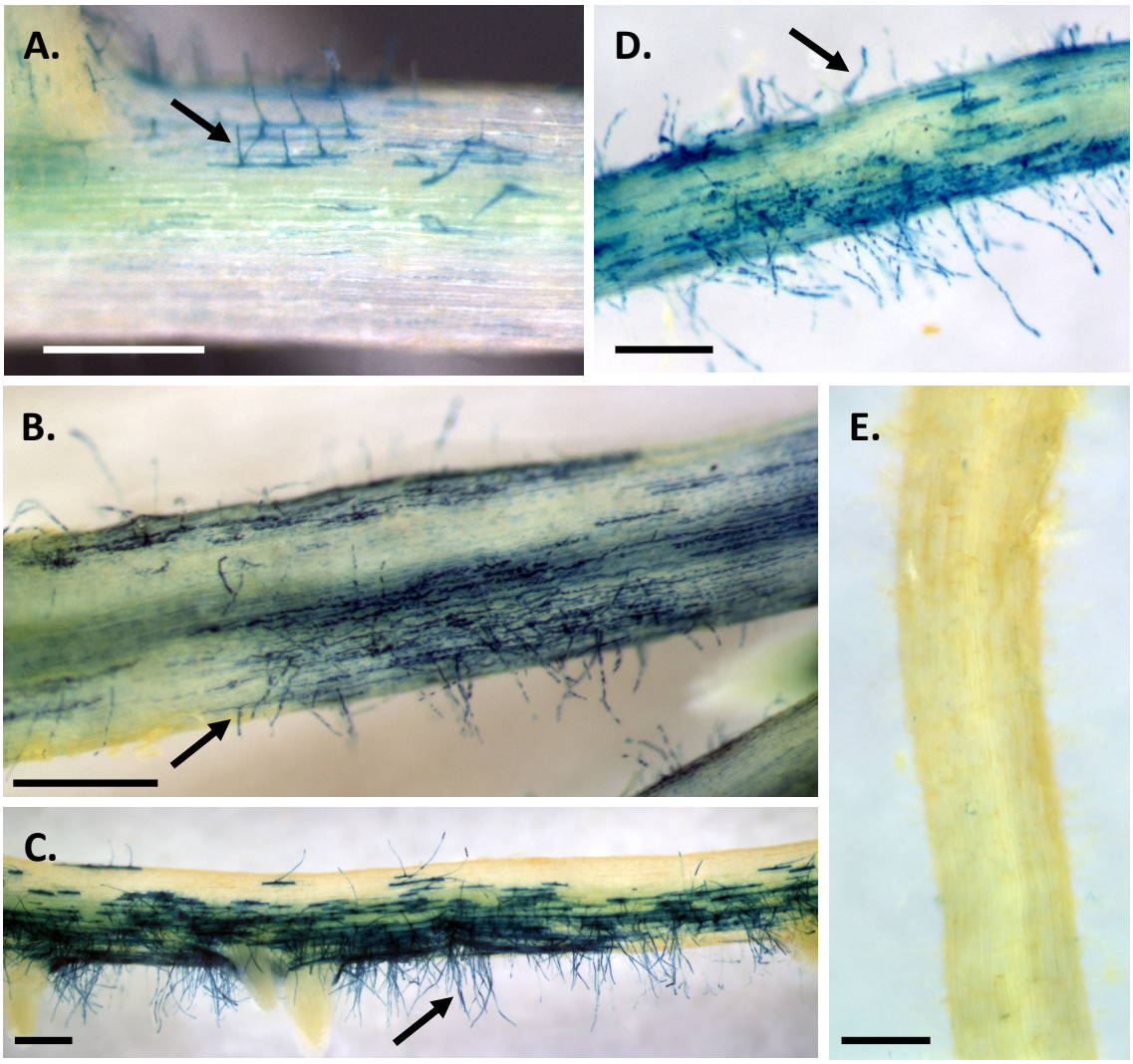


Figure 3- 4 Transcriptional activity of four soybean promoter sequences in soybean hair cells. The β-glucuronidase activity (black arrows) was used as a reporter of the transcriptional activity of the promoter sequence from four different soybean genes (A: Glyma.18G025200; B: Glyma.15G020700; C: Glyma.08G115000; D: Glyma.03G188300). The pYXT1 binary vector was used as a negative control (E). Bar = 0.5 mm.

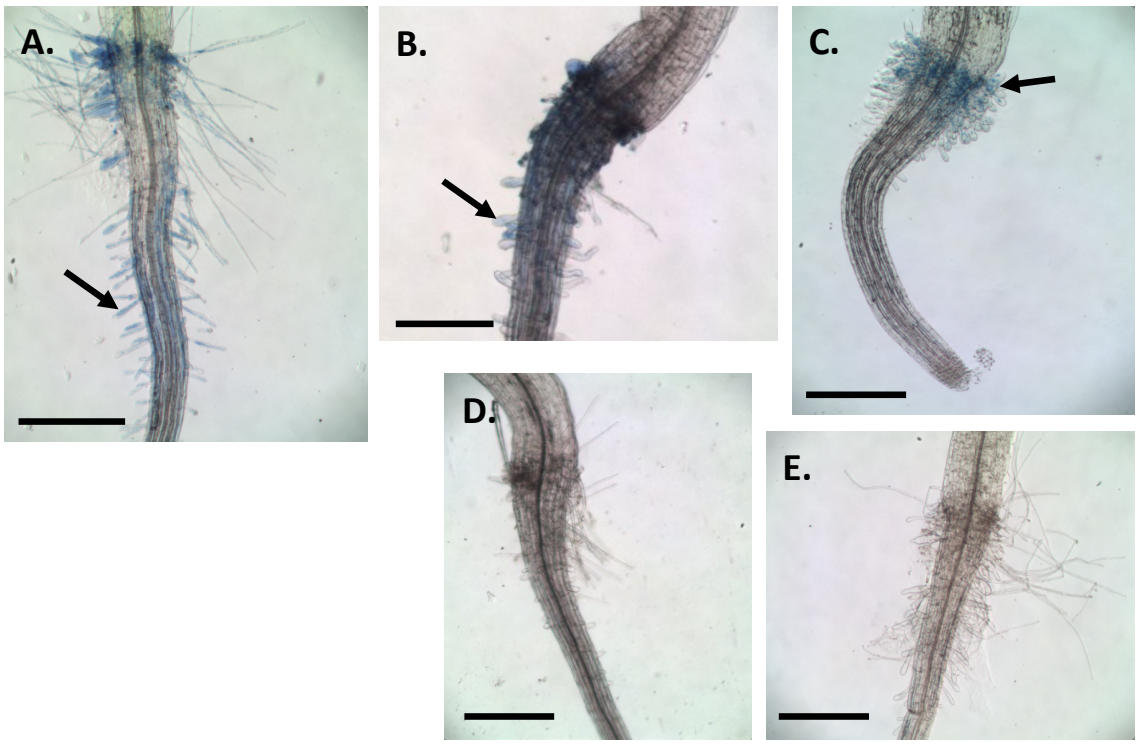


Figure 3- 5 Transcriptional activity of four soybean promoter sequences in *Arabidopsis* root hair cells. The β-glucuronidase activity (black arrows) was used as a reporter of the transcriptional activity of the promoter sequence from four different soybean genes (A: Glyma.18G025200; B: Glyma.15G020700; C: Glyma.08G115000; D: Glyma.03G188300). The pYXT1 binary vector was used as a negative control (E). Bar = 0.5 mm.

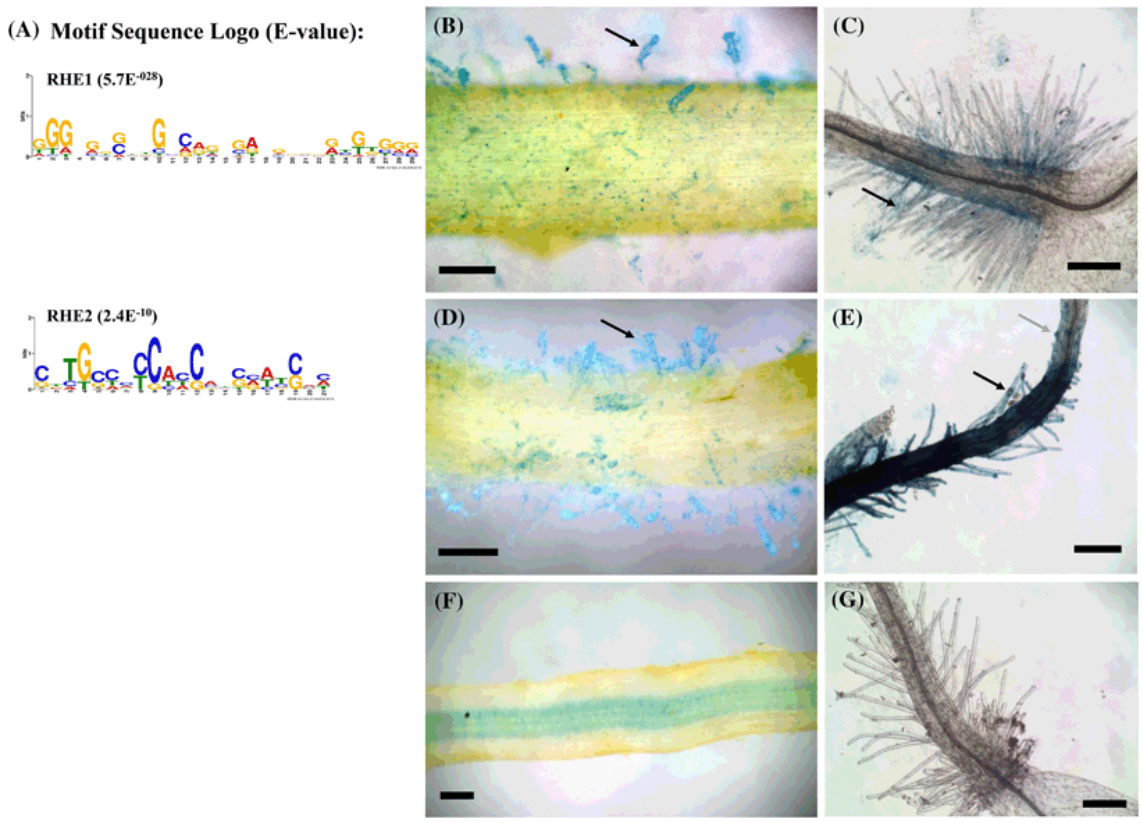
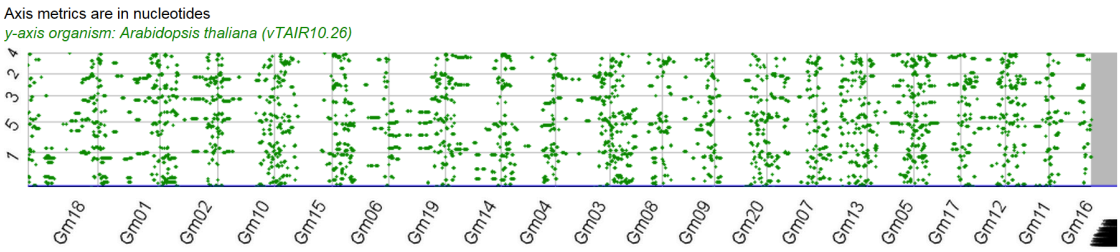
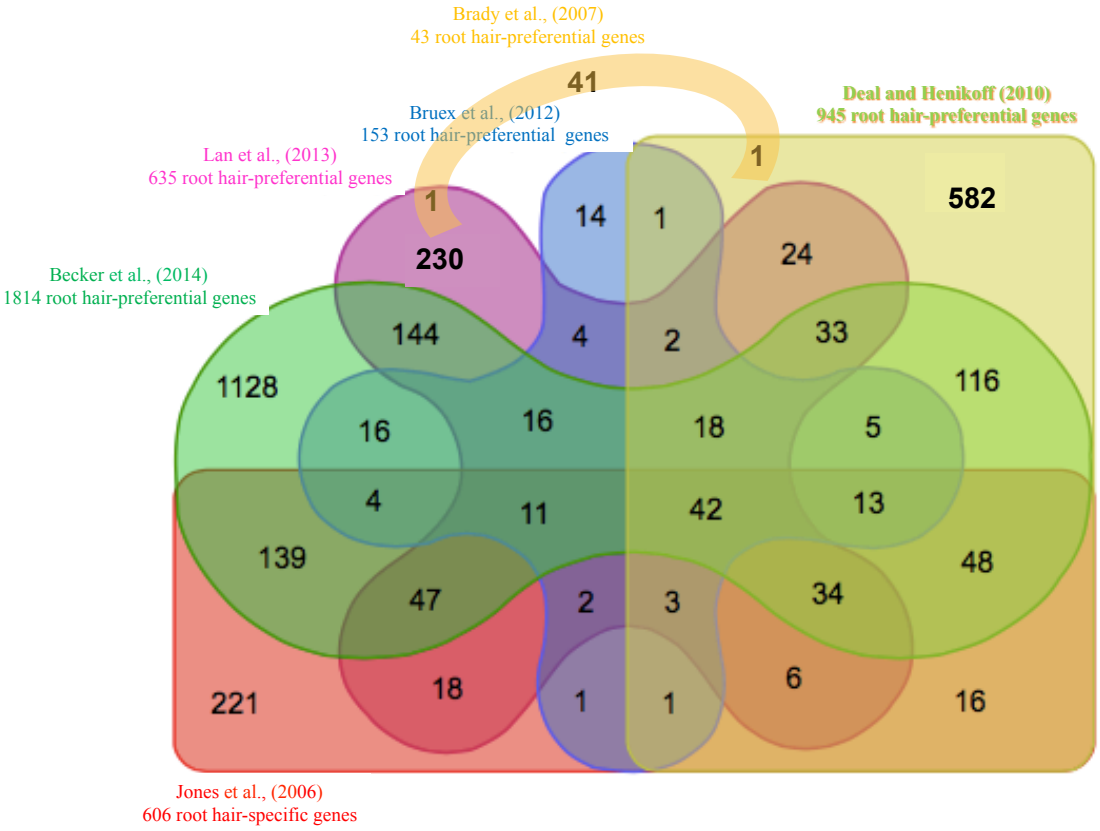


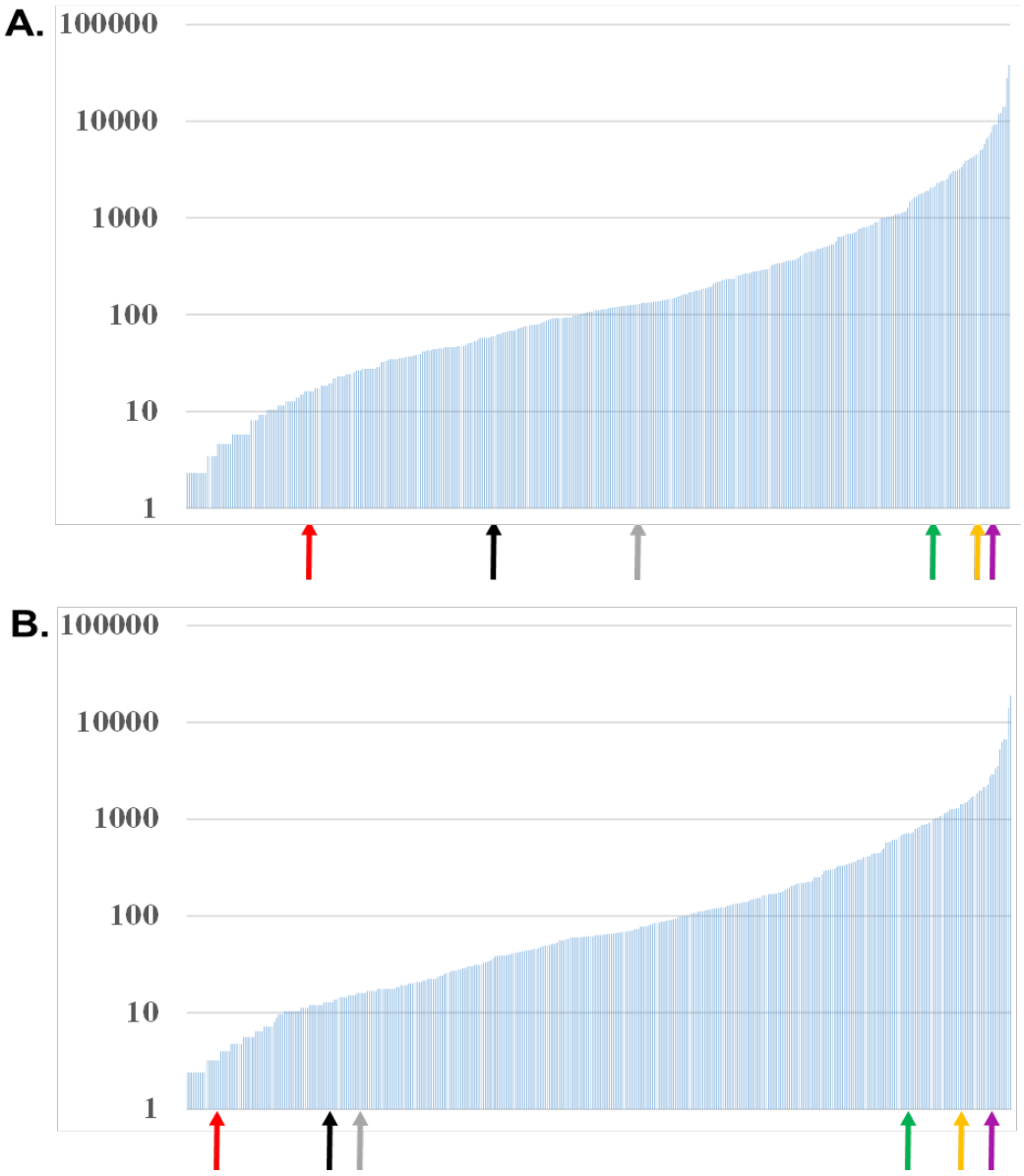
Figure 3- 6 Transcriptional activity of two RHEs cloned upstream to the minimal 35S in soybean (B, D, and F) and A. thaliana (C, E, and G) root hair cells. A RHE1 and two motif sequence logos and their e-values are provided. The 4 bp (i.e., A, C G, and T) are highlighted indifferent colors (i.e., red, blue, yellow, and green, respectively). The transcriptional activities of two regulatory elements, RHE1 (consensus sequence: TGTTGTTGCCGTAGGTGGTGGCGTTGGTC) (B and C) and RHE2 (consensus sequence: CGCCTCCGCCAACGAAGCATTCAGCAAC) (D and E), fused upstream to the minimal 35S promoter, were observed through the detection of activity of the β-glucuronidase reporter protein (black arrows). The minimal 35S cloned upstream to the UidA reporter gene was used as a negative control (F and G). Bar 0.2 mm



Supplemental Figure 3- 1 Synteny relationships existing between the 20 soybean chromosomes (x-axis) and the 5 *Arabidopsis* chromosomes (y-axis). Each green dots reflects the conservation of microsyntenic blocks between chromosomes of the two plant species. This syntenic map was generated using CoGe resources (Lyons and Freeling 2008; Lyons et al. 2008).



Supplemental Figure 3- 2 Venn diagram of the *Arabidopsis* genes preferentially expressed in root hairs and characterized from six independent studies (Becker et al. 2014; Brady et al. 2007; Bruex et al. 2012; Lan et al. 2013; Deal and Henikoff 2010; Jones et al. 2006).



Supplemental Figure 3- 3 The relative level of expression of the 399 soybean genes preferentially expressed in the root hair cells. Gene expression levels (RPKM; y-axis; logarithmic scale) are normalized Illumina reads number generated from root hair cells isolated 84 and 120 hours after soybean seed sowing (A and B, respectively). The 399 genes are displayed on the x-axis. The arrows highlighted the six soybean genes selected for functional analyses of their promoter sequences [i.e., Glyma.08G115000 (red), Glyma.15G020700 (Young et al.), Glyma.18G025200 (orange), Glyma.03G188300 (purple), Glyma.12G012500 (black) and Glyma.19G226300 (grey)]. These genes were selected to cover a large range of expression levels. The root hair-specificity of soybean genes highlighted with color with high chromaticity was experimentally confirmed.

Supplemental Figure 3- 4 Promoter sequences of 6 soybean genes preferentially expressed in root hairs and cloned upstream the *UidA* reporter gene.

Supplemental Table 3- 1 Annotation of the soybean root hair genes and loci according to the Wm82.a2.v1 version of the soybean genome sequence.

Supplemental Table 3- 2 Expression levels of soybean genes preferentially expressed in the root hair cells and their paralogous genes.

Supplemental Table 3- 3 Characterization of *Arabidopsis* genes preferentially expressed in root hairs upon mining and integration of six different root hair transcriptomic data sets

Supplemental Figure 3-4 and Supplemental Table 3-1, 3-2 and 3-3 are not included in this dissertation duo to their large sizes, which can be accessed by the link:

https://link.springer.com/article/10.1007%2Fs11103-017-0630-8#SupplementaryMaterial

Supplemental Table 3- 4 List of the 6 soybean genes preferentially expressed in root hair cells and selected for the analysis of the activity of their promoter sequence.

GeneID V1.1	Glyma. 12G012500	Glyma. 08G115000	Glyma. 19G226300	Glyma. 15G020700	Glyma. 18G025200	Glyma. 03G188300
Roor hair 84HAS	76.15	43.84	124.60	1795.22	5149.15	7644.69
Root hair 120HAS	13.59	15.18	17.58	573.03	1511.29	2758.04
Root Tip	0.00	0.00	0.00	49.51	92.96	153.58
Root	0.85	0.85	0.00	8.49	47.54	66.22
Nodule	0.00	0.00	0.00	0.00	0.00	0.00
SAM	0.00	0.00	0.00	2.50	0.00	0.00
Leaf	0.00	0.00	0.00	0.00	0.00	0.00
Flower	0.00	0.00	1.93	0.97	0.00	0.00
Green Pod	0.00	0.00	0.00	0.00	0.00	0.00

Supplemental Table 3- 5 Primers designed to quantify gene transcription ("qRT-PCR").

Gene ID	Forward primer	Reverse primer
Glyma01g42370	ttgtggtggtcactccttctc	cgaggaaaactcattttgcag
Glyma11g03000	cggccactcctttctaaaact	tagctacatgaacctgctgctc
Glyma17g14230	ccacaaatgaagcaaatggtc	gtgccaactgtgggagtacat
Glyma05g03720	ccaaggaaatgggtattggtt	attctgttccactccccttca
Glyma01g34770	gaaggcatatgcaagctgaaa	tggctagcaagaatgaaatcac
Glyma09g32630	tgtcccatttgtttgtgtttc	gcatggaatgaagatgaaactg
Glyma08g04020	aatcaatgtgcctccaatgtc	cccaatttagctgtccgatct
Glyma02g16800	tctgttaaagtggattccaagg	atagcacctccccaacaaaac
Glyma10g03000	gcctctcgatatcaattattctg	aagagatagcacctccccaac
Glyma19g33650	tcggtcacagaaaaatccatc	ccctttgtaggtgctgaggtt
Glyma03g30800	ggcattggtttgagttgtgtt	aagtgatgcaacttcctcgac
Glyma20g03120	gggtcaattgaagtaacggata	gcagccaaggagtgcatagta
Glyma07g35240	ggagcagtgaagttcatttgg	cgcaggttgaacttttctagg
Glyma01g07070	caatagggactgtggttggtg	ggtgcagatcttctagctttgg
Glyma02g12930	ctcctgacctgaagattgcac	cattggaaccttggggttatt
Glyma19g33180	ggtttgcttgctctccttttt	cctttgattctcataggccaaa
Glyma03g30260	tgatttggtttgcttgctctc	ccttgattctcataggccaac
Glyma09g16640	gccaatctaatctttctttcattg	cattatgcacaaaatacatgtcg
Glyma09g33650	cacatcaccattcgttgtctg	acaatcacatcagaatcgttcg
Glyma01g02330	cacaagtgttcccttctgcat	ctcttccccggtatgcattat
Glyma08g36820	ttggatgcggagtacaagaag	ggcttgtttgtctgaccatgt
Glyma08g41320	accatcaaatttggggagtgt	aggcttgtgatgaactcagga
Glyma08g41310	cccctcccccaatatttagtt	tgggaatgacagcgaatttac
Glyma14g03890	tctttctggcctttcaggatt	caactcagaaaaagtggaacaa
Glyma02g44890	tccatttcttcccttggatct	tccatcagagcccttcaagta
Glyma06g10460	aaacctgatgacccgtctttt	caatcaaattcggagaacgag
Glyma04g10610	tcacgtgacaagaaacagcac	ttccccgacatcatccttaat
Glyma14g35620	tgaggttggatgtggaaaaa	gcaatccaaaaattaaaatgaaa
Glyma02g37340	ggatacgaggtttgatgtgga	tccaaatccaaaaattaaactga
Glyma19g28770	tcatctcatgtgttgggtgaa	ggtaaatggattcttccccttt
Glyma16g04560	tgggtgaatcgtattttcacg	gctgtcacaagaagtggcatag
Glyma16g26940	tgccttagattggtccacttg	aacagcacaggaaaagggaac
Glyma02g07940	gaaattcctttccactttatgattg	attgcaattggaccaacaaag
Glyma05g22760	ggtagtcaaatgccgacaaaa	aaggcggttggacaagtttac
Glyma17g17210	gggtagtcaaatggcaacaaa	tcctagatcggtcacttgtgc
Glyma11g04970	cgtccggtggatatagctttt	ccggagctcagatctagatagc
Glyma01g40320	gtgcacagtttgtttcgtttg	aaaaacggtgtaactagcagca
Glyma02g35770	tttggggtcttctatggatca	aaagggatggtgtatggtcct
Glyma10g09480	tttggggtcttctatggatca	gggatgttgtatggtcctacg
Glyma19g38540	tgaatcctctttgtggacgat	atgcccgtatatgatggatga

Glyma03g35880	atcctctttgtggacgattca	atgcccgtatatgatggatga
Glyma09g36690	gggcaatgctctaatagatgg	cacatcacgttgcatctaaaat
Glyma12g00670	caatacttttgccatggagga	aagcttccagcggaagataac
Glyma16g05710	taggcctgcataatgaattgc	cacaagatcaacctccactca
Glyma19g26830	cacccctaacacgctacata	tcgatcttggagtggtctttg
Glyma02g04010	tgcccaaacactttctcctaa	tggtgctttgtaagttgatgc
Glyma01g03690	gccaaaactgatttcaccaaac	cgagtcatggttgcaatttct
Glyma08g39480	ggagttagaggctgctgcttt	tacaggatgggatgtggtttg
Glyma07g05390	tggtttttgaatacattttgatga	tgaagcattacagtttatggattt
Glyma19g45050	tctccatcagaataagccacaa	ccgactaaatacaaaagggatg

Supplemental Table 3- 6 Primers designed to clone selected soybean root hair promoter sequences ("Promoter cloning")

Primer	Sequence
Glyma.03G188300-forward	acagtttgatgatatttaaccaaaac
Glyma.03G188300-reverse	catctctctatctatctttctctc
Glyma.03G188300-forscreen	cgtacccgttgttccgcact
Glyma.08G115000-forward	gttggtcagttataggaagaaatg
Glyma.08G115000-reverse	cattttggttttgatactccatac
Glyma.08G115000-forscreen	gattgggctaatctataagtctc
Glyma.12G012500-forward	acaaatttagtcatatttacatgaatt
Glyma.12G012500-reverse	agcctagcttcacaaactaagg
Glyma.12G012500-forscreen	gcactgatagtttgacaattagat
Glyma.15G020700-forward	atgtactttagtagttactgcttta
Glyma.15G020700-reverse	catcttgaactcaactcaaccta
Glyma.15G020700-forscreen	cctatggggatggtaataatcaa
Glyma.18G025200-forward	gactgggtgaatatgacttggt
Glyma.18G025200-reverse	catttttccttgaggggtatact
Glyma.18G025200-forscreen	gaccaaagtggtggcaaaaag
Glyma.19G226300-forward	ttttaattttaatttgttaccaaatatt
Glyma.19G226300-reverse	accatgaagtttttggtggttaa
Glyma.19G226300-forscreen	gacccgtaaggccataaccc

Chapter 4: Comprehensive comparative genomic and transcriptomic analyses of the legume genes controlling the nodulation process

Authors: **Zhenzhen Qiao**, Lise Pingault, Mehrnoush Nourbakhsh-Rey and Marc Libault

Publication status: This chapter was published in Frontiers in Plant Science (DOI: 10.3389/fpls.2016.00034), which allowed incorporation into this dissertation.

Author contributions:

Dr. Marc Libault designed the research. Mehrnoush Nourbakhsh-Rey characterized and updated the annotation of the legume nodulation genes with my assistance. I performed the syntenic and transcriptomic analyses between legume genes. Dr. Lise Pingault analyzed the distribution of the nodulation genes on the legume chromosomes. Dr. Marc Libault wrote the article.

Abstract

Nitrogen is one of the most essential plant nutrients and one of the major factors limiting crop productivity. Having the goal to perform a more sustainable agriculture, there is a need to maximize biological nitrogen fixation, a feature of legumes. To enhance our understanding of the molecular mechanisms controlling the interaction between legumes and rhizobia, the symbiotic partner fixing and assimilating the atmospheric nitrogen for the plant, researchers took advantage of genetic and genomic resources developed across different legume models (e.g. *Medicago truncatula*, *Lotus japonicus*, *Glycine max* and *Phaseolus vulgaris*) to identify key regulatory protein coding genes of the nodulation process. In this study, we are presenting the results of a comprehensive comparative genomic analysis to highlight orthologous and paralogous relationships between the legume genes controlling nodulation. Mining large transcriptomic datasets, we also identified several orthologous and paralogous genes characterized by the induction of their expression during nodulation across legume plant species. This comprehensive study prompts new insights into the evolution of the nodulation process in legume plant and will benefit the scientific community interested in the transfer of functional genomic information between species.

Introduction

Legumes (family Fabaceae) are the 2nd most important crop family after grass species. Legumes are an important source of oil and proteins for human and animal consumption, and they also fix atmospheric nitrogen leading to a sturdy supply of nitrogen fertilizers, which can benefit other plant species. This unique feature of legumes is the result of their symbiotic relationship with soil bacteria involved in nodulation (e.g., *Bradyrhizobium japonicum* and *Sinorhizobium meliloti* are the *Glycine max* and *Medicago truncatula* symbiotic partners, respectively). Ultimately, upon infection of the legume by the symbiotic bacteria, a novel plant root lateral organ called a nodule is formed where rhizobia fix and convert atmospheric nitrogen into nitrite/ammonia that can be used by plants. In exchange, acting as a carbon sink, the symbionts receive plant photosynthates.

Legume nodulation is initiated by the recognition of plant flavonoids by rhizobia. In response, rhizobia secretes the Nod Factor (NF), a lipochito-oligosaccharide signal molecule which is recognized by plant lysine motif (LysM) receptor kinase, such as *Lotus japonicus* NFR1/NFR5 (Madsen et al., 2003a; Radutoiu et al., 2003), *Pisum sativum* SYM10 (Madsen et al., 2003a), and *M. truncatula* NFP (Amor et al., 2003) and LYK3 (Limpens et al., 2003). Downstream of the recognition of the NF, a signaling cascade is activated leading to oscillations in calcium concentration within the nucleus of the root hair cell (Capoen et al., 2009; Miwa et al., 2006; Sieberer et al., 2009a). Ultimately, this molecular recognition of the two partners will lead to root hair cell deformation, curling, and formation of a shepherd hook (Kijne, 1992). Upon root hair curling. rhizobia infect the root hair cell through the formation of a tube-like apoplastic

compartment called infection thread (VandenBosch et al., 1989). Concomitantly, the root cortex cells are actively dividing leading to the formation of the nodule primordium (Yang et al., 1994).

Through an examination of legume mutants defective in nodulation, root hair deformation, and calcium spiking, the root hair regulatory pathway activated in response to rhizobia inoculation or NF treatment was characterized. For instance, while the *M. truncatula* mutants *dmi1* and *dmi2 (does not make infections)* are not affected by their calcium flux and root hair deformation, the *nfp* (*nod factor perception*) mutant is impaired for both phenotypes suggesting that the *DMI1* and *DMI2* genes are acting downstream to *NFP* (Amor et al., 2003). As a part of this regulatory pathway, CCaMK protein, a nuclear protein sensitive to the calcium oscillations, interacts with and phosphorylates CYCLOPS, a nuclear coiled-coil transcription factor, directly inducing the expression of *NODULE INCEPTION (Sieberer et al.),* encoding a RWP-RK transcription factor (Marsh et al., 2007; Singh et al., 2014). In another model legume, *L. japonicus*, LjNIN targets two *Nuclear Factor-Y* (*NF-Y*) genes, Lj*NF-YA1* and Lj*NF-YB1*, to control nodule development (Soyano et al., 2013*)*. NF-Y TFs play a central role during the nodulation process. For instance, in *M. truncatula*, MtNF-YA1 and MtNF-YA2 redundantly act to control the early stage of rhizobial infection via the transcriptional activation of Mt*ERN1* (*Ethylene Response Factor Required for Nodulation 1*) (Laloum et al., 2014). In *Phaseolus vulgaris*, PvNF-YC1 is also controlling nodule organogenesis as well as the selection process of the symbiosis partner (Zanetti et al., 2010). More recently, Baudin et al. (2015) demonstrated that both MtNF-YA1, MtNF-YB16 and MtNF-YC2, and PvNF-YA1, PvNF-YB7, and PvNF-

YC1 proteins form heterotrimers recognizing the CCAAT box of the Mt*ERN1* and Pv*ERN1* promoter sequences, respectively (Baudin et al., 2015). Interestingly, while Mt*ERN2*, the Mt*ERN1* paralogous gene, also regulates the nodulation process, its activity does not overlap with Mt*ERN1* supporting the existence of additional regulatory pathways (Cerri et al., 2012). Together, these legume TFs control the expression of various genes regulating the infection of the plant root hair cell by rhizobia then its progression in the root cells through the development of the infection thread. MtFLOT4 and LjSYMREM1, plasma membrane microdomain proteins, seem also to play integral roles during the early stages of infection of the root hair cells by rhizobia such as the formation of the infection thread (Haney and Long, 2010a; Toth et al., 2012).

While the genetic resources developed on model legumes, such as *M. truncatula* have a considerable impact on the characterization of nodulation genes, additional resources have been more recently developed taking advantage of –omic technologies. For instance, the access to large quantities of *M. truncatula* (Breakspear et al., 2014) and soybean root hair cells inoculated by rhizobia (Brechenmacher et al., 2010b; Brechenmacher et al., 2012a; Libault et al., 2010b; Libault et al., 2010c; Nguyen et al., 2012a; Yan et al., 2015) are now allowing a more global understanding of the molecular processes controlling the early stages of legume nodulation. In this manuscript, we are taking advantage of the knowledge gained during the past two decades and the more recent release of genomic and transcriptomic datasets to perform a comprehensive analysis of the evolution of legume protein coding genes controlling the nodulation process. Our results confirm the strong conservation of a core set of legume protein

coding genes controlling nodulation between species and also highlight the divergence of a subset of the paralogs. Interestingly, the nodulation genes are also characterized by their high density along the legume chromosomes suggesting that these genes share the same biological function and are physically co-localized on the chromosomes. This study represents a new resource to better understand the evolution of legume nodulation, maximize the transfer of the scientific information between legumes and to open perspectives regarding the role and the conservation of gene modules in controlling the nodulation process.

Results

Syntenic relationships reveal the orthologous relationships between legume nodulation-related protein-coding genes

To perform the most comprehensive evolutionary analysis of the protein coding genes involved in nodulation, we mined the scientific literature allowing us to list 110 functionally characterized genes from *M. truncatula*, *L. japonicus*, *G. max* and *P. vulgaris* (Supplemental Table 4-1). To date, most protein coding genes involved in nodulation genes have been characterized in the model legumes *G. max* and *M. truncatula* notably upon the development of the *Tnt1* retrotransposon insertion mutant population (Cui et al., 2013; Pislariu et al., 2012; Tadege et al., 2008b). This observation supports the need to identify orthologous genes between legume species to facilitate the transfer of the scientific knowledge.

To update the annotation of these 110 genes, we first blasted their published nucleotidic sequences against the most recent release of the legume genome sequences

[i.e., Phytozome v10.3 (http://phytozome.jgi.doe.gov/pz/portal.html) for *M. truncatula* (Mt4.0v1), *G. max* (Wm82.a2.v1) and *P. vulgaris* (v1.0); 2- Miyakogusa v3.0 (http://www.kazusa.or.jp/lotus/) for *L. japonicus* (v2.5)] (e-value<10^{-12} and score>100) (Supplemental Table 4-1).

To identify orthologous and paralogous genes in and between *L. japonicus, M. truncatula, G. max* and *P. vulgaris*, we examined syntenic relationships between corresponding genomic regions based on gene content, order, and orientation. To perform this analysis, we took advantage of the release of the sequence of various legume genomes and the development of comparative genomic resources such as CoGe (Lyons and Freeling, 2008; Lyons et al., 2008b). Upon our evolutionary analysis, we repetitively observed strong syntenic relationships between genes from the 4 legume species including a large number of soybean paralogs, a consequence of the most recent whole genome duplication (WGD) of the soybean genome (Schmutz et al., 2010b) (Figure 4-1; Supplemental Table 4-1 and Supplemental Figure 4-1 for access to the entire datasets). Together, this comparative genomic analysis led to the characterization of 191, 92, 65 and 91 soybean, *M. truncatula*, *L. japonicus* and common bean genes orthologous and paralogous to functionally described nodulation genes, respectively. These genes belong to 81 orthologous/paralogous groups (Supplemental Table 4-1).

Comparative transcriptomic analysis reveals conservation and neo-/sub-functionalization between M. truncatula, G. max, L. japonicus and P. vulgaris nodulation genes

While microsynteny relationships clearly revealed the orthology existing between nodulation genes from different species, they are not sufficient to conclude about the conservation of their function. To provide a first insight into the conservation and divergence of the function between orthologous genes, we mined transcriptomic databases and integrated them into our comparative genomic analysis. Specifically, we took advantage of the release of the *M. truncatula* and soybean root hair transcriptomes and their perturbation in response to rhizobia inoculation (Breakspear et al., 2014; Libault et al., 2010c) as well as the access to the *M. truncatula*, *G. max*, *L. japonicus* and *P. vulgaris* transcriptome atlases (Benedito et al., 2008; Libault et al., 2010g; O'Rourke et al., 2014; Verdier et al., 2013). In addition, we included in our analysis more focused transcriptomic studies on the nodulation process in the model plant *M. truncatula* (Larrainzar et al., 2015; Roux et al., 2014)

To identify the entire set of legume genes transcriptionally induced during nodulation, we independently analyzed their expression pattern during both the early (i.e., root hair response to rhizobia inoculation) and late events of the nodulation process (i.e., nodule specific expression compared to other plant organs). A total of 18 and 19 *M. truncatula* genes were induced in root hair cells in response to wild-type *Sinorhizobium meliloti* inoculation or were preferentially expressed in nodules compared to other plant tissues (fold-change>2; Supplemental Table 4-2), respectively. Among those genes, eleven were both induced in root hair cells in response to *S. meliloti* and preferentially

expressed in nodules including Mt*NF-YA1/2* (Figures 4-2B and 3B), Mt*NIN* (Figures 4-2D and 3F), Mt*FLOT2/4* and Mt*NSP1* (Figures 4-2F and 3J). To provide a more complete understanding of the *M. truncatula* genes transcriptionally induced during nodulation, we mined recently published *M. truncatula* RNA-seq data sets (Larrainzar et al., 2015; Roux et al., 2014). We identified a total of 30 and 15 *M. truncatula* genes differentially expressed in specific zones of the *M. truncatula* nodule [i.e., (Roux et al., 2014)] and during the early stages of nodulation in root tissue [i.e., from 30 minutes to 2 days after *S. meliloti* inoculation; (Larrainzar et al., 2015)], respectively (Supplemental Table 4-2). Analyzing the soybean transcriptome, the expression of 47 genes was induced in root hair cells in response to rhizobia inoculation while 38 soybean genes were preferentially expressed in nodules versus other soybean tissues. Among those, 20 were both up-regulated in root hairs in response to *B. japonicum* and preferentially expressed in nodules such as Gm*NF-YA1/2* (Figures 4-2A and 3A), Gm*NIN* (Figures 4-2C and 3E), and Gm*FLOT2/4* (Figures 4-2E and 3I). Mining the common bean and *L. japonicus* transcriptome atlases, we also characterized 19 and 16 genes preferentially expressed in common bean and *L. japonicus* nodules, respectively such as Pv*NF-YA1/2* (Figure 4-3C), Pv*NIN* (Figure 4-3G), Lj*NF-YA1/2* (Figure 4-3D), Lj*NIN* (Figure 4-3H). Unexpectedly, Pv*NSP1*, PvFLOT2*/4* and Lj*NSP1* are not preferentially expressed in common bean and *L. japonicus* nodules (Figure 4-3D, H, K and L).

To better evaluate the impact of soybean WGDs on the population of genes controlling nodulation, we classified and compared the 47 and 18 soybean and *M. truncatula* genes induced in root hair cells in response to rhizobia inoculation based on their paralogous relationships. A total of 32 and 16 paralogous groups were identified

for each plant species, respectively (Figure 4-4A; Supplemental Table 4-1, Supplemental Table 4-3). Performing a similar analysis on the 38, 19, 19 and 16 soybean, *M. truncatula,* common bean and *L. japonicus* genes preferentially expressed in nodules, we characterized 27, 18 18 and 14 groups of paralogous genes, respectively (Figure 4-4B, Supplemental Table 4-3). These results suggest that a large number of differentially expressed soybean genes during nodulation is not just the consequence of the WGDs.

Upon WGDs, paralogs can share the same expression profiles leading to functional redundancy [e.g., induction of the expression of the 4 soybean *NF-YA1/2* paralogs in root hair cells in response to *B. japonicum* inoculation (Figure 4-2A)], or could be a source for sub- and neo-functionalization [(e.g., Gm*NIN*s orthologs can be divided into two groups: *Glyma.14G001600* and *Glyma.02G311000*, which are induced during nodulation and *Glyma.04G017400* and *Glyma.06G017800*, characterized by their more constitutive expression (Figures 4-2C and 3E)]. To further explore the consequences of WGDs on the transcriptional regulation of paralogous genes, we compared the expression of soybean and *M. truncatula* genes that belong to paralogous groups that are induced during nodulation.

We identified 11 overlapping root hair inducible groups between the two plant species (34% and 69% of the *G. max* and *M. truncatula* root hair-inducible groups, respectively; Figure 4-4C, Supplemental Table 4-3). Twenty-nine and 18 soybean and *M. truncatula* genes are represented in these 11 groups, respectively, including the *NOD100*, *NMNa*, *RIP1*, *MtPUB1*, *NODULIN-26a*, NIN, *FLOT2/4*, *NSP1*, *MtNramp1*, *NF-YA1/2* and *ERN1/2* genes (Supplemental Table 4-3). Performing a similar analysis

on the 27, 18, 18 and 14 *G. max*, *M. truncatula*, *P. vulgaris* and *L. japonicus* groups preferentially expressed in nodules, we characterized 4 overlapping orthologous groups (11%, 22%, 22% and 29% of the *G. max, M. truncatula, P. vulgaris* and *L. japonicus* groups preferentially expressed in nodules, respectively; Figure 4-4D). These 4 groups are composed of 14, 7, 7 and 8 *G. max, M. truncatula, P. vulgaris,* and *L. japonicus* genes, respectively, such as *ENOD20, NIN, NIN2,* and *NF-YA1/2* genes (Supplemental Table 4-3).

Analyzing the expression patterns of these genes, we only observed a slight increase in the number of soybean genes differentially expressed during the early (Figure 4-4E; grey bar) and late events of the nodulation process compared to *M. truncatula*, *P. vulgaris* and *L. japonicus* genes (Figure 4-4F, grey bar). Oppositely, many soybean paralogs are not transcriptionally responsive to rhizobium, but are characterized by their broad expression patterns (Figure 4-4E and F, orange bars). These results suggest that upon soybean WGDs, a subset of the paralogous genes were transcriptionally restricted to their role during nodulation based on the conservation of their induction in response to rhizobia while other paralogs gained new function after alteration of their expression patterns.

Legume genes specialized in the nodulation processes are forming gene modules on legume chromosomes

Previous studies revealed the correlation existing between gene function and the organization of the euchromatin. Specifically, gene modules on chromosomes are characterized by the presence of genes involved in the same biological function and

sharing similar expression patterns (Pingault et al., 2015; Williams and Bowles, 2004; Zhan et al., 2006). Hypothesizing that nodulation genes are also located in modules along the legume chromosomes, we analyzed their chromosomal distribution and density. Accordingly, we mapped the annotated legume genes controlling nodulation (Supplemental Table 4-1) on the 20, 8, 6 and 11 chromosomes of *G. max*, *M. truncatula*, *L. japonicus* and *P. vulgaris*, respectively. To take in consideration the differences in the size of the genomes of the four legume plants, we analyzed the distribution of nodulation genes on the chromosomes generating window sizes of 10 Mb, 6 Mb, 5 Mb and steps of 1 Mb, 0.6 Mb and 0.5 Mb for *G. max*, *P. vulgaris* and *L. japonicus*/*M. truncatula* respectively. To highlight the significant increase of the density of nodulation genes on chromosomes as reflected by their Z-score, we also normalized these results based on a random distribution of the nodulation genes (Figure 4-5). A significant enrichment in nodulation genes was repetitively observed on the *M. truncatula*, *L. japonicus* and *P. vulgaris* chromosomes and on 12 out of the 20 soybean chromosomes (Kolmogorov-Smirnov test; p-value<0.05; Supplemental Table 4-4). Together, these results confirmed the clustering of nodulation genes in functional modules on the chromosomes. The fact that these modules are conserved between model legumes suggests their essential roles in controlling the nodulation process.

Discussion

Applying comparative genomic and transcriptomic analyses to reveal the conservation and divergence between nodulation orthologous/paralogous genes

Functional genomic studies led to the characterization of 110 genes controlling the early and late events of the nodulation process. As described by Chen et al. (2015), a core set of *M. truncatula* genes such as *ERN1/2*, *NSP2* and several *NF-Y* genes are playing a central role in plant cell infection by rhizobia, both early (i.e., root hair cell infection) and late (i.e., nodule cell infection) during the nodulation process (Chen et al., 2015). In this study, our analyses have been conducted in four different legume models having the objective to better transfer scientific knowledge between species. Accordingly, we took advantage of the genomic information now available to perform a global and comprehensive analysis of the evolution of nodulation genes based on: 1- their orthology and paralogy upon demonstration of their microsyntenic relationships and; 2- the conservation and divergence of their transcriptomic patterns.

Applying various bioinformatics tools available on the CoGe platform (https://genomevolution.org/coge/) such as Synfind and GEvo, we repetitively analyzed the evolutionary relationship between nodulation genes across four model legume plants. Upon the availability of functional genomic datasets, we were able to confirm the orthologous relationships existing between genes known to control nodulation across different species. For instance, using the soybean NF receptor gene Gm*NFR1a* (Indrasumunar et al., 2011) as a query, we clearly highlighted its orthology with the *M. truncatula* Mt*LYK3/HCL* (Limpens et al., 2003) and the *L. japonicus* Lj*NFR1* genes (Radutoiu et al., 2003). To further support our analysis, we revealed the orthology

existing between 65 functionally characterized genes across different species [e.g., *SYMRK /NORK*, *SUNN/NARK/HAR1*, *RDN2*, *POLLUX/DMI1*, *NSP2*, *NSP1*, *NIN*, *NFR/NFP*, *NFR1/LYK3/HCL*, *MtENOD20/ENOD55-2*. *ENOD16/ENOD55-1* and *CCaMK/DMI3* genes (Supplemental Figure 4-1)].

While the conservation of the function between orthologous genes is often assumed, biological evidence to validate these assumptions are limited. Transcriptomes provide a first insight into the conservation of the function of orthologous and paralogous genes based on the conservation of a similar expression pattern. For instance, previous studies focusing on the fate of soybean paralogs, products of the successive duplications of the soybean genome, concluded about the differential expression of 50% of the paralogs leading to their sub-functionalization (Roulin et al., 2013). Taking advantage of the release of legume transcriptome atlases as well as the analysis of the transcriptomic response of the root hair cells to rhizobia inoculation, we compared the nodulation related protein coding genes' expression patterns between soybean, *M. truncatula*, common bean and *L. japonicus* orthologous genes. These analyses were also conducted under the context of paralogy. Our study clearly revealed the conservation of the transcriptomic patterns between *M. truncatula*, *G. max*, *L. japonicus* and *P. vulgaris* orthologous genes during the nodulation process supporting the conservation of the biological function of at least one paralog. Interestingly, a significant number of soybean paralogs can display very distinctive expression patterns suggesting a gain of function or a sub-functionalization (e.g., soybean *NIN* genes; Figure 4-2 and 3; Figure 4-4). Differential epigenome and the evolution of the promoter regions of paralogous genes (i.e., presence and absence of *cis*- and *trans*-regulatory

elements) should be carefully investigated to reveal the evolutionary mechanisms controlling these transcriptomic changes.

Plant genes acting together in the same biological process are physically closely located on the chromosomes (Pingault et al., 2015; Williams and Bowles, 2004; Zhan et al., 2006). Taking advantage of the identification of a large number of nodulation genes from 4 different species (i.e., 191, 92, 65 and 91 *G. max*, *M. truncatula*, *L. japonicus*, *P. vulgaris* genes, respectively), we analyzed their distribution along the 20, 8, 6 and 11 chromosomes of each species, respectively. We observed that the legumes genes controlling the nodulation process are characterized by their high density on the chromosomes rather than being randomly located on the chromosomes. This close relationship might be beneficial to their co-expression during nodulation. The relationships existing between chromosome territories, position of gene modules on the chromosome, their epigenomic context and their transcriptional activities should be investigated.

Material and Methods

Identification of the nodulation genes across legume species

To properly update the annotation of the functionally characterized nodulation genes, published nucleotidic sequences were used as a query for a BLAST search against the four different genome sequences [i.e., *M. truncatula* v4.0 (Young et al., 2011a), *G. max* Wm82.a2.v1 (Schmutz et al., 2010b), *P. vulgaris* v1.0 (Schmutz et al., 2014) and *L. japonicus* v2.5 (Sato et al., 2008b) available on the Phytozome v10.3 (http://phytozome.jgi.doe.gov/pz/portal.html) and Miyakogusa v3.0

(http://www.kazusa.or.jp/lotus/) websites]. Hits with an e-value<10^{-12} and a score>100 were considered for further analysis.

Syntenic analysis between nodulation genes

The Accelerating Comparative Genomic database [CoGe (https://genomevolution.org/coge/) (Lyons and Freeling, 2008; Lyons et al., 2008b)] was mined to characterize microsynteny relationships between legume genes. The most recent versions of the genome sequences available on the CoGe database were selected when highlighting microsynteny relationships [i.e., *M. truncatula* v4.0*, G. max* v9.0, *P. vulgaris* v1.0 and *L. japonicus* v2.5]. Only syntelog genes were used for further analysis. To better connect the v9.0 annotations of the soybean nodulation genes characterized by CoGe with the most recent release of the soybean genome (Wm82.a2.v1), we included both annotation systems in our analysis (Supplemental Table 4-1). The GEvo (genome evolution analysis) tool was applied to visualize the collinearity and/or rearrangement between syntenic regions.

Transcriptomic analysis of soybean and M. truncatula nodulation genes

Gene expression atlases were mined to characterize the transcriptional patterns of the G. max, *M. truncatula*, *L. japonicus* and *P. vulgaris* nodulation genes including in root hairs in response to rhizobia inoculation (Benedito et al., 2008; Breakspear et al., 2014; Libault et al., 2010c; Libault et al., 2010g; O'Rourke et al., 2014; Verdier et al., 2013). The expression pattern of orthologous genes was compared based on the

induction of gene expression in root hair cells upon rhizobia inoculation and based on their specific expression in nodules compared to other plant tissues.

Gene density analysis

For each species, the annotation file (gff3) were collected from the Phytozome v10.3 (http://phytozome.jgi.doe.gov/pz/portal.html#) and the Miyakogusa v3.0 websites (http://www.kazusa.or.jp/lotus/):

M. truncatula: Mtruncatula_285_Mt4.0v1.gene.gff3.gz

G. max: Gmax_275_Wm82.a2.v1.gene.gff3.gz

P. vulgaris: Pvulgaris_218_v1.0.gene.gff3.gz(Pislariu et al., 2012)

L. japonicus: Lj2.5_gene_models.gff3

The distribution of nodulation genes on the chromosomes was performed with sliding window sizes of 10 Mb, 6 Mb, 5 Mb and a step of 1 Mb, 0.6 Mb and 0.5 Mb for *G. max*, *P. vulgaris,* and *L. japonicus*/*M. truncatula* respectively. The gene density was normalized with a Z-score calculation [Z-score = (gene density-u)/st, u and st are the mean and the standard deviation of the gene density for each chromosome, respectively], and R package ggplot2 (http://ggplot2.org/) was used to draw the plot. The sample function in R was used to produce random nodulation gene distribution. The Kolmogorov-Smirnov test was used to validate the specific distributions of the genes on the chromosomes.

Figures and Tables

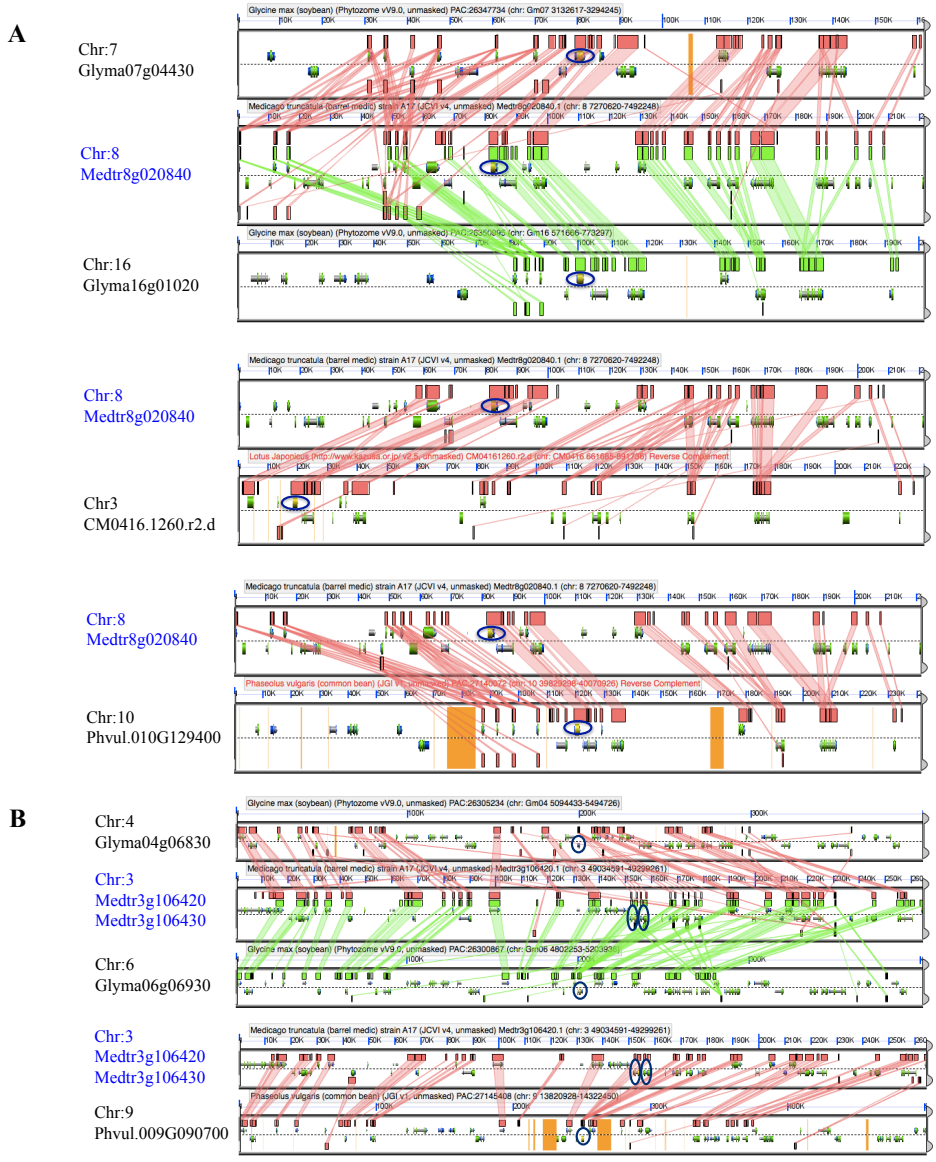


Figure 4- 1 Syntenic relationships between Mt*NSP1* (A) and Mt*FLOT2/4* (B) loci and chromosome regions from *Glycine max*, *Medicago truncatula*, *Lotus japonicus* and *Phaseolus vulgaris*. Each panel is a visualization of chromosome region showing the gene models on positive and negative strands. Green and red blocks in each panel highlight microsyntenic regions between legumes based on gene function and orientation. Genes highlighted in blue (e.g., *Medtr8g020840* and *Medtr3g106420/Medtr3g106430* for *NSP1* and *FLOT2/4* genes, respectively) were used as a query when performing microsynteny analysis against the four legume species. The *NSP1* and *FLOT2/4* orthologs based on their microsyntenic relationships are highlighted with blue circles.

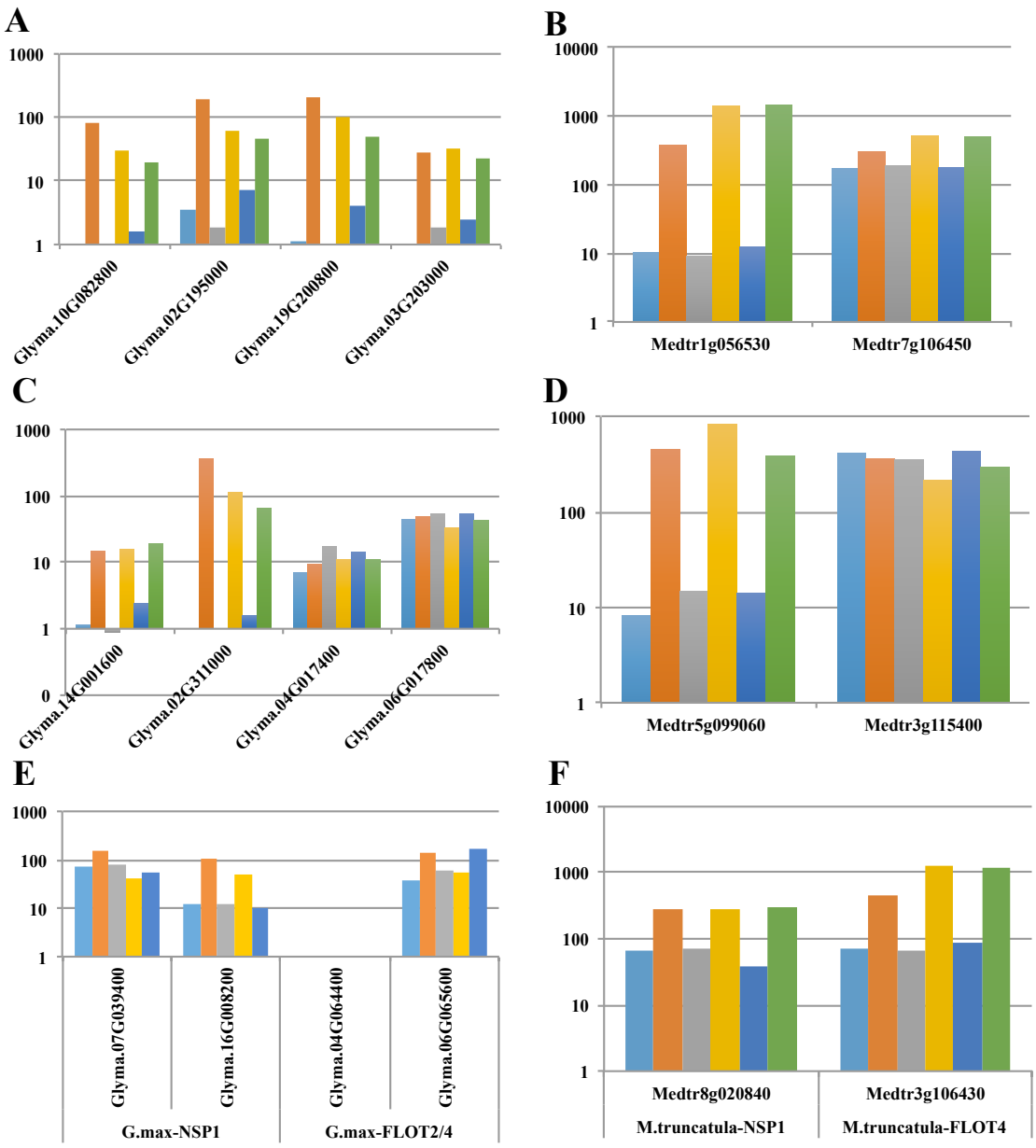


Figure 4- 2 Relative expression levels of *NF-YA1/2* (A, B), *NIN* (C, D), *NSP1* and *FLOT2/4* (E, F) *M. truncatula* genes (B, D, F) and their *G. max* orthologs (A, C, E) in inoculated and mock-inoculated root hair cells. Gene IDs are highlighted on the x-axis. The relative expression levels of the genes (log10 scale) are indicated on the y-axis. Gene expression datasets were mined from the *G. max* and *M. truncatula* root hair transcriptomic datasets (Libault et al., 2010b; Breakspear et al., 2014).
For *G. max*: blue: 12H UN (Uninoculated), orange: 12H IN (Inoculated), grey: 24H UN, yellow: 24H IN, dark blue: 48H UN and green: 48H IN;
For *M. truncatula*: blue: 1D UN, orange: 1D IN, grey: 2D UN, yellow: 2D IN, dark blue: 3D UN and green: 3D IN).

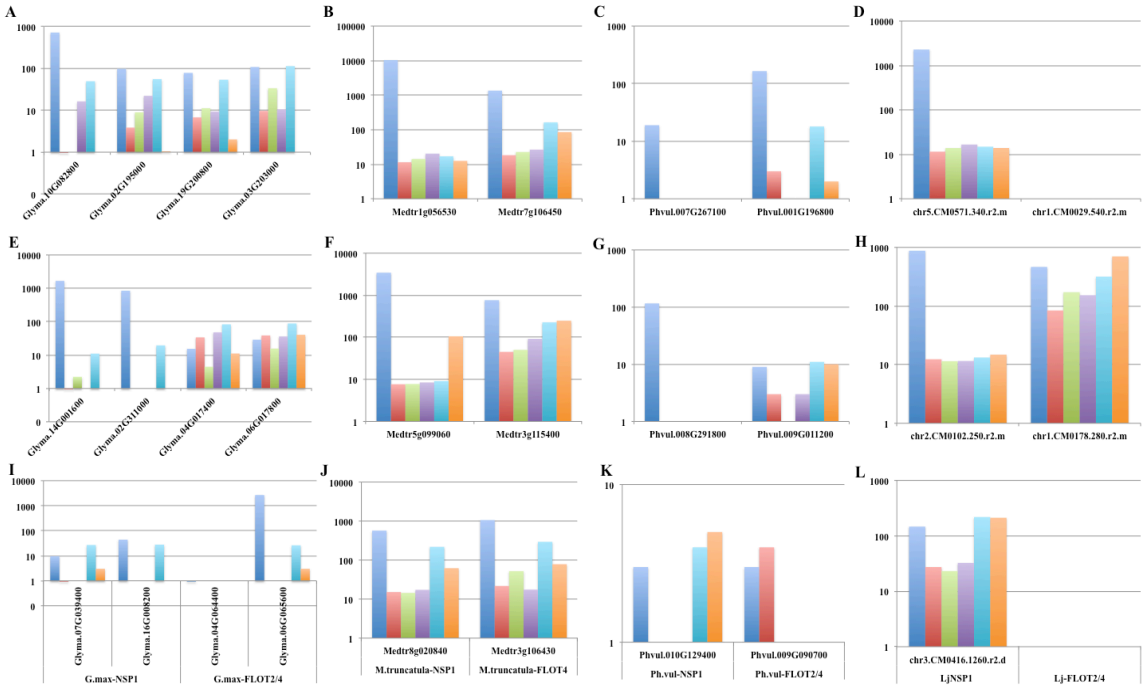


Figure 4- 3 Relative expression levels of *NF-YA1/2* (A, B, C, D), *NIN* (E, F, G, H), *NSP1* and *FLOT2/4* (I, J, K, L) of *M. truncatula* genes (B, F, J) and their *G. max* orthologs (A, E, I), *P. vulgaris* orthologs (C, G, K) and *L. japonicus* (D, H, L) in various plant organs (i.e., blue: nodule, orange: flower, grey: pod, yellow: leaf, dark blue: root and green: root tip). Gene IDs are highlighted on the x-axis. The relative expression levels of the genes (log10 scale) are indicated on the y-axis. Gene expression datasets were mined from the *G. max, M. truncatula, P. vulgaris* and *L. japonicus* atlases datasets (Libault et al., 2010c; Benedito et al., 2008 (http://mtgea.noble.org/v3/); Verdier et al, 2013; O'Rourke et al., 2014).

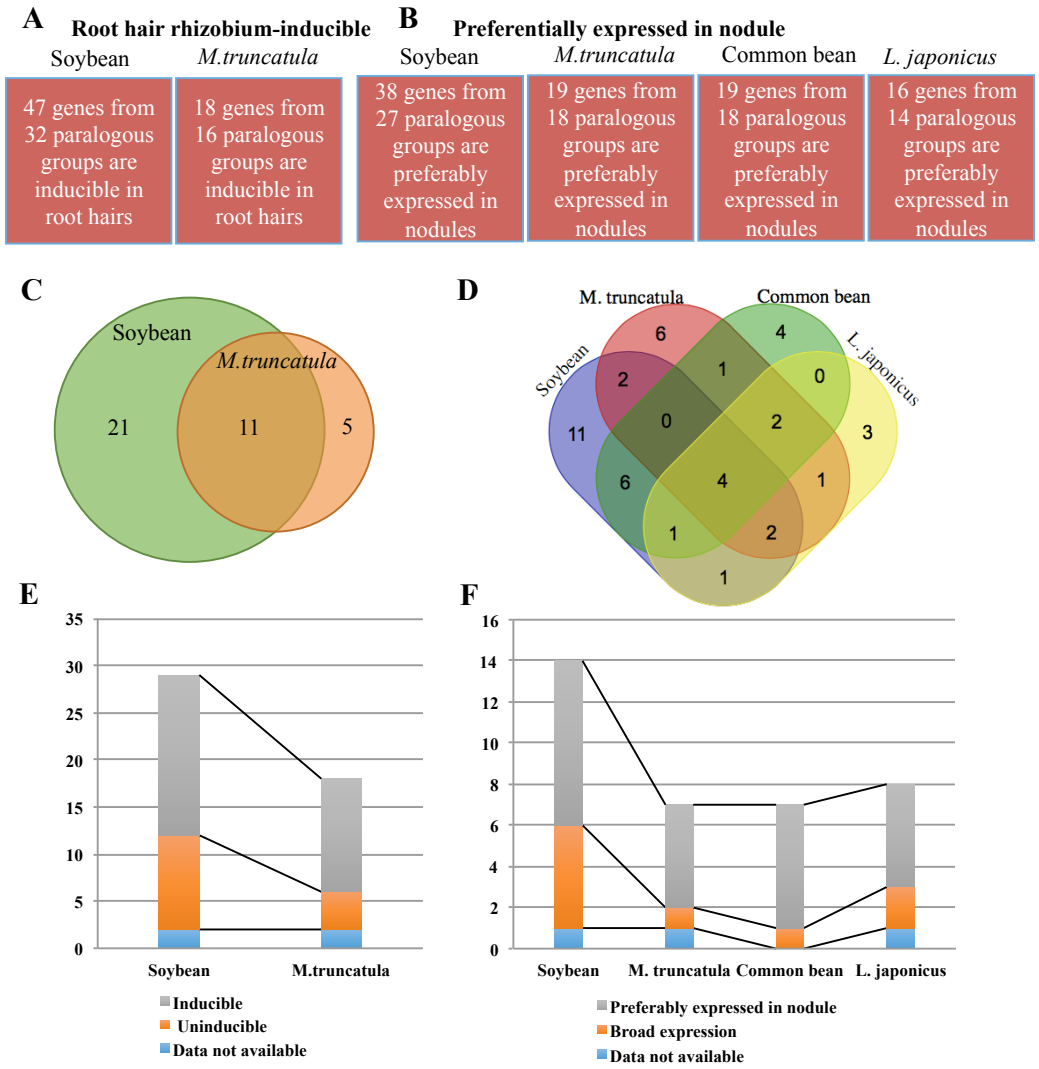


Figure 4- 4 Comparison of the expression profiles between *G. max*, *M. truncatula*, *P. vulgaris* and *L. japonicus* orthologous genes and groups during the early (i.e., root hair response to rhizobia inoculation; A, C, E) and late stages of the nodulation process (i.e., preferential expression in mature nodules; B, D, F).
A and B: numbers of *G. max* and *M. truncatula* genes and orthologous groups induced in root hairs upon rhizobia inoculation (A) and *G. max*, *M. truncatula*, *P. vulgaris* and *L. japonicus* genes preferentially expressed in nodules (B) (fold-change>2; Supplemental Table 4-3 for details).
C and D: Venn diagrams showing the overlaps between *G. max* and *M. truncatula* root hair inducible (C) and overlaps between *G. max*, *M. truncatula*, *P. vulgaris* and *L. japonicus* nodule-specific (D) orthologous groups;
E and F: Distribution of the number of *G. max* and *M. truncatula* genes which belong to the 11 root hair-inducible (E) and the number of *G. max*, *M. truncatula*, *P. vulgaris* and *L. japonicus* genes which belong to 4 nodule preferential (F) orthologous groups, respectively, according to their expression patterns.

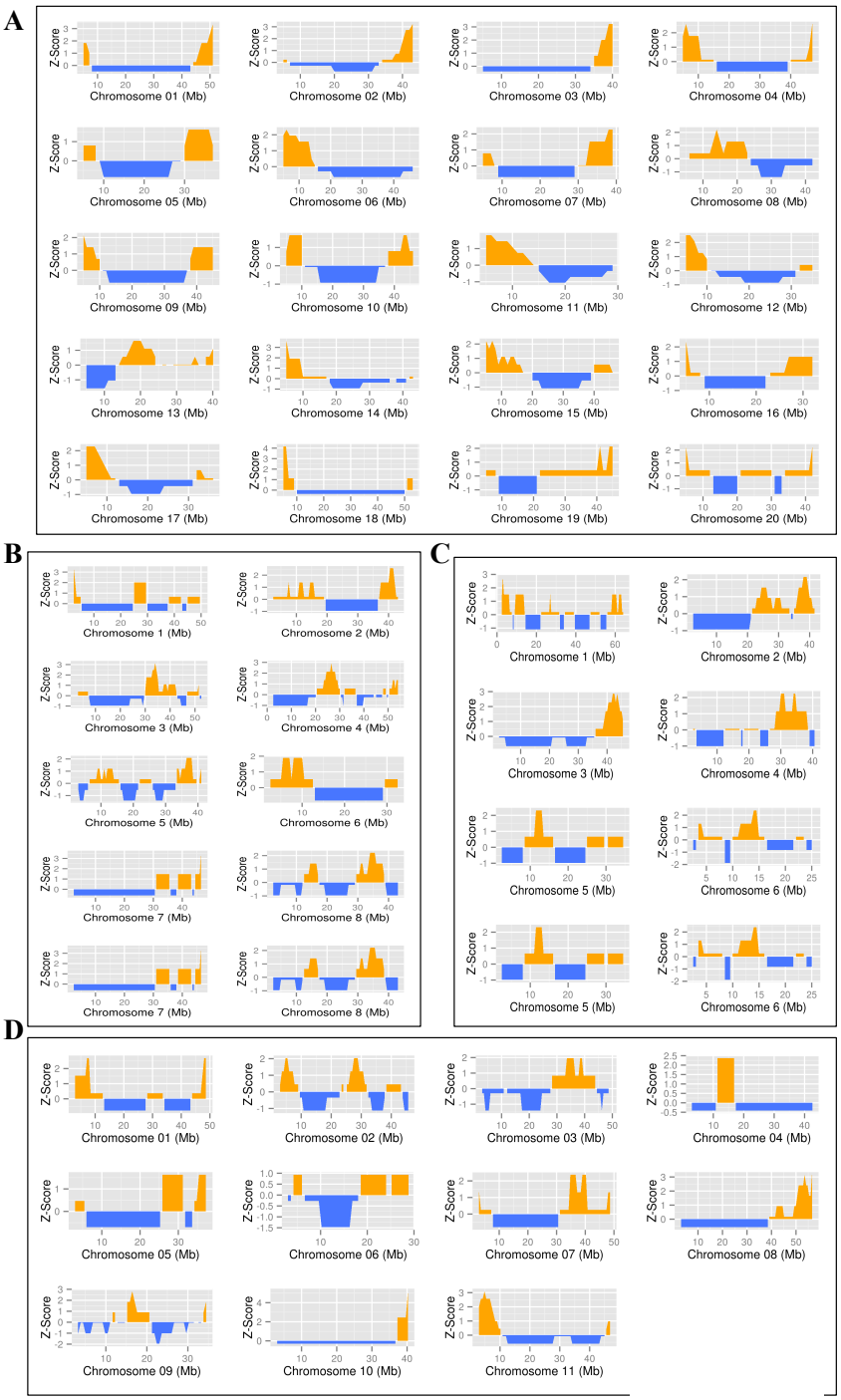


Figure 4- 5 Representation of the density of the legume nodulation genes along each legume chromosome. Gene density is represented by a z-score in a 10 Mb sliding window (step 1 Mb) along each chromosome for each legume genome. Positive values are in orange, negative value in blue. The position on the chromosomes is indicated in Mb. The genomes represented are *Glycine max* (A), *Medicago truncatula* (B), *Lotus japonicus* (C), and *Phaseolus vulgaris* (D).

Supplemental Figure 4- 1 Syntenic relationships between *Glycine max*,*Medicago truncatula*, *Lotus japonicus* and *Phaseolus vulgaris* to reveal orthology and paralogy between nodulation genes.

Supplemental Table 4- 1 Legume orthologous genes controlling nodulation.

Supplemental Table 4- 2 Expression levels of *G. max*, M. *truncatula*, *P. vulgaris* and *L. japonicus* genes controlling nodulation or orthologous to legume genes controlling nodulation.

Supplemental Table 4- 3 *G. max*, M. *truncatula*, *P. vulgaris* and *L. japonicus* paralogous and orthologous genes induced in root hairs upon rhizobia inoculation and/or specifically in nodules.

Supplemental Table 4- 4 Reference genomes, their annotation and the annotation of legume nodulation genes were used to calculate gene density for each legume chromosome.

Supplemental Figure 4-1 and Supplemental Table 4-1, 4-2, 4-3, and 4-4 are not included in this dissertation duo to their large sizes, which can be accessed by the link:

http://journal.frontiersin.org/article/10.3389/fpls.2016.00034/full

Chapter 5: The Gm*FWL1* (*FW2-2-like*) nodulation gene encodes a plasma membrane microdomain-associated protein

Authors: **Zhenzhen Qiao**, Laurent Brechenmacher, Benjamin Smith, Gregory W. Strout, William Mangin, Christopher Taylor, Scott D. Russell, Gary Stacey and Marc Libault

Publication status: This chapter was published in Plant Cell and Environment (DOI: 10.1111/pce.12941), which allowed incorporation into this dissertation.

Author contributions:

Dr. Marc Libault and Gary Stacey designed the research. Dr. Marc Libault and I performed cloning for Co-IP, confocal microscopy and TEM assays. Dr. Christopher Taylor generated vectors. Dr. Marc Libault and Dr. Laurent Brechenmacher carried out Co-IP and proteomic analysis. I performed confocal and transmission electron microscopy with the assistance from Dr. Benjamin Smith, Gregory W. Strout, and Dr. Scott D. Russell. William Mangin, Dr. Marc Libault，and I wrote the article published in Plant cell and Environment.

Abstract

The soybean gene Gm*FWL1* (*FW2.2-like1*) belongs to a plant-specific family that includes the tomato *FW2-2* and the maize *CNR1* genes, two regulators of plant development. In soybean, Gm*FWL1* is specifically expressed in root hair cells in response to rhizobia and in nodules. Silencing of *GmFWL1* expression significantly reduced nodule numbers supporting its role during soybean nodulation. While the biological role of Gm*FWL1* has been described, its molecular function and, more generally, the molecular function of plant FW2.2-like proteins is unknown. In this study, we characterized the role of GmFWL1 as a membrane microdomain-associated protein. Specifically, using biochemical, molecular and cellular methods, our data show that GmFWL1 interacts with various proteins associated with membrane microdomains such as remorin, prohibitins, and flotillins. Additionally, comparative genomics revealed that GmFWL1 interacts with GmFLOT2/4 (FLOTILLIN2/4), the soybean ortholog to *Medicago truncatula* FLOTILLIN4, a major regulator of the *M. truncatula* nodulation process. We also observed that similarly to MtFLOT4 and GmFLOT2/4, GmFWL1 was localized at the tip of the soybean root hair cells in response to rhizobial inoculation supporting the early function of GmFWL1 in the rhizobium infection process.

Introduction

Nodulation is the result of an intimate relationship between rhizobia and legume plants. The interaction between the two partners requires the exchange of chemical signals that include recognition of legume flavonoids by bacteria, which in turn initiates the production of specific *lipo-chitooligosaccharide* signals, the Nod factors (NFs) (Oldroyd et al., 2011a). Upon recognition of the NFs by the plant, a complex molecular and cellular response occurs in the root hair to promote the infection via the formation of the infection thread. Concomitantly, a nodule primordium develops in the root cortex through the initiation of *de novo* cell division. This primordium ultimately grows to form a new organ, the nodule, composed of uninfected cells and cells infected by rhizobia. Ultimately, upon root hair infection, differentiated bacteria, named bacteroids, colonize the nodule in order to fix and provide nitrogen to the plant while receiving photosynthates from the plant.

During the past two decades, researchers have used forward and reverse genetic tools to characterize the role of many legume genes involved in the nodulation process (Oldroyd, 2013). One of the most notable outcomes of this research is the characterization of the symbiotic pathway in legume root hair cells. This pathway is initiated by the recognition of bacterial NFs by the NOD FACTOR RECEPTOR (NFR5), NOD FACTOR PERCEPTION (NFP) and LYSM DOMAIN-CONTAINING RECEPTOR-LIKE KINASE 3 (LYK3) plant membrane associated receptor proteins (Amor et al., 2003; Madsen et al., 2003b; Smit et al., 2007b). More recently, membrane microdomain-associated proteins were also found to regulate the *Medicago truncatula* nodulation process, including *M. truncatula* remorin SYMREM1 (Medtr8g097320) and

flotillins FLOT2 (Medtr3g106420) and FLOT4 (Medtr3g106430) (Haney and Long, 2010b; Lefebvre et al., 2010b). In 2010, Lefebvre *et al.*, (2010) demonstrated the localization of MtSYMREM1 in the plasma membrane and gave direct evidence of its interaction with symbiotic receptors NFP, LYK3 and DOES NOT MAKE INFECTIONS 2 (DMI2) (Amor et al., 2003; Ane et al., 2002; Smit et al., 2007b). Similar to Mt*SYMREM1*, the silencing of Mt*FLOT2* and Mt*FLOT4* genes led to defects in nodule number and development (Haney and Long, 2010b). In addition, these mutants were also affected in the initiation and development of the infection thread. The localization of MtFLOT4 to the infection thread membrane strongly suggests that membrane microdomains act in the signaling cascade involved in infection thread elongation (Haney and Long, 2010b). Altogether, these studies provide strong evidence for a role of membrane microdomains as regulators of legume nodulation.

Previously, we reported that RNAi induced silencing of the Gm*FWL1* (*FW2.2-LIKE1*; Glyma.09G187000) gene resulted in a significant reduction in soybean nodule development (Libault et al., 2010h). This soybean gene belongs to the *FW2-2* (*FRUIT WEIGHT 2.2*) family. *FW2-2* was initially identified as a negative regulator of cell division during tomato fruit development (Frary et al., 2000; Nesbitt and Tanksley, 2001). Subsequently, the *Zea mays* homolog *CNR1* (*CELL NUMBER REGULATOR1*) was found to have a role in cell division, impacting plant development (Guo et al., 2010). More recently, several groups have highlighted the critical role of *FW2-2-like* genes as negative regulators of cell division and fruit development in *Oryza sativa* [*FWL3;* (Xu et al., 2013)], *Physalis floridana* [*CNR1;* (Li and He, 2015)], *Prunus avium* [*CNR12* and *CNR20* (De Franceschi et al., 2013)], and *Persea americana* (Dahan et al.,

2010). In soybean, Gm*FWL1* gene expression is strongly induced during the nodulation process and silencing of this gene reduced nodule cell size and the number of symbionts per nodule (Libault et al., 2010h), suggesting its importance in nodulation. However, our understanding of the molecular function of Gm*FWL1* or Gm*FWL1*-homologs, such as Sl*FW2.2* and Zm*CNR1* (Frary et al., 2000; Guo et al., 2010), is limited. The interaction between SlFW2.2 and the β subunit of casein kinase II is the only prior molecular evidence supporting the role of SlFW2.2 as a negative regulator of cell division during tomato fruit development (Cong and Tanksley, 2006). The amino acid sequence of GmFWL1 did not predict any putative molecular functions and, therefore, in this study, we decided to identify GmFWL1 binding partners using co-immunoprecipitation (co-IP) in order to better understand its molecular function. We found that a large fraction of GmFWL1 binding partners belongs to protein families previously characterized as membrane microdomain-associated proteins. The direct interaction between GmFWL1 and three different soybean prohibitins also confirmed GmFWL1 as a membrane microdomain-associated protein. In addition, GmFLOT2/4, the soybean protein orthologous to MtFLOT2 and MtFLOT4, membrane microdomain-associated proteins controlling *M. truncatula* nodulation (Haney and Long, 2010b), interacts with GmFWL1. Additional microscopic observations revealed punctate plasma membrane localization of GmFWL1 and GmFLOT2/4 in soybean, supporting their membrane microdomain localization. This localization is affected in root hair cells upon rhizobial inoculation with GmFWL1 and GmFLOT2/4 translocated to the tip of the root hair cell in response to inoculation.

Results

GmFWL1 interacts with membrane microdomain-associated proteins

In order to identify GmFWL1 protein partners and to characterize the biochemical function of GmFWL1, co-IP assays were conducted on mature nodules, in which Gm*FWL1* is highly and specifically expressed [i.e., the maximum level of expression of Gm*FWL1* was recorded 32 days after *B. japonicum* inoculation (Libault et al., 2010h)]. Specifically, applying the hairy root transformation method, we expressed both N- and C-terminal HA-tagged GmFWL1 proteins as well as the HA tag alone in mature soybean nodules. Thirty-two days after *B. japonicum* inoculation, wild-type and transgenic nodules expressing Green Fluorescent Protein (GFP) as a marker were collected from three independent biological replicates.

Bioinformatic analysis (Bernsel et al., 2009; Omasits et al., 2014) predicts that GmFWL1 carries one to two transmembrane domains (Supplemental Figure 5-3). In addition, we previously demonstrated that GmFWL1 is localized within the plasma membrane of plant cells (Libault et al., 2010h). Therefore, to maximize the extraction of the GmFWL1-protein complexes from soybean nodules, we used a protein extraction buffer supplemented with 1% Triton-X100 (see Material and Methods) before protein co-IP using an anti-HA-tag antibody.

Prior to performing mass spectrometry analysis of putative GmFWL1 binding partners, we performed western blot analysis with an anti-HA-tag antibody as the primary antibody on each sample and each replicates to validate the expression of the fusion GmFWL1-HA tag proteins (Supplemental Figure 5-2). Using the same protein samples, we conducted mass spectrometry analyses on trypsin-digested proteins

obtained after co-IP using three independent biological replicates (Supplemental Table 5-2). Our analysis revealed 178 proteins specifically co-immunoprecipitated with the N- and C-terminal GmFWL1-HA tagged proteins, but not against the HA tag alone and wild type nodules, the two negative controls.

Among these 178 putative GmFWL1 protein partners, 62 belong to known membrane microdomain-associated protein families, such as membrane intrinsic proteins (i.e., aquaporins), SPFH/Band 7/PHB domain-containing membrane-associated proteins (i.e., flotillins and prohibitins) (Browman et al., 2007; Rivera-Milla et al., 2006), remorins, proton-ATPases, vacuolar-ATPases, phospholipase Ds, receptor kinases, leucine-rich repeat proteins and proteins involved in vesicle trafficking (Figure 5-1) (Brechenmacher et al., 2012a; Jarsch et al., 2014; Lefebvre et al., 2007; Morel et al., 2006; Takahashi et al., 2013a, b; Takahashi et al., 2012). In addition, independent of the direction of the fusion between the HA tag and the GmFWL1 cDNA, we consistently identified three membrane intrinsic proteins, one proton-ATPase, three vacuolar-ATPases, and three SPFH/Band 7/PHB domain-containing membrane-associated proteins (i.e., two prohibitins and one flotillin), as GmFWL1 protein partners across the three independent biological replicates (Supplemental Table 5-2; highlighted in red). These results suggest that GmFWL1 preferentially interacts with microdomain-associated proteins; findings consistent with the plasma membrane localization of the GmFWL1 protein and the prediction of at least one transmembrane domain in the FWL1 protein (Supplemental Figure 5-3).

GmFWL1 directly interact with prohibitin proteins, markers of plasma membrane microdomains

Independent of the tag or promoter used to express tagged proteins, co-IP technology can be prone to artifacts [e.g., some proteins, such as ribosomal proteins, are hypothesized to stick to the microdomain fraction during protein extraction (Alexandersson et al., 2004; Lefebvre et al., 2007)]. Accordingly, putative interactions between proteins must be validated using other methods (Swatek et al., 2014). Accordingly, based on the repeated identification of different prohibitins as GmFWL1 interactors (Supplemental Table 5-2) and previous studies establishing prohibitins as markers of plasma-membrane microdomains (Browman et al., 2007; Rivera-Milla et al., 2006), we applied the split luciferase assay to validate the interaction between GmFWL1 and prohibitins.

In soybean, we identified a total of 21 genes encoding prohibitins (Supplemental Figure 5-4). Among them, based on our co-IP assays, five encode proteins that interact with GmFWL1 (i.e., Glyma.05G029800, Glyma.13G065000, Glyma.16G204900, Glyma.17G097000 and Glyma.19G020000). Mining the soybean transcriptome atlas (Libault et al., 2010g), Glyma.05G029800, Glyma.13G065000 and Glyma.19G020000 are the strongest expressed soybean prohibitin genes in nodules. In addition, Glyma.05G029800 is predominantly expressed in soybean nodules compared to other soybean organs (Supplemental Figure 5-4). The strong expression of these 3 prohibitin genes and Gm*FWL1* in nodules support their potential to directly interact. Fusion proteins with the N- and C-terminal domains of the luciferase fused with prohibitins or GmFWL1 were transiently expressed in tobacco epidermal cells under the control of the

CaMV 35S promoter. To overcome the likely variation in expression of the various split-luciferase transgenes, we generated seven independent biological replicates. While the negative controls [i.e., co-transformed tobacco leaves with: 1) the N or C-terminal fusion protein and the N or C-terminal domain of the luciferase alone; and 2)- with N or C-terminal fusion proteins between the N or C-terminal domain of the luciferase, and GmFWL1 and the Glyma.13G293600 prohibitin. The latter was not identified as a GmFWL1 interactor according to our co-immunoprecipitation assay (Supplemental Table 5-2)] lacked detectable luciferase activity, we observed luciferase activity upon expression of GmFWL1 and the three prohibitins independent of the placement of the luciferase subunits (Supplemental Figure 5-5). These data support the direct interaction in and between GmFWL1 and the three prohibitin proteins.

Microscopic evidence for membrane microdomain localization of GmFWL1 and its role during the early stages of the nodulation process

Biochemical assay is not sufficient to conclude about the membrane microdomain localization of proteins (Raffaele et al., 2009; Tanner et al., 2011). In order to further validate the role of GmFWL1 as a membrane microdomain-associated protein, we used laser scanning confocal microscopy (LSCM) to determine the sub-cellular localization of GmFWL1. Our previous observations using tobacco epidermal cells and protoplasts revealed punctate localization of GmFWL1 in the plasma membrane (Libault et al., 2010h). Using LSCM on tobacco protoplasts and epidermal cells, we confirmed this punctate plasma membrane localization of N- and C-terminal fusion proteins between GmFWL1 and the Green Fluorescent Protein (GFP) (Figure 5-

2; Supplemental Videos 5-1 and 2; green channel). This same punctate localization has been shown to be characteristic of other known FWL proteins, such as the rice FWL proteins (Xu et al., 2013). In addition, similarly to *M. truncatula* SYMREM1 (Lefebvre et al., 2010b), GmFWL1 localization upon tobacco cell plasmolysis remains between the PM and cell wall (Libault et al., 2010h), which is compatible with plasmodesmata and membrane microdomain localization.

A previous study highlighted the translocation of MtFLOT4, a membrane microdomain-associated protein controlling *M. truncatula* nodulation, to the tip of the root hair cells upon rhizobial inoculation (Haney and Long, 2010b). We hypothesized that GmFWL1 might show a similar translocation in soybean root hairs in response to *B. japonicum* inoculation based on the interaction between soybean flotillins and GmFWL1, the specific induction of GmFWL1 expression in inoculated root hairs and the role of GmFWL1 during soybean nodulation (Libault et al., 2010h). To validate our hypothesis, we generated transgenic soybean roots expressing the GFP-GmFWL1 fusion proteins under the control of the native Gm*FWL1* promoter (Libault et al., 2010h) or the cassava vein mosaic virus (CVMV) promoter, and inoculated these plants with *B. japonicum*. When expressed under the control of the CVMV promoter, the GFP-GmFWL1 fusion protein showed a punctate localization in mock-inoculated root hairs (Figure 5-3A, B, E and F; Supplemental Figure 5-6). When expressed under the control of the Gm*FWL1* promoter sequence, we observed a weak GFP signal in mock-inoculated root hair cells (Supplemental Figures 5-6, 7A, 7B, 7I and 7J). This result is likely the consequence of the low level of activity of the Gm*FWL1* promoter sequence in uninoculated root hair cells (Libault et al., 2010h). Within 24 hours after *B.*

japonicum inoculation, the fusion protein aggregated to the tip of the root hair cell independently of the promoter used to drive the expression of the GFP-GmFWL1 fusion proteins (Figure 5-3C and D; Supplemental Figures 5-6, 7C and 7D). Seven days after inoculation, similar to MtFLOT4 (Haney and Long, 2010b), GFP-GmFWL1 was highly concentrated at the tip of the root hair cells (Figure 5-3G and H; Supplemental Figures 5-6, 7K and 7L). Placement of the GFP at the C-terminus (i.e., GmFWL1-GFP) produced a similar punctuate plasma membrane localization and a specific root hair tip translocation in response to rhizobial inoculation (Supplemental Figure 5-8B and D). In contrast, control plants expressing only the unfused GFP protein showed localized in the root hair cytosol and nucleus (Supplemental Figure 5-8A and C).

GmFWL1 punctate localization is conserved in nodule plasma membranes

GmFWL1-silenced nodules are characterized by fewer bacteroids suggesting a role of GmFWL1 in controlling the infection of the nodule cells by the bacteroids (Libault et al., 2010h). In addition, our initial proteomic analysis revealing GmFWL1 protein partners was conducted using mature nodule extracts, suggesting a putative role of membrane microdomains during the late events of nodulation. Therefore, in order to validate the localization of GmFWL1 as a plasma membrane microdomain-associated protein in the nodule, we applied high-resolution transmission electron microcopy (Indrasumunar et al.) combined with immunogold-labeling.

We chose to express N- and C-terminal HA-tagged GmFWL1 fusion proteins in soybean nodules under the control of the native GmFWL1 promoter sequence, since a specific antibody to GmFWL1 is unavailable. We observed a diffuse distribution of a

limited number of gold nanoparticles when expressing the HA-tag alone (Figure 5-4). In contrast, we repeatedly observed an accumulation of gold particles in clusters in a limited number of discrete regions of the plasma membrane, infection threads and the symbiosome membrane independent of the direction of the fusion between GmFWL1 and the HA tag (Figure 5-4). These obervations support the membrane microdomain localization of GmFWL1 in mature nodules. Taking into consideration the limited infection of the GmFWL1-silenced nodules by rhizobia (Libault et al., 2010h), our results suggestathe role of GmFWL1 and, more globally, the role of membrane microdomains during the mutualistic interaction between plant cells and bacterial symbionts.

Legume comparative genomics reveal the direct interaction between GmFWL1 and GmFLOT2/4, the soybean ortholog to MtFLOT2 and MtFLOT4

Among the various GmFWL1 protein partners, several belong to the SPFH/Band 7/PHB domain-containing membrane-associated protein and to the remorin families, two microdomain markers (Supplemental Table 5-2). Members of these protein families were previously characterized as regulators of legume nodulation, such as the *M. truncatula* flotillins FLOT2 (Medtr3g106420) and FLOT4 (Medtr3g106430) and the remorin Mt SYMREM1 (Haney and Long, 2010b; Lefebvre et al., 2010b). More recently, taking advantage of the release of the *M. truncatula* and *L. japonicus* genome sequences (Sato et al., 2008b; Young et al., 2011b), LjSYMREM1 was identified and its role during *L. japonicus* nodulation was validated (Toth et al., 2012).

We hypothesized that the soybean flotillin (Glyma.06G065600) and remorin proteins (Glyma.07G073100 and Glyma.16G189500) that interact with GmFWL1 are orthologous to MtFLOT2/4 and MtSYMREM1 and, therefore, looked for microsynteny relationships between these genes in the genomes of soybean and medicago (Schmutz et al., 2010b; Young et al., 2011b).

Using comparative genomic resources (Lyons and Freeling, 2008), we identified four soybean genes (i.e., Glyma.05G205900, Glyma.08g012800, Glyma.01g218000, Glyma.11g025200) displaying significant microsynteny with MtSYMREM1 (Supplemental Figure 5-9). While Glyma.05G205900 and Glyma.08g012800 are highly expressed in nodules (Supplemental Figure 5-10), none of these 4 soybean genes encode a GmFWL1 binding protein based on our co-IP results.

Performing a similar analysis, we characterized Glyma.06G065600 and Glyma.04G064400 as Mt*FLOT4* and Mt*FLOT2* soybean orthologous genes based on their macro- and microsynteny relationships (Supplemental Figure 5-11). Taking in consideration the tandem duplication of Mt*FLOT2* and Mt*FLOT4* genes, we also conclude that the emergence of the two medicago flotillin genes occurred after the speciation of *Glycine max* and *Medicago truncatula*.

To provide more evidence of the putative conservation and divergence of the biological function between Glyma.06G065600, Glyma.04G064400, Mt*FLOT4* and Mt*FLOT2*, we mined the soybean transcriptome atlas (Libault et al., 2010g). As expected, similarly to Mt*FLOT2* and Mt*FLOT4* (Haney and Long, 2010b), Glyma.06G065600 is preferentially expressed during nodule development. However, the Glyma.04G064400 transcript was not detected in nodules (Supplemental Figure 5-

11). This result suggests that the function of Glyma.04G064400 diverged from Glyma.06G065600 after the duplication of the soybean genome. Comparing the amino acid sequence of the proteins encoded by Glyma.06G065600, Mt*FLOT2* and Mt*FLOT4*, these proteins share 81% and 84% sequence identity, respectively. Accordingly, we hypothesized that the protein encoded by Glyma.06G065600 serves as a microdomain-associated protein in soybean nodules (Raffaele et al., 2009; Haney and Long, 2010; Li et al., 2012; Jarsch et al., 2014). To verify this hypothesis, we performed LSCM observations on tobacco epidermal cells expressing a fusion between the Glyma.06G065600 protein and the red fluorescent mCherry protein. We observed that the Glyma.06G065600 protein localized in a punctuate pattern in the plasma membrane which is typical of microdomain-associated proteins (Supplemental Video 2; red channel). Together, these results strongly support that Glyma.06G065600 is the Mt*FLOT2* and Mt*FLOT4* functional ortholog in soybean and, as a consequence, is named Gm*FLOT2/4*.

GmFWL1 and GmFLOT2/4 accumulate at the tip of soybean root hair cells in response to B. japonicum inoculation

The interaction between GmFWL1 and GmFLOT2/4 (Supplemental Table 5-2) suggests that at least a subset of these proteins should co-localize in the root hair cell. Similar to GmFWL1, we hypothesize that GmFLOT2/4 accumulates at a certain level at the tip of the soybean root hair cell in response to *B. japonicum* inoculation. To verify our hypothesis, we took advantage of hairy root transformation to express a mCherry-GmFLOT2/4 fusion protein in soybean root hairs under the control of the CvMV or

Gm*FWL1* promoters. This transgenic material was treated with a suspension of *B. japonicum* or with the plant nutritive solution (i.e., mock-inoculated condition; (Broughton and Dilworth, 1971a)). Similarly to GmFWL1 (Figure 5-3G-H; Supplemental Figures 5-6 and 7), GmFLOT2/4 under the control of the CvMV promoter showed diffuse, punctate localization in mock-inoculated soybean root hair cells (Figure 5-5A, B, E and F; Supplemental Figure 5-6). However, within 24 h after inoculation with *B. japonicum*, GmFLOT2/4 could be seen concentrated at the tip of the root hair cells (Figure 5-5C and D; Supplemental Figures 5-7, 7G and 7H), leading to high tip accumulation 7 days after inoculation (Figure 5-5G-I; Supplemental Figures 5-6, 7O and 7P). We repeatedly observed these accumulations independently of the promoter used to drive the expression of the protein fusions (Figure 5-5 and Supplemental Figures 5-6 and 7). The similar localization and timing of the translocation of the GmFWL1 and GmFLOT2/4 proteins in response to rhizobia at the tip of the soybean root hair are consistent with their interaction. The roles of Gm*FWL1* and Mt*FLOT4* during legume nodulation (Haney and Long, 2010b; Libault et al., 2010h), their similar translocation at the tip of the root hair cells in response to rhizobia inoculation, and the interaction between GmFWL1 and GmFLOT2/4 reinforces the global role of membrane microdomain-associated proteins during the nodulation process.

Discussion

GmFWL1 encodes a membrane microdomain-associated protein

Our biochemical assays and microscopic observations demonstrate that Gm*FWL1* encodes a membrane microdomain-associated protein. Notably, we identified 62 membrane microdomain-associated proteins interacting with GmFWL1 (i.e., 8 membrane intrinsic proteins, 6 SPFH/Band 7/PHB domain-containing membrane-associated proteins, 2 remorins, 8 proton-ATPases, 14 vacuolar-ATPases, 3 phospholipase Ds, 2 receptor kinases, 4 leucine-rich repeat proteins and 15 proteins involved in vesicle trafficking).

Hypothesizing that any plant *FWL* genes encode plasma membrane microdomain-associated proteins, we mined the scientific literature looking for plant FWL proteins as part of plant microdomain proteomes, being cautious about the fact that detergent resistant membrane (DRM) proteins would require additional microscopic observations before unequivocally concluding a membrane microdomain localization (Raffaele et al., 2009; Tanner et al., 2011). In addition to repeated identification of members of protein families characteristic of the detergent-insoluble fraction of *M. truncatula* root plasma membrane, Lefebvre et al. (2007) identified MtC00726_1, a Cys-rich protein (Lefebvre et al., 2007). Further sequence analysis of the MtC00726_1 nucleotide sequence allowed us to characterize this annotation as Medtr8g104870, a gene encoding a 191 amino acid protein sharing 40% identity with GmFWL1. While macro- and micro-synteny analyses did not reveal an orthologous relationship between Medtr8g104870 and Gm*FWL1*, we identified Medtr8g104870 as orthologous to Gm*FWL3* (Supplemental Figure 5-12), a second soybean *FWL* gene

strongly induced in root hairs upon *B. japonicum* inoculation and in nodules (Libault et al., 2010h). GmFWL3 also co-immunoprecipitated with GmFWL1 (Supplemental Table 5-2). A second proteomic study performed on *Arabidopsis thaliana* detergent-resistant membranes also highlighted the presence of a 151 amino acid protein (At1g14880) sharing 32.9% of identity with GmFWL1 and previously identified as a GmFWL1 homolog (Libault et al., 2010h). These two independent studies reinforce our conclusion that FWL proteins are plasma membrane-associated proteins including GmFWL1, consistent with a central role for membrane microdomains as regulators of legume nodulation.

To complement our biochemical assays, both confocal and electron microscopic observations of GmFWL1 protein revealed a punctate localization in the plasma membrane from various heterologous (i.e., tobacco epidermal cells and protoplasts) and homologous plant systems (soybean root hair and nodule cells). This specific sub-plasma membrane localization is one of the attributes of membrane microdomain-associated proteins (Brechenmacher et al., 2012a; Haney and Long, 2010b; Jarsch et al., 2014; Raffaele et al., 2009). In addition, the similar translocation of GmFWL1 and MtFLOT4 (Haney and Long, 2010b) at the tip of the soybean and medicago root hair cell in response to rhizobia combined with our macro/micro-syntenic and biochemical analyses (i.e., our data support the interaction between GmFWL1 and GmFLOT2/4, a protein orthologous to MtFLOT4), support the formation of a microdomain protein complex at the root hair cell tip. The co-localization of GmFWL1 and GmFLOT2/4 in soybean root hair cells upon rhizobial inoculation also supports the interaction between

these two proteins. This protein complex may act to catalyze the initial endocytotic penetration of the root hair cell by the infecting rhizobia.

One striking feature of Gm*FWL1* is its strong and specific expression in nodules and in *B. japonicum*-inoculated root hair cells. Mining the soybean transcriptome atlas (Libault et al., 2010g), 14 out of the 178 GmFWL1 protein partners are specifically expressed in nodules (Supplemental Figure 5-13). Among these genes, we identified two SPFH/Band 7/PHB proteins characterized by their apparent co-localization with GmFWL1 (i.e., Glyma.05G029800 and Glyma.06G065600; Supplemental Table 5-2 and Supplemental Figure 5-13) as well as Gm*FWL3* (Glyma.08G043500) (Libault et al., 2010h). Based on the hypothesis that the proteins encoded by these 14 genes have no or a limited rate of organ-to-organ translocation upon their biosynthesis, the transcriptomic data suggest that the proteomic composition of the nodule membrane microdomains is unique compared to other plant organs. This specificity might be essential for the efficient establishment of the symbiosis between legumes and rhizobia as demonstrated by the key role of GmFWL1, MtFLOT2, and MtFLOT4 during the nodulation process (Libault et al., 2010h); (Haney and Long, 2010b).

Function of plasma membrane microdomains during legume nodulation

GmFWL1 is important for efficient nodulation of soybean (Libault et al., 2010h). In this study, we highlighted the microdomain localization of GmFWL1 leading us to revisit the role of membrane microdomains in plant cells and more specifically in the context of their infection by mutualistic symbiotic bacteria.

SYMREM1 and FLOT2/4 were previously described as regulators of *M. truncatula* and *L. japonicus* nodulation (Haney and Long, 2010b; Lefebvre et al., 2010b; Toth et al., 2012). Hence, the current data indicates that GmFWL1, a third membrane microdomain-associated protein, is critical for legume nodulation. We assume that additional players interacting with GmFWL1 and preferentially expressed during the nodulation process, such as GmFWL3, should also be included in this list.

It was proposed that MtFLOT2 and MtFLOT4 control the invagination of the root hair plasma membrane, leading to the development of the infection thread (Haney and Long, 2010b). This hypothesis is supported by previous studies across various eukaryotic models showing the role of membrane microdomain-associated proteins in promoting endocytosis (Brechenmacher et al., 2012a). We propose that GmFWL1 also participates in this process. Rhizobia inoculation would be necessary to activate the clustering of membrane microdomains as demonstrated by the accumulation of GmFWL1, GmFLOT2/4 and MtFLOT4 protein at the tip of the soybean and medicago root hair cells [Figure 5-3, Figure 5-5, Supplemental Figures 5-6 and 7 (Haney and Long, 2010b)]. Following membrane microdomain clustering, the plasma membrane will invaginate and the infection thread will start to progress through the root hair cell. It is also possible that GmFWL1 and membrane microdomains play a role in the endocytosis of bacteroids into the infected cells. This second function is consistent with our findings that the GmFWL1-HA tag fusion protein was localized to the plasma membrane of infected cells in the nodule, the infection threads and the symbiosome membrane (Libault et al., 2010h) (Figure 5-4). The marked reduction of bacteroid numbers in nodules formed on plants where Gm*FWL1* expression was silenced is also

consistent with the notion that FWL1 is required for bacterial release from the infection thread (Libault et al., 2010h).

Role of plant FWL proteins

While the biological importance of the *FWL/CNR/PCR* genes has been noted, the molecular role of these genes in plants is less defined. To date, yeast-two-hybrid assays revealed an interaction between the SlFW2-2 protein and a subunit of casein kinase II (Cong and Tanksley, 2006). The current data provide a list of other proteins that interact with GmFWL1 and clearly point to the localization of this protein in membrane microdomains such as prohibitins, flotillins, and remorins. Our analysis also revealed the homodimerization of both GmFWL1 and soybean prohibitins (Supplemental Figure 5-5). Taking in consideration that prohibitins do not carry a transmembrane domain, we propose a model where a network of hetero- and homo-dimers of prohibitins are anchored to the plasma membrane microdomains through, at least in part, their interactions with GmFWL1. The dimerization of GmFWL1 and prohibitins will insure the propagation of this network and the formation of plasma membrane microdomains. Similarly, independent studies mention the homodimerization of flotillin proteins in human cells (Babuke et al., 2009; Fernow et al., 2007; Frick et al., 2007; Neumann-Giesen et al., 2004; Solis et al., 2007). We hypothesize that flotillin homodimers and flotillin-FWL1 heterodimers might also be part of this microdomain-associated protein network.

Plant membrane microdomains have multiple biological functions: cell-to-cell communication as reflected by their accumulation around plasmodesmata, signal

transduction such as auxin signaling by controlling the accumulation of auxin transporter PIN1 in the plasma membrane (Titapiwatanakun et al., 2009), plant response to pathogenic (Raffaele et al., 2009) and mutualistic microbes (Haney and Long, 2010b), and endocytosis (Fan et al., 2015). These various molecular functions could reflect the formation of membrane complexes with unique compositions that would then mediate different aspects of plant biology, including plant organ development and legume nodulation. Accordingly, the characterization of the protein partners of plant FWL proteins and the downstream cascades activated by these plasma membrane microdomains are likely two important avenues to better understand the role of *FWL* genes and their critical roles in plant development. As an example, the identification of several genes encoding GmFWL1 protein partners specifically expressed in nodules suggests at least the potential to form a nodule-specific membrane microdomain complex. The disruption of the biological activity of one or several of these proteins such as MtFLOT4 (Haney and Long, 2010b) and GmFWL1 (Libault et al., 2010h) would then lead to a defect in the role of the membrane microdomains and a disruption of the nodulation process. Haney and Long (2010) suggested a potential role of MtFLOTs in the endocytosis and trafficking of rhizobia. We suggest that GmFWL1 could play a similar role based on its discrete localization in the ITs of the nodule and in the symbiosome membrane.

Material and Methods

Bacterial culture

Escherichia coli (DH5α), *Agrobacterium rhizogenes* (K599) and *Agrobacterium tumefaciens* (GV3101) were grown in LB medium supplemented with appropriate antibiotics at 37 °C for *E. coli* and 30°C for agrobacteria. *Bradyrhizobium japonicum* USDA110 was grown in HM medium and washed in the plant nutritive solution ((Broughton and Dilworth, 1971a)) previously to any inoculation as previously described (Libault et al., 2009c).

Clonings

Co-IP assay: The cDNA sequence of Gm*FWL1* (See Supplemental Table 5-1 for primer sequences) was cloned downstream and upstream to the HA tag into CGT3304 and CGT3305 vectors (Supplemental Figure 5-1), which were previously digested with *BglII* and *SacI*, and by *BamHI* and *SacI*, respectively. Upon validation of the conservation of the frames between the HA tag and the GmFWL1 cDNA, the *pFMV-HAtag-GmFWL1 cDNA-tNOS* and *pFMV- GmFWL1cDNA-HAtag-tNOS* cassettes were integrated into the pAKK1467B vector upon digestion by *SdaI* which is also known as *Sse8387I* (Collier et al., 2005; Govindarajulu et al., 2009).

Split-luciferase assay: To test the direct interaction between GmFWL1 and three soybean prohibitin proteins (i.e., Glyma.05G029800, Glyma.13G065000, and Glyma.19G020000), we applied the split-luciferase system. First, the GmFWL1 and prohibitin cDNA sequences surrounded by *KpnI*, *BamHI* and *SalI* restriction sites after

PCR reaction (see Supplemental Table 5-1 for primer sequences) were cloned into the PUC19-cLUC and PUC19-nLUC vectors (Chen et al., 2008). The various transgenes (i.e., *CaMV35S:cDNA-n/cLUC* as well as *CaMV35S:n/cLUC* negative controls) were amplified by PCR to include *PmeI* restriction sites on 5' and 3'. The DNA fragments of interest were purified on a gel before being cloned into the *PmeI*-digested pE3185 binary vector.

Epifluorescent microscopy assay: Translational fusions between Gm*FWL1* and *GFP* cDNA under the control of the dual 35S promoter were performed using the Gateway system (Invitrogen, Carlsbad, CA, USA). Gm*FWL1* cDNA was amplified using GmFWL1 cDNA AttB1-AttB2 Forward and Reverse primers followed by the AttB1 Forward and AttB2 Reverse primers (Supplemental Table 5-1). The Gateway compatible PCR product was introduced first in pDONR-Zeo (Invitrogen, Carlsbad, CA, USA) and then into pMDC43 and pMDC83 vectors (Curtis and Grossniklaus, 2003) using the Gateway BP and LR Clonase II enzyme mixes (Invitrogen, Carlsbad, CA, USA). The phase between GFP and Gm*FWL1* cDNA sequences and the integrity of Gm*FWL1* cDNA sequence were verified by sequencing.

The cloning of *pCvMV: GFP-GmFWL1, pCvMV: GmFWL1-GFP* and *pCvMV:mCherry-GmFLOT2/4* into pDONR221 vectors was performed using the 3-fragments multisite Gateway® system (Invitrogen, http://www.invitrogen.com/). The cloning of the respective controls (i.e., *pCvMV: GFP*, *pCvMV:mCherry*) into pDONR221 vectors were generated using the 2 fragments multisite Gateway® system (Invitrogen, http://www.invitrogen.com/). Gateway compatible PCR products were

generated by two independent and successive PCR reactions to generate 50% then 100% of the corresponding AttB box. The entry vectors product of the BP Clonase reaction (Invitrogen, http://www.invitrogen.com/) and containing a promoter, tag, and cDNA sequences were recombined into the Gateway compatible pAKK1467B binary vector using the Gateway® LR Clonase® II plus enzyme mix. To modify the pAKK1467B vector, an AttR1-CmR-ccdb-AttR2-tNOS cassette was amplified by PCR using pMDC43 as a template (Curtis and Grossniklaus, 2003) (see Supplemental Table 5-1 for CGT3304-AttR1-tNos-for and CGT3304-AttR1-tNos-rev primer sequences) and cloned into pAKK1467B preliminary digested by *SbfI*.

TEM assay: The *pFWL1: HA-cFWL1*, *pFWL1:cFWL1-HA* and *pFWL1: HA* transgenes were cloned into a modified AKK1467B binary vector carrying the attR1-CmR-ccdB-attR2 Gateway cassette. In order to generate these three transgenes and directly insert them into our Gateway compatible binary vector, we took advantage of the 3- and 2-fragments multisite Gateway® system (Invitrogen, http://www.invitrogen.com/). The Gm*FWL1* promoter sequence used to drive the expression tagged GmFWL1 protein has been characterized for its nodule specific activity (Libault et al., 2010h). The HA tags surrounded by selected AttB Gateway cassettes were synthetized by Integrated DNA Technologies using gBlock technology. The various DNA fragments were cloned into the appropriate pDONR vectors before to perform a multiple recombinations into the Gateway compatible AKK1467B vector. The fidelity of the generated clones was verified by sequencing. The plasmids were then electroporated into *A. rhizogenes* strain

K599 and *A. tumefaciens* strain GV3101 to generate transgenic plant material (see below).

Co-IP assay and mass-spectrometry analysis

Soybean transgenic nodules were generated using the hairy root transformation protocols as describe by Libault et al., (2009) (Libault et al., 2009c). Thirty-two days after *B. japonicum* inoculation of the transgenic soybean root, nodules were collected and used fresh when purifying GmFWL1 protein complexes.

Preliminary to any co-IP assay, total protein extractions were conducted using the following buffer: 50mM Tris-MES pH 7.5, 300mM sucrose, 150mM NaCl, 10 mM potassium acetate, 5mM EDTA, Sigma plant protease inhibitor cocktail, 1% Triton-X 100. To maximize our yield in protein and to remove contaminants, the chopped soybean nodules were incubated in the extraction buffer on ice for 30 minutes, filtrated through Miracloth and then centrifuged (13,000 rpm, 10 min, 4°C). The supernatant containing the resuspended proteins was used for immunoprecipitation assay using anti-HA microbeads and the μMACS Epitope Tag Protein Isolation Kit according to the manufacturer's protocol (Miltenyi Biotech). 250 to 500 μg of soluble proteins were loaded for each co-IP assay. For elution, the native protein elution protocol was used, using 50ul 0.1M Triethylamine pH11.8/0.1% TX-100. Samples were neutralized with 3ul of 1M MES pH3.

Proteins resulting from the co-IP were separated by SDS-PAGE (12%). The gel lane containing the proteins was divided into 8 gel bands. In gel trypsin digestion was performed according to Brechenmacher et al., (Brechenmacher et al., 2009b) with the

following modifications: proteins were reduced in 10 mM DTT for 30 min at 56°C and alkylated in 50mM IAA for 30 min before proceeding to the trypsin digestion. After enzymatic digestion, peptides were extracted twice using a 60% acetonitrile, 39% water and 1% trifuoroacetic acid. Peptides were then lyophilized and resuspended in 1% formic acid in water.

The tryptic peptides were analyzed by liquid chromatography (Petersson et al.) tandem mass spectrometry (MS/MS) on an Agilent 6520 quadrupole time-of-flight (Q-TOF) mass spectrometer using a chip cube interface. The LC and the Q-TOF were both controlled by MassHunter. The solvents used for the LC system are the A1 (97% water, 2.9% acetonitrile, 0.1% formic acid) and B1 (90% acetonitrile, 9.9% water; 0.1% formic acid) solutions. Five µl of tryptic peptides were loaded onto a C18 trap column of an Agilent ProtID-Chip (ref# G4240-62001) operating in enrichment mode using the capillarity pump of the LC system at a flow rate of 3 µl min^{-1} and 3% B1. Once the sample was loaded onto the enrichment column, the chip valve was switched from enrichment to analysis mode and the elution of the peptides was performed using a 25 min linear gradient from 3% to 50% B1 at 0.6 µl min^{-1}. The peptides were electrosprayed into the Q-TOF using an ionization voltage of 1950V. The Q-TOF was operated in positive auto MS/MS mode. The precursor ions with a m/z comprised between 300 and 2500 were acquired at a scan rate of 250 ms/spectrum and the 5 most abundant precursors for each cycle having a charge higher than +1 and an intensity of at least 2000 counts was fragmented by collision induced dissociation (CID). Fragment ions having an m/z comprised between 70 and 2500 were acquired at a scan rate of 500 ms/spectrum. An active exclusion which was released after 1 spectrum and 0.1 min was

applied to avoid re-acquiring the same precursor. The generated data files were imported into Agilent MassHunter qualitative analysis software and converted into a Mascot Generic Format (MGF) file using the default parameters.

Tryptic peptides were also analyzed on an LTQ Orbitrap XL mass spectrometer operated with Xcalibur (version 2.0.7) and coupled to a nanoflow Proxeon-EasynLC system (Thermo Scientific, Waltham, MA). Five ul of tryptic peptides were loaded onto a C18 peptide Cap Trap (Michrom Bioresources, Auburn, CA). Peptides were eluded using a 25 min linear gradient from 5 to 45% of solvent B (0.1% formic acid in acetonitrile) at a flow rate of 0.4 µl/min. and separated on fused silica analytical column (150 µM ID x 100 mm) packed with C18 (5 µM, 100 Å; Michrom Bioresources). Solvent A was 0.1% formic acid in MilliQ water. Peptides were then electroprayed in the LTQ Orbitrap operating in positive mode. The Orbitrap first performed a full MS scan at a resolution of 30000 FWHM to detect the precursor ion having an m/z comprised between 300 and 2000 and a +2 or higher charge. CID (35% normalized collision energy) was used to fragment the nine most intense precursor ion (1000 counts minimum) of each full scan which was analyzed by the LTQ linear trap.

Xcalibur raw data (Thermo) or mgf files (Agilent) were imported into Sorcerer (version 3.5; Sage-N research, Milipitas, CA) which extracted the peaks and used a Sequest algorithm (version 27) to match the MS and MS/MS spectra to protein sequences. The database (Glyma 1.0) used in this study contains 75778 soybean protein sequences deduced from the soybean genomic sequence. Search parameters for tryptic peptides included carbamidomethylation of cysteine and oxidation of methionine as fixed and variable medications, respectively. One miscleavage for trypsin was allowed.

Mass tolerances were set at 10 ppm and 1.0 Da for the precursor ion and fragmented ions, respectively. The results of the different searches were further analyzed using Scaffold (version 3.4.9; Proteome software, Portland, OR) to validate MS and MS/MS data. Only proteins identified with at least 2 peptides and a confidence higher than 99% were kept for further analysis. Protein abundance was determined by mass spectral counting and normalized in Scaffold using the quantitative value function. Three independent biological replicates were generated and analyzed.

Plant transformation, treatment, and observation

To ensure the reproducibility of the results, microscopic experiments were repeated at least three times on different sets of tobacco or soybean plants grown at different times.

Split-Luciferase assay: *Nicotiana benthamiana* leaves were infiltrated with *A. tumefaciens* transformed with pE3185 vectors carrying the N and Cter domain of the luciferase fused to GmFWL1 and three soybean prohibitins (i.e., Glyma.05G029800, Glyma.13G065000, and Glyma.19G020000). To optimize the transient expression of these various transgenes, we also transform tobacco leaves to express the viral protein HC-Pro.

Epifluorescent microscopy: *Nicotiana benthamiana* leaves were infiltrated with *A. tumefaciens* carrying the following constructs: *2x35S: GFP-GmFWL1cDNA*, *2x35S: GFP*, *pCvMV:mCherry-GmFLOT2/4* and *pCvMV: GFP-GmFWL1* vectors. To optimize

the transient expression of these various transgenes, we also transformed tobacco leaves to express the viral protein HC-Pro. Tobacco leaf cells co-expressing *pCvMV: GFP-GmFWL1* and *pCvMV:mCherry-GmFLOT2/4* fusions were observed using an Olympus FluoView 500 LSCM with an argon 488 nm and a helium-neon 543 nm laser lines, and a 60x/1.00NA water immersion objective. Three dimensional images were acquired by taking a series of optical sections along the Z-axis, with an over-all voxel dimension of 410 nm × 410 nm ×410 nm.

To produce tobacco leaf protoplasts, leaf epidermal cells were incubated overnight, 2 days after infiltration, under dark conditions at room temperature, in MOPS medium (9% mannitol, 0.037%KCl, 0.2M MOPS, pH 6.0) supplemented with 0.05% driselase, 0.02% macerozyme R10 and 0.1% onozuka R10 (all Sigma, http:// www.sigmaaldrich.com/). Epidermal tobacco protoplasts expressing the translational fusions (*2x35S: GFP-GmFWL1cDNA*, and *2x35S: GFP*) were observed using a Leica SP2 confocal microscope 3 days after leaf infiltration (488 nm excitation wavelength and 500-550 nm emission filter). The oil objective used was 63x/1.4NA.

Soybean hairy root transformation was performed as described by Libault et al. (2009) (Libault et al., 2009c) including the following changes. Soybean seeds were surface-sterilized according to Wan et al. (2005) (Wan et al., 2005a) then sowed on pro-mix. Ten days after transformation, the transgenic plants were transferred into the ultrasound aeroponic system (Qiao and Libault, 2013a). When required, 3 weeks old transgenic soybean root system were treated with nitrogen-free B & D solution (Broughton and Dilworth, 1971a) or by a suspension of *B. japonicum* (OD_{600}=0.1) (Qiao and Libault, 2013a). Transgenic *pCvMV: GmFWL1-GFP, pCvMV: GFP-*

GmFWL1, *pCvMV: GFP*, *pCvMV:mCherry-GmFLOT2/4* roots were observed at 1day and 7 days after inoculation using an Olympus FluoView 500 laser scanning confocal microscope with an argon 488 nm laser line and a helium-neon 543 nm laser line, and a with 60x/1.00NA water immersion objective.

TEM: Soybean hairy root transformation was performed as described by Libault et al. (2009) (Libault et al., 2009c) using *A. rhizogenes* strains to carry three different binary Gateway compatible vectors: *AKK1467B-Gateway-pFWL1:cFWL1-HA*, *AKK1467B-Gateway-pFWL1: HA-cFWL1* and *AKK1467B-Gateway-pFWL1: HA* plasmids. Three weeks and three days after transformation, the composite soybean plants expressing the fusion proteins HA tag- GmFWL1 and GmFWL1-HAtag and the HA tag alone were inoculated with *B. japonicum*. Thirty-two days after inoculation, the transgenic roots expressing the GFP were selected under an epifluorescent stereomicroscope and the nodules attached to these roots were isolated before tissue preparation for TEM. Three independent replicates were generated.

For each transgene and each biological replicate, 80 μm thick sections of 3 independent transgenic root nodules *pFWL1: HA-cFWL1*, *pFWL1:cFWL1-HA* and *pFWL1: HA*, were frozen by propane jet according to (Ding et al., 1991), substituted with pure acetone at -85°C for 2 days and brought to -20°, followed by a change of pure acetone at 4 °C overnight. The samples were substituted with ethanol on ice in 50, 70, 80, 100(3X)% ethanol and pure acetone for 15min each step. Then the samples were infiltrated through a graded series (30, 60, 100%) of medium grade LR White resin (Electron Microscopy Sciences, https://www.emsdiasum.com) and absolute ethanol. At

each step samples were infiltrated under vacuum for 1 hour followed by an additional 1 hour at 4°C. After a final 100% LR White exchange overnight, the samples were dispensed into gelatin capsules filled with fresh LR White resin and polymerized at 50°C for 48 hours.

Immunogold labeling was performed on 70 nm sections picked up on nickel grids. A total of 15 to 18 sections were analyzed for each transgene and each biological replicate. Grids with sections were blocked on PBS, 0.02% (v/v) Tween 20 and 1% (v/v) fish gelatin (FG) for 15min. Samples were labelled with anti-HA Epitope Tag (EMD-MILLIPORE) diluted 1:50 in (PBS-Tween20-FG) overnight at 4 °C. After 3 washes in PBS, the labeled samples were incubated for 2hours in secondary antibody-conjugated 15nm gold (Protein A Colloidal Gold, EY Laboratories, Inc.) diluted 1:50 in (PBS-Tween20-FG). The sections were rinsed with PBS (3X) and deionized water (2X) and subsequently observed on a Zeiss-10 (Carl Zeiss, Oberkochen, Germany) transmission electron microscope (Indrasumunar et al.) operated at 80kV. Samples for quantification of gold particles included tissues from the various biological replicates, transgenic nodules and ultrathin sections.

Figures and Tables

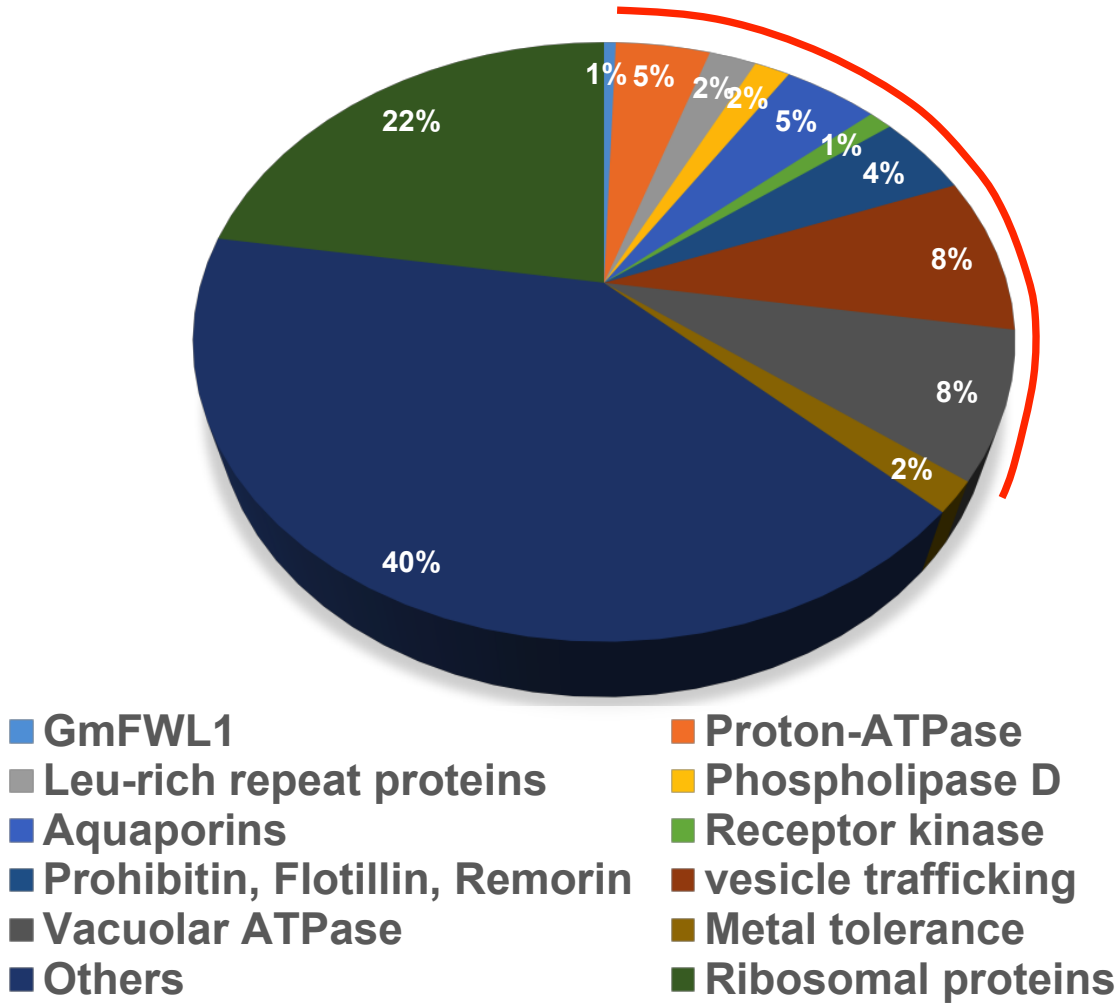


Figure 5- 1 Distribution of soybean proteins interacting with GmFWL1 in nodules according to their gene ontology. Protein described as membrane microdomain-associated proteins are highlighted with a red line.

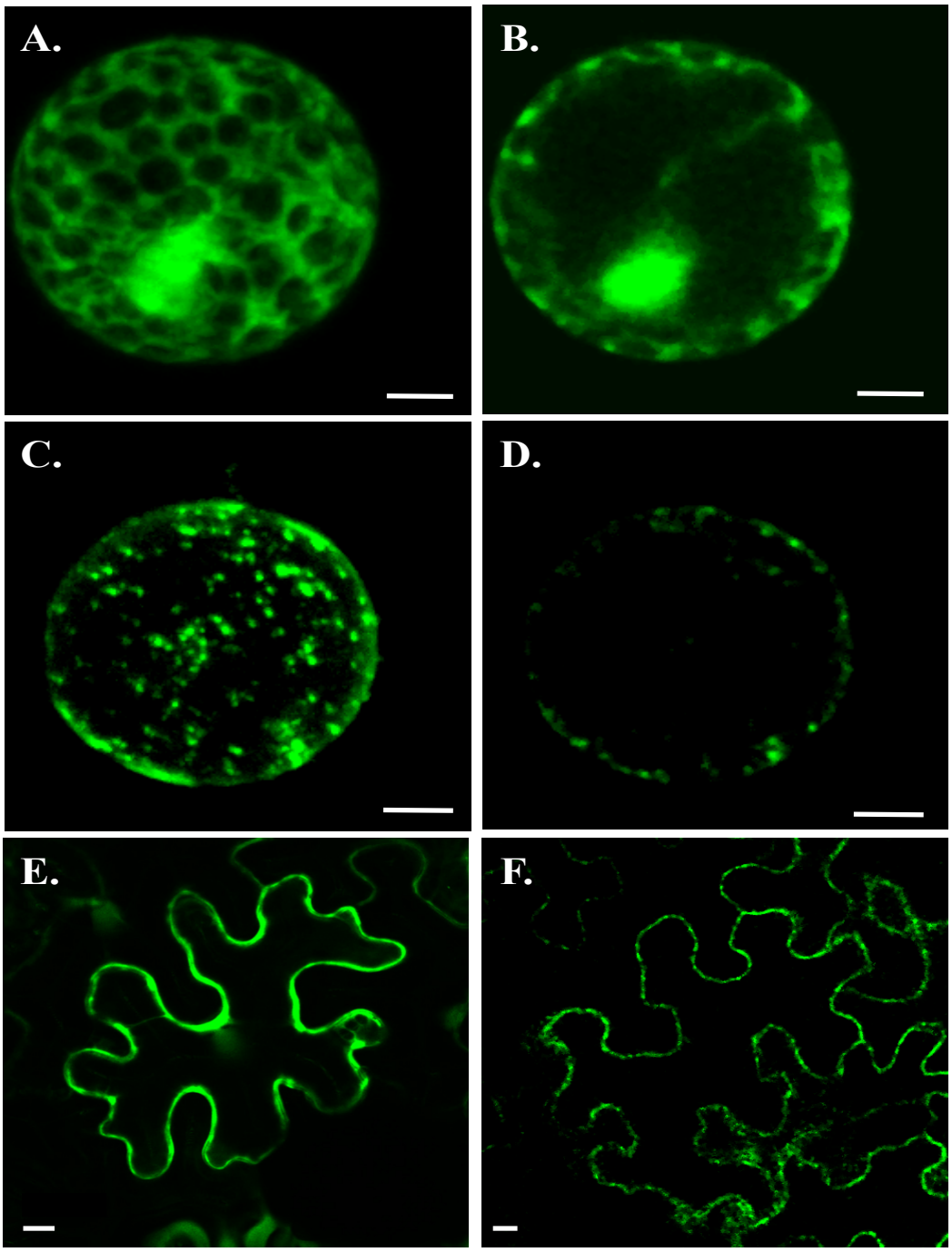


Figure 5- 2 Subcellular localization of GmFWL1 in tobacco leaf protoplasts (A-D) and in tobacco leaf cells (E-F). Maximum intensity projections (A and C) and optical sections (B, D, E and F) of tobacco protoplasts and tobacco leaf cells transformed with a 2Xp35S:GFP (A, B, E) and the 2Xp35S:GFP-cFWL1 (C, D, F) revealed the cytosolic and nuclear localization of the GFP (A, B, and E) and the punctate localization of the GFP-GmFWL1 fusion in the plasma membrane of the tobacco leaf protoplast and cell (C, D and F). Bar: 10 μm

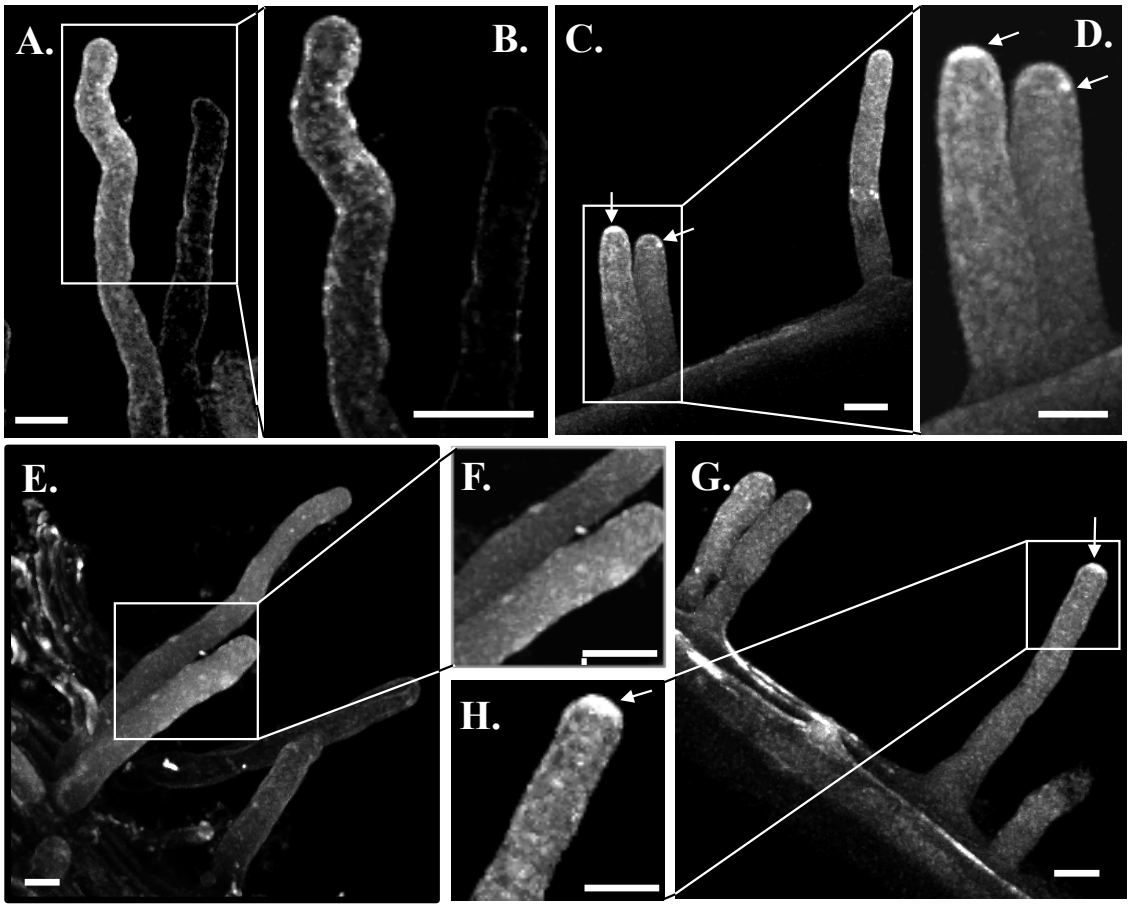


Figure 5- 3 Subcellular localization of the GFP-GmFWL1 chimeric protein in soybean root hair cell in response to *B. japonicum* inoculation. Soybean root hair cells were mock-inoculated (A, B, E, and F) or inoculated with *B. japonicum* (C, D, G and H). One day (A-D) and seven days (Lee et al.) after inoculation, the transgenic root hair cells were observed under an epifluorescent confocal microscope. Compared to mock-inoculated conditions, the translocation of the GmFWL1 protein at the tip of the root hair cells upon *B. japonicum* inoculation was observed as soon as 24 hours after inoculation and continues 7 days after inoculation.
White arrows highlight the accumulation of the GFP-GmFWL1 protein at the tip of the root hair cells upon *B. japonicum* inoculation. Supplemental Figure 5-6 shows the quantification of the GFP signal in inoculated and mock-inoculated transgenic root hair cells. Bar: 10 μm

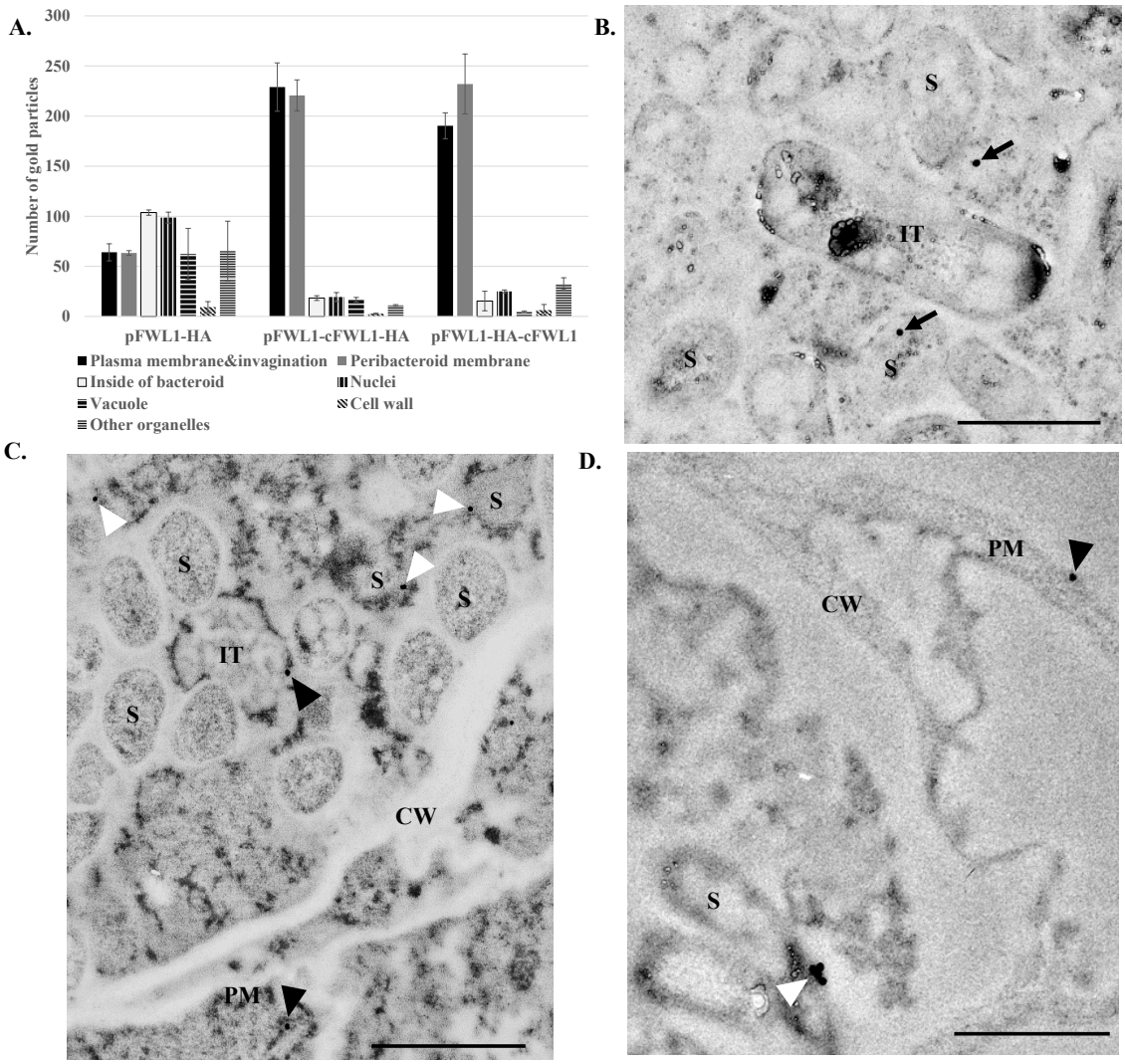


Figure 5- 4 The punctuate plasma membrane localization of GmFWL1 was revealed in nodules using transmission electron micrographs after immunogold labeling.
A. Distribution of the number of gold particles across different plant cell compartments (i.e. plasma membrane, peribacteroid membrane, cell wall, vacuole, nucleus, etc.) after immunogold labelling against the HA tag alone, the GmFWL1-HA tag and the HA tag-GmFWL1 chimeric proteins. While the HA tag alone was randomly distributed in the cell (B), the HA tag-GmFWL1 (C) and the GmFWL1-HA tag (D) chimeric proteins were strongly detected in the plasma and the peribacteroid membranes.
Black and white arrow heads (C and D) point at the gold particles located in the plasma membrane, the infection threads, and the peribacteroid membrane. Black arrows (N) highlight those randomly distributed.
Legend: CW: cell wall; PM: plasma membrane; IT: infection thread; S: symbiosome; Scale bar: 1 μm

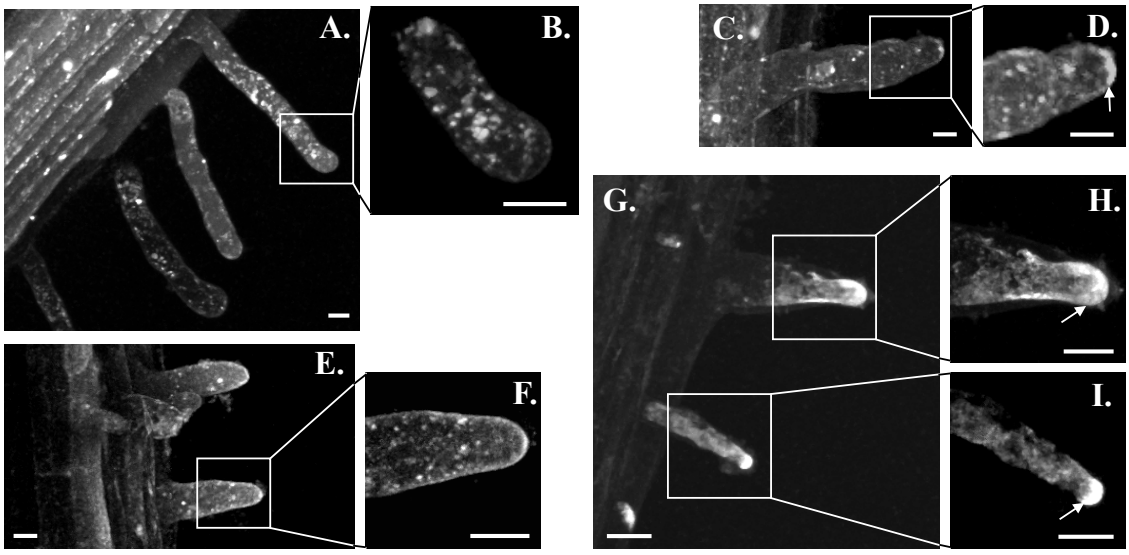


Figure 5- 5 Subcellular localization of mCherry-GmFLOT2/4 fusion protein in soybean root hair cell in response to *B. japonicum* inoculation. Soybean root hair cells were mock-inoculated (A, B, E, and F) or inoculated with *B. japonicum* (C, D, G, H and I). One day (A-D) and seven days (E-I) after inoculation, the transgenic root hair cells were observed under an epifluorescent confocal microscope. Compared to mock-inoculated conditions and similarly to GmFWL1 (Figure 5-3), GmFLOT2/4 accumulation at the tip of the root hair cells in response to *B. japonicum* inoculation was observed as soon as 24 hours after inoculation and continued 7 days after inoculation. White arrows highlight the accumulation of the mCherry-GmFLOT2/4 protein at the tip of the root hair cells upon *B. japonicum* inoculation. Supplemental Figure 5-6 shows the quantification of the mCherry signal in inoculated and mock-inoculated transgenic root hair cells.
Bar: 10 μm

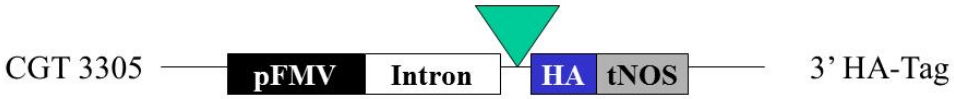
Supplemental Video 5-1 3-dimensional confocal imaging of a tobacco leaf protoplast expressing the GFP-GmFWL1 fusion protein (Young et al.). The auto-fluorescence of the chloroplasts (red) has been recorded to provide information about the shape of the protoplast. As supported by additional observations (Figure 4), GmFWL1 is a plasma membrane protein. In this tobacco protoplast, GmFWL1 shows a punctate localization at the periphery of the cell supporting its function as a membrane microdomain-associated protein.

Supplemental Video 5-2 3-dimensional confocal imaging of a tobacco leaf cell co-expressing the GFP-GmFWL1 (Young et al.) and mCherry-GmFLOT2/4 fusion protein (red). Both proteins are located at the periphery of the cell accordingly to their plasma membrane localization. In addition, the two proteins are showing a similar punctate localization supporting their membrane microdomain localization. Only a small fraction of the red and green signal overlap suggesting that membrane microdomains differ in their protein composition.

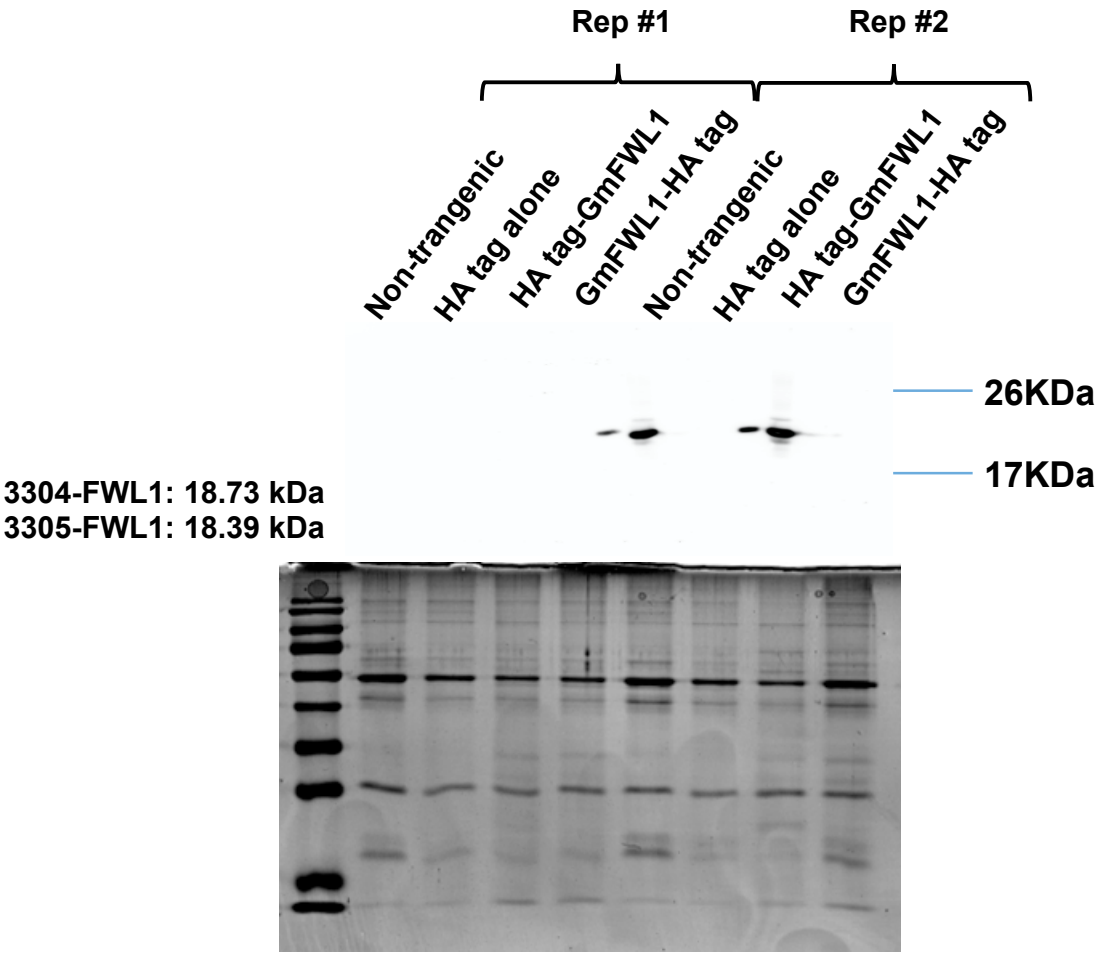
Supplemental Video 5-1 and Supplemental Video 5-2 are not included in this dissertation duo to their large sizes, which can be accessed by the link:
http://onlinelibrary.wiley.com/doi/10.1111/pce.12941/full

A.

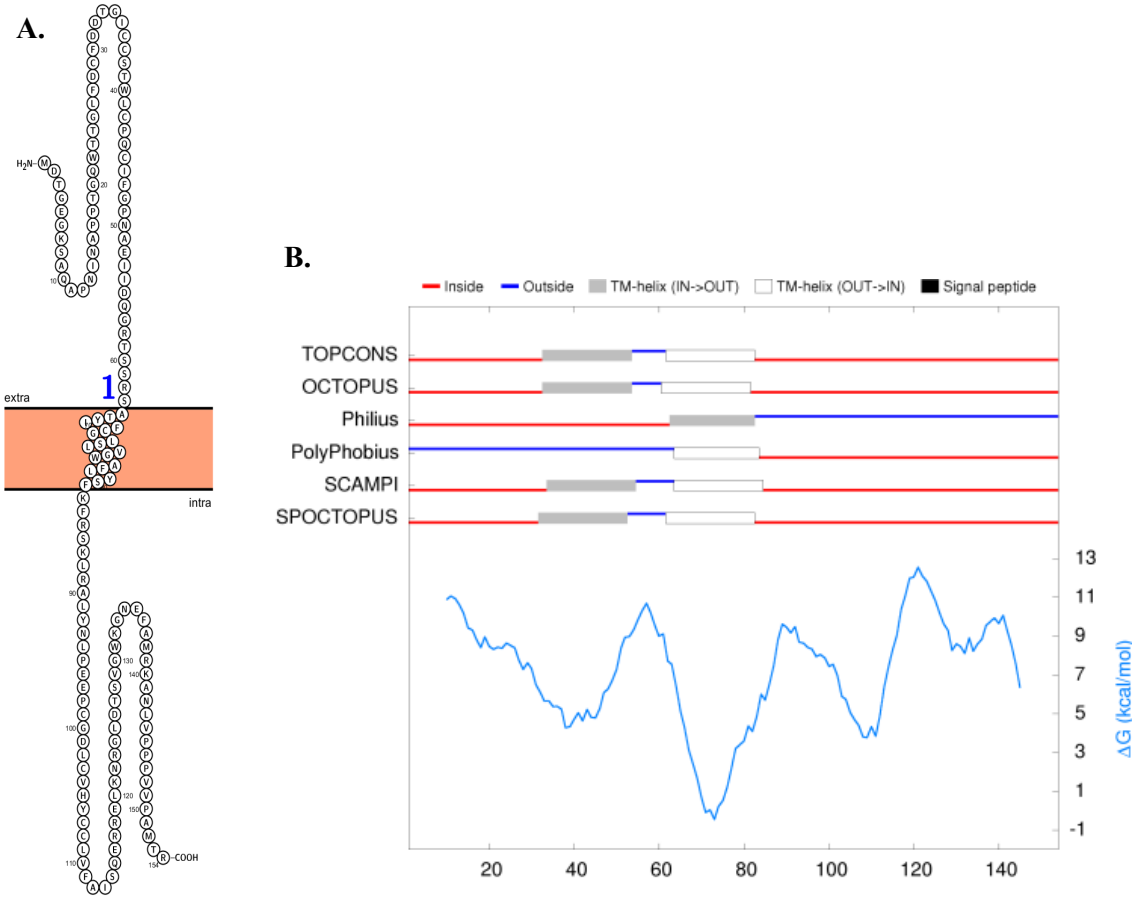
B.



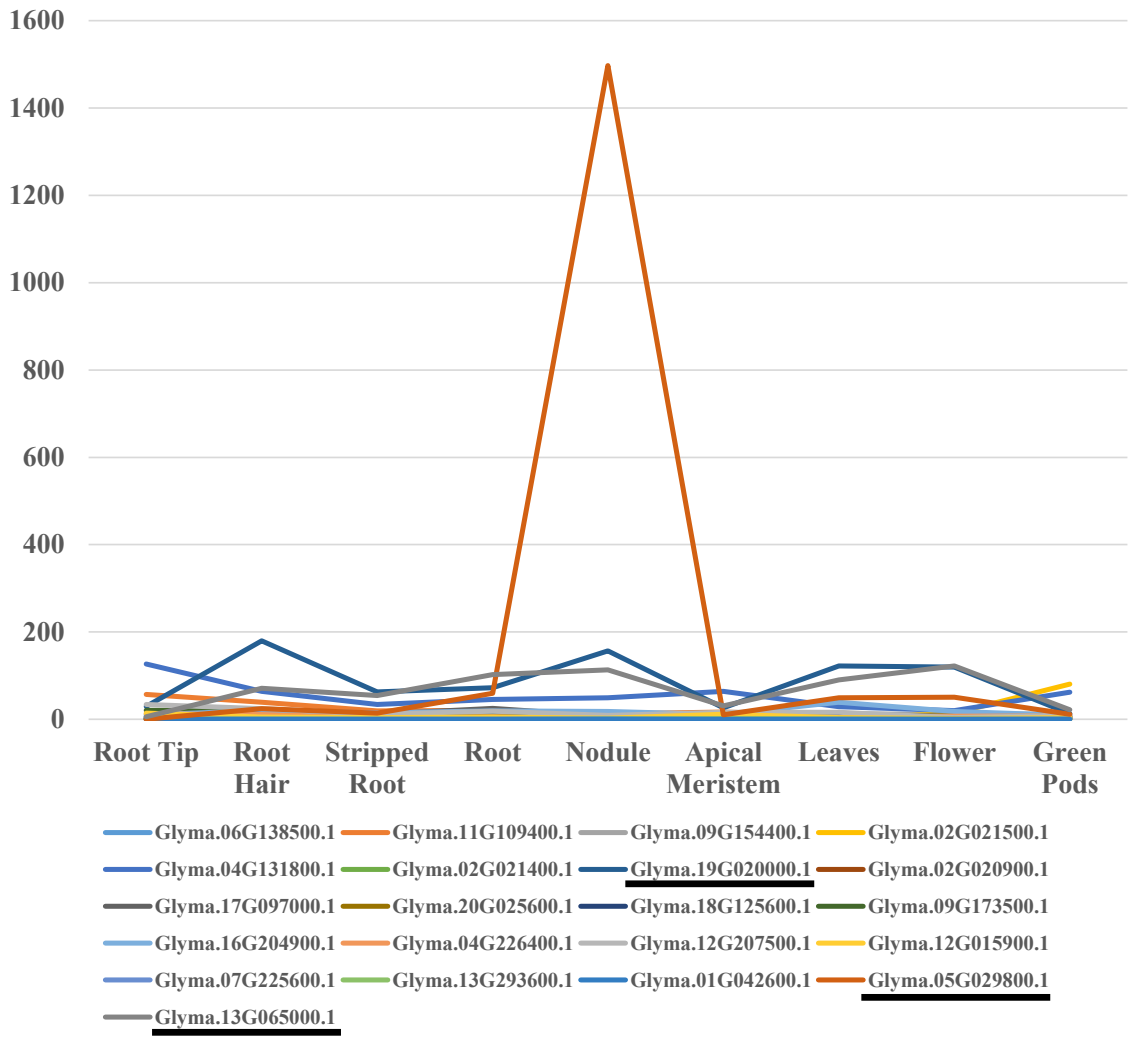
Supplemental Figure 5- 1 Construction of the CGT3304 (A.) and CGT3305 (B.) vectors. Two primer-based adaptors containing the hemagglutinin (HA)-epitope tag were cloned into the *FMV: Intron:uidA:tNOS* shuttle vector (CGT 3302; Govindarajulu et al., 2008) to create a 5' HA-tag (CGT 3304) and 3' HA-tag (CGT 3305) shuttle plasmid. HA-epitope adaptor (Adapt 48; 5' ccccatggaaTACCCGTACGACGTTCCGGACTACGCTctagatctgatatcgagctccc-3') was cloned into CGT 3302 using restriction sites *NcoI* and *EcoICRI* to create the 5' HA-tag (capitol letters) shuttle plasmid, CGT 3304. In this plasmid, the ATG translational start site (in bold) begins at the NcoI site and leads into the HA-tag followed by unique restriction enzyme sites *XbaI*, *BglII*, *EcoRV* and *EcoICRI* for cloning your favorite gene. The 3' HA-epitope shuttle plasmid (CGT 3305) was created by cloning a 3' HA-epitope adaptor (Adapt 49; 5'-cccccatggattctagatctgagctcTACCCGTACGACGTTCCGGACTACGCTtaagatatcccc-3') into the shuttle vector CGT 3302 using restriction enzymes *NcoI* and *EcoRV* for adaptor digestion and *NcoI* and *EcoICRI* for digestion of the shuttle. In this plasmid, your favorite gene is cloned 5' of the HA-tag using unique restriction enzyme sites, *NcoI*, *XbaI* and/or *EcoICRI*. A stop codon is included after the HA-tag sequence (in bold).



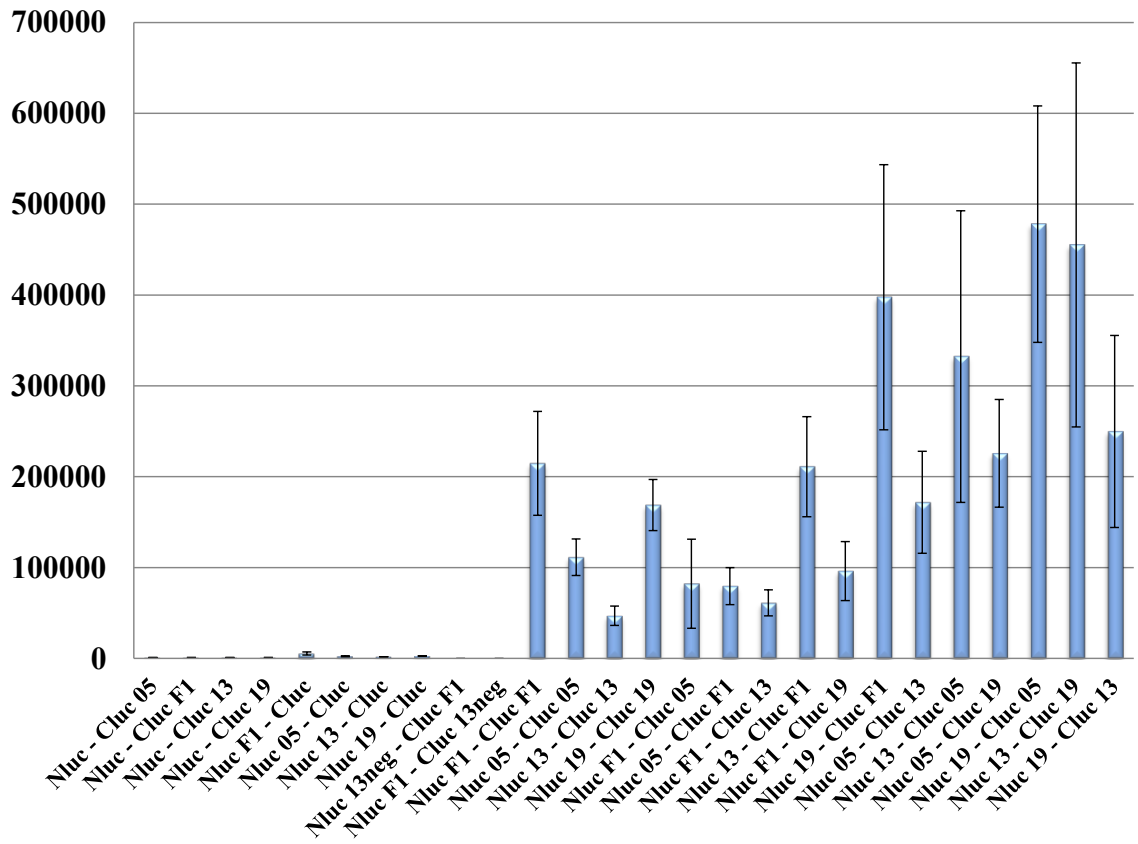
Supplemental Figure 5- 2 Western blot analysis of the GmFWL1 co-immunoprecipitations. The expression of the GmFWL1-HA tag and HA tag-GmFWL1 fusion proteins was validated by western blot (two independent biological replicates) using an anti-HA tag primary antibody followed by a horseradish peroxidase-conjugated secondary antibody. The lower intensity of the horseradish peroxidase activity detecting the HA tag-GmFWL protein might reflect a lower stability of the fusion protein. Two negative controls are included: HA-tag-only nodules and wild-type nodules. The absence of detection of the HA tag alone (1.1 KDa) is likely the consequence of its rapid electrophoresis through the gel. As a control, a Coomassie Blue staining protein gel shows equal amounts of protein loaded in every lane.



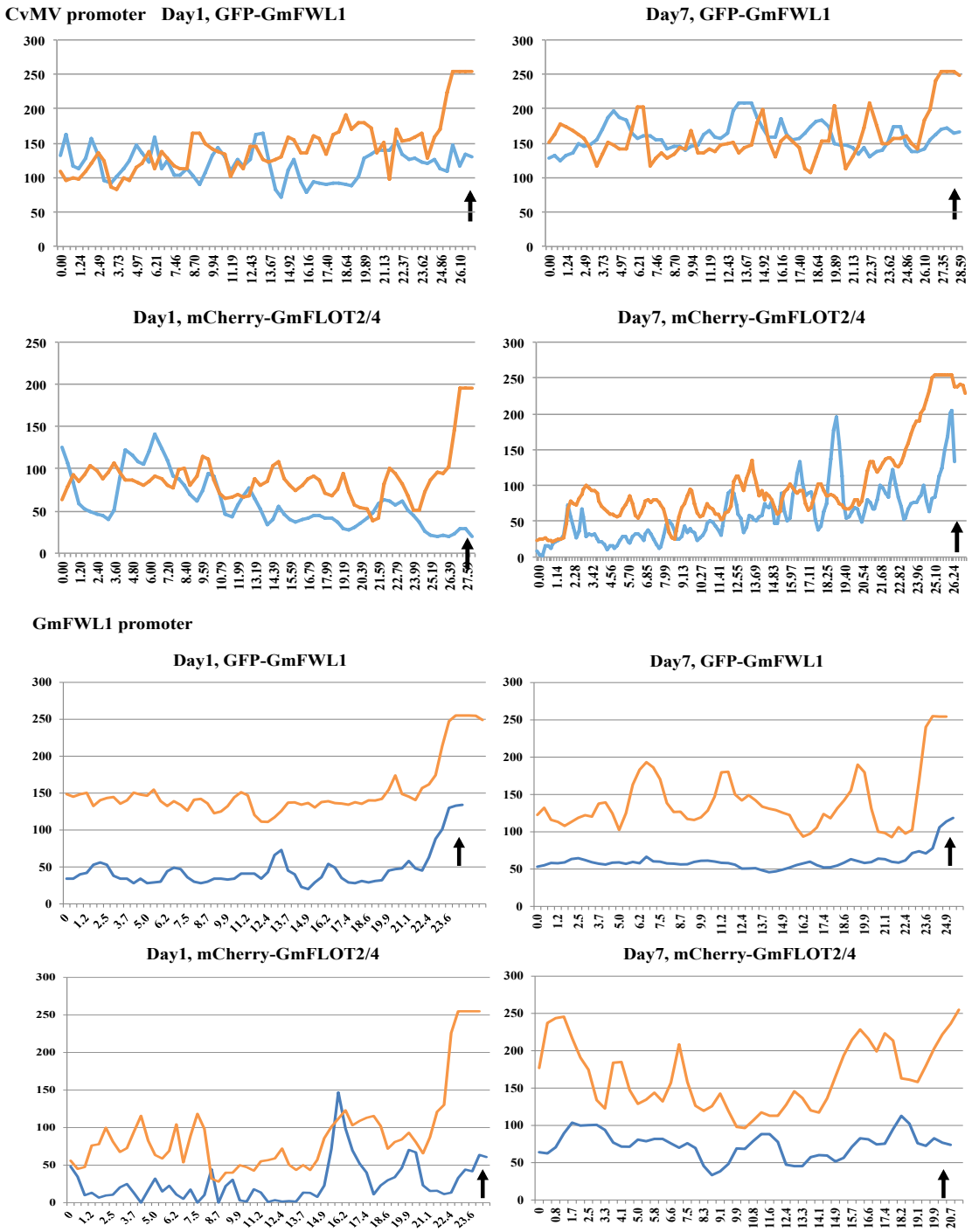
Supplemental Figure 5- 3 A. Schematic representation of the organization of the GmFWL1 protein across the plasma membrane predicted by Protter (Omasits et al., 2014; http://wlab.ethz.ch/protter/start/). One transmembrane domain is predicted. As a note, the N-terminal and C-terminal domains of GmFWL1 are predicted to be extracellular and intracellular, respectively. B. Consensus prediction of membrane topology of GmFWL1 by TOPCONS (Bernsel et al., 2009; http://topcons.cbr.su.se/). Depending prediction tools, one to two transmembrane domains are predicted.



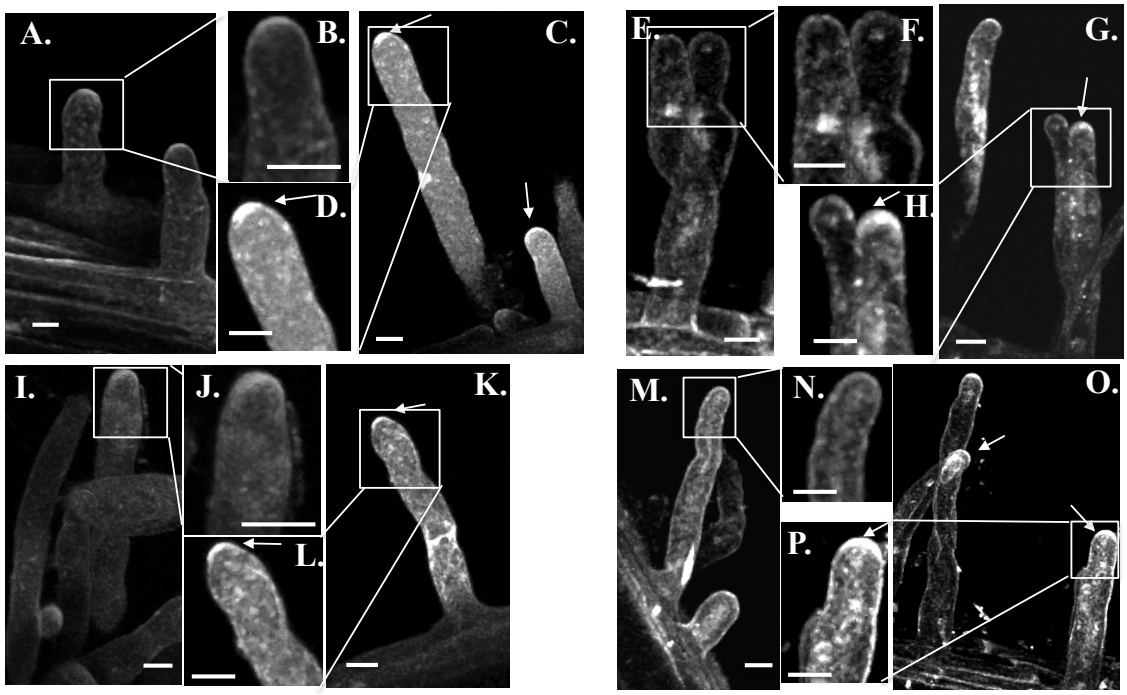
Supplemental Figure 5- 4 Normalized level of expression of the 21 soybean prohibitins genes across 9 different soybean cell type and organs. “Stripped roots” refers to roots devoid in root hair cells. The data sets used to generate this figures were mined from the soybean transcriptome atlas (Libault et al., 2010c). The three prohibitin genes the most expressed in nodules (underlined genes) are encoding proteins interacting with GmFWL1.



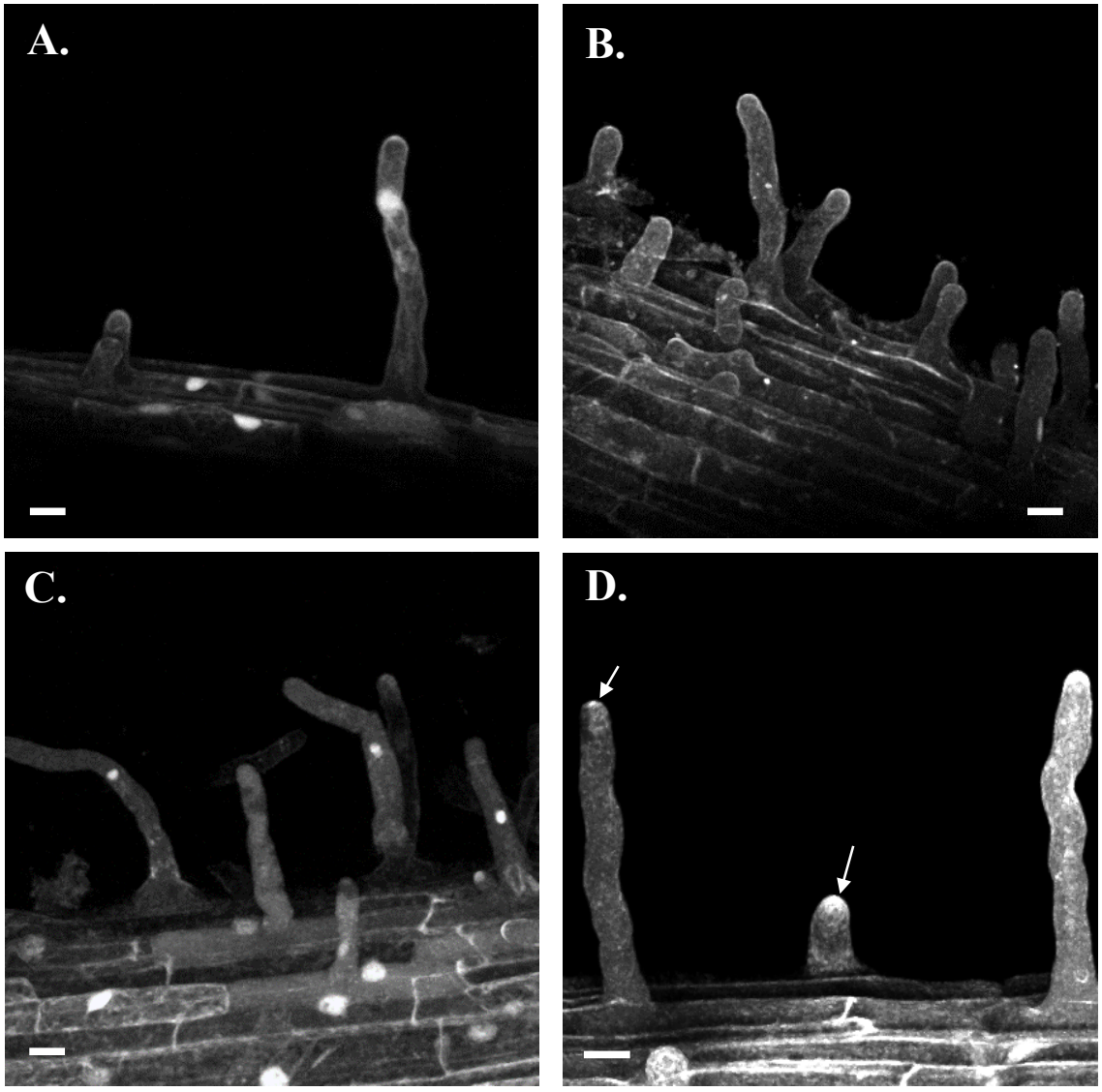
Supplemental Figure 5- 5 Split luciferase assays reveal the homo- and heterodimerization of GmFWL1 (F1) and prohibitins (05: Glyma.05G029800; 13: Glyma.13G065000; 19: Glyma.19G020000). Mean luciferase activity (y-axis) recorded from five independent biological replicates after co-expression of fusion proteins with the Nter and Cter domains of the luciferase (Nluc; Cluc, respectively) in tobacco leaves. Various homo- and hetero-dimerization were tested as well as a series of negative controls including the prohibitin protein Glyma.13G293600 (13neg), which is not characterized as a GmFWL1 interactor according to our co-immunoprecipitation assay (Supplemental Table 5-2).



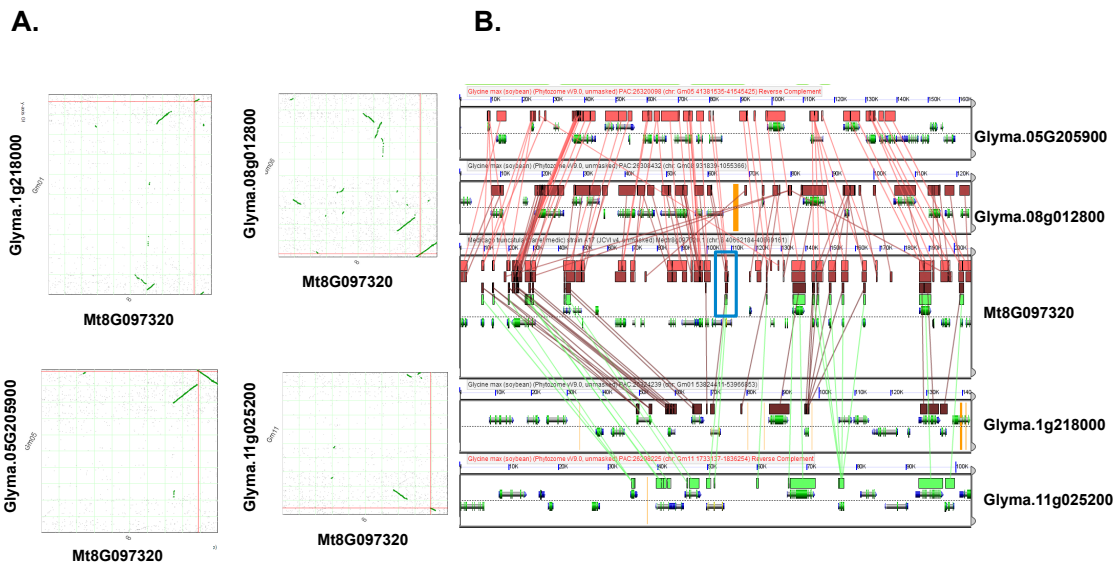
Supplemental Figure 5- 6 Quantification of the intensity of the fluorescence of the GFP-GmFWL1 and mCherry-GmFLOT2/4 proteins (y-axis) across the soybean root hair cells (x-axis, µm) in mock-inoculated (blue) and *B. japonicum* inoculated (orange) conditions. The fluorescent proteins are expressed under the control of the CvMV or the Gm*FWL1* promoters (see Figures 5-3 and 5, Supplemental Figure 5-7) (the arrows highlight the position of the root hair tip).



Supplemental Figure 5- 7 Subcellular localization of GFP-FWL1 (A-D, I-L) and mCherry-GmFLOT2/4 (E-H, M-P) fusion proteins driven by the Gm*FWL1* promoter in soybean root hair cell in response to *B. japonicum* inoculation. Soybean root hair cells were mock-inoculated (A, B, E, F, I, J, M, and N) or inoculated with *B. japonicum* (C, D, G, H, K, L O and P). One day (A-H) and 7 days (I-O) after inoculation, the transgenic root hair cells were observed under an epifluorescent confocal microscope. We confirmed the accumulation at the tip of the root hair cells of GmFWL1 and GmFLOT2/4 in response to *B. japonicum* inoculation as soon as a day after inoculation. White arrows highlight the accumulation of the GFP-GmFWL1 and mCherry-GmFLOT2/4 proteins at the tip of the root hair cells upon *B. japonicum* inoculation. Bar: 10 μm

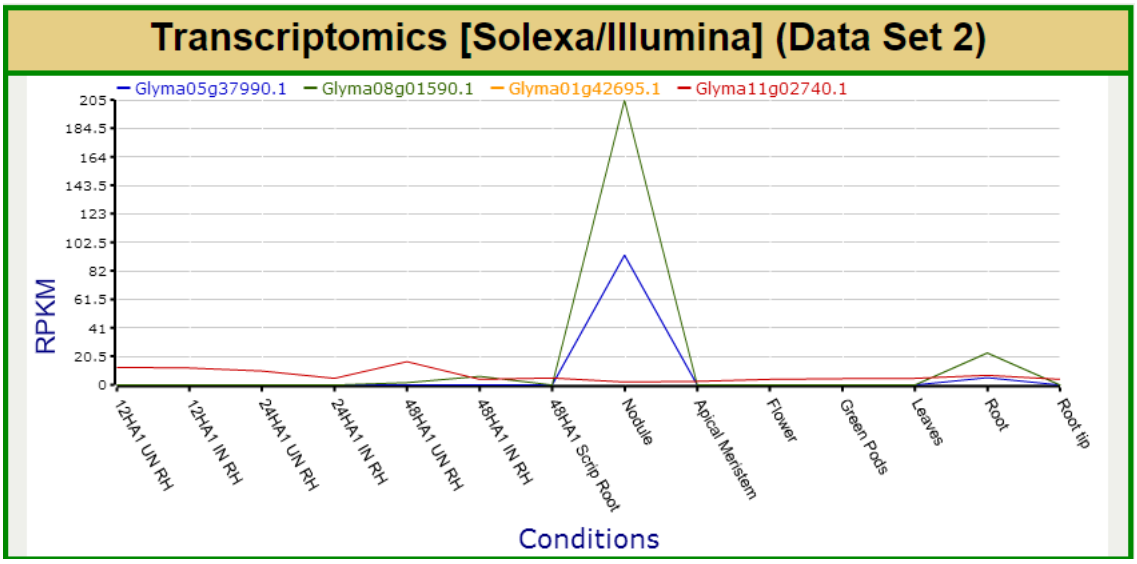


Supplemental Figure 5- 8 Subcellular localization of pCvMV-GFP (A and C) and pCvMV-GmFWL1-GFP (B and D) in soybean root hair cell in response to *B. japonicum* inoculation. Soybean root hair cells were mock-inoculated (A and B) or inoculated with *B. japonicum* (C and D). Seven days after inoculation, the transgenic root hair cells were observed under an epifluorescent confocal microscope.
White arrows highlight the accumulation of the GmFWL1 protein at the tip of the root hair cells upon *B. japonicum* inoculation. Bar: 10 μm

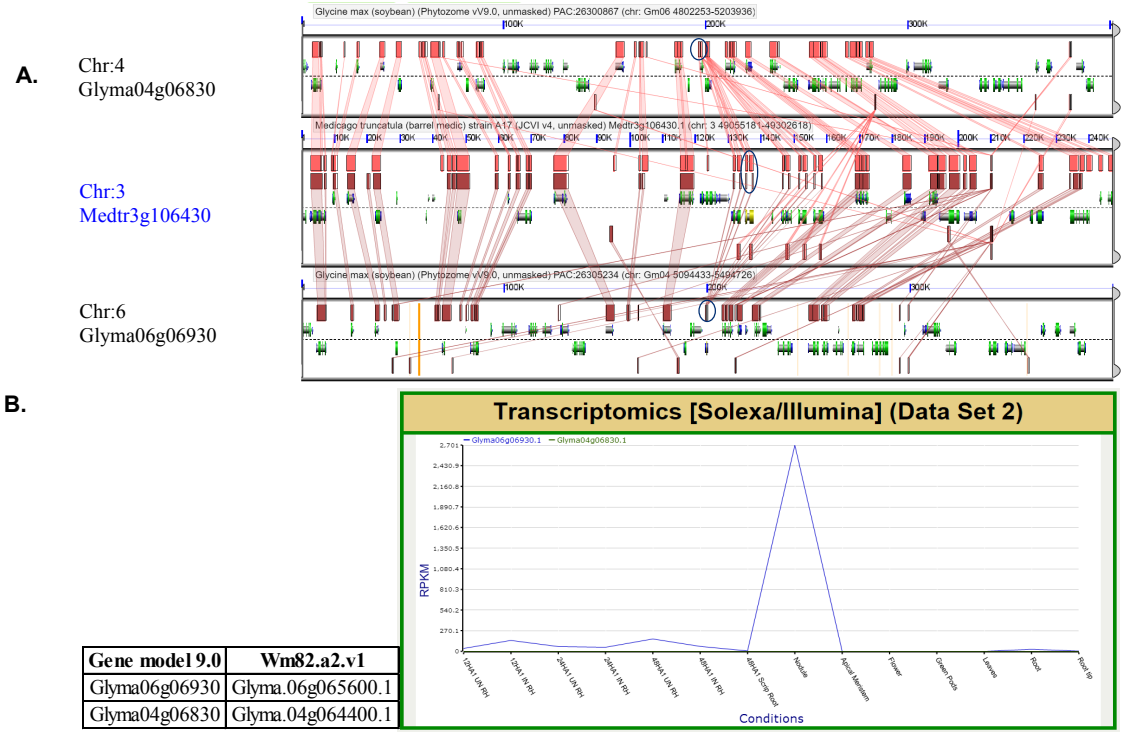


Supplemental Figure 5- 9 Macro (A.) and microsyntenic (B.) relationships between MtSYMREM1 (i.e., Mt8g097320) and 4 soybean orthologous genes located on the chromosome 01, 05, 08 and 11. A. Macrosyntenic relationships are highlighted in green between soybean and medicago chromosomes. B. In addition, soybean and medicago genes showing microsyntenic relationships are linked together. These analyses were performed using CoGe resources.

Gene model V9.0	Wm82.a2.v1
Glyma05g37990	Glyma.05G205900
Glyma08g01590	Glyma.08G012800
Glyma01g42695	Glyma.01G218000
Glyma11g02740	Glyma.11G025200

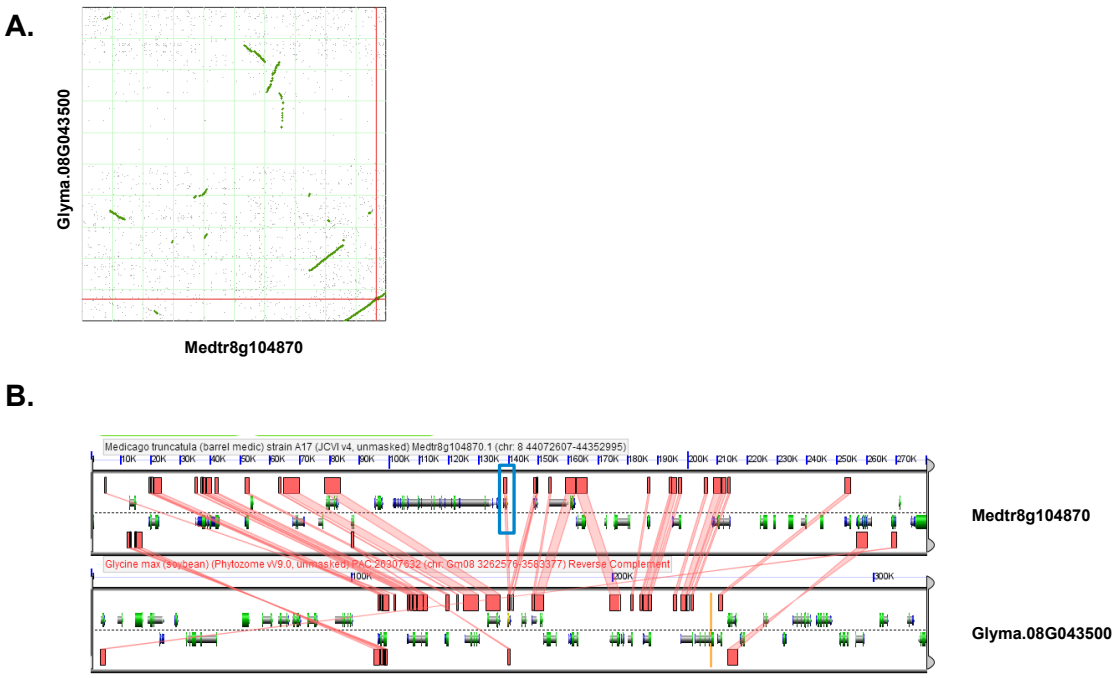


Supplemental Figure 5- 10 Expression patterns of 4 soybean genes encoding remorin proteins orthologous to Mt*SYMREM1*. Among these 4 genes, Glyma.05G205900, Glyma.08g012800 are highly expressed in nodules. This result suggests a specific protein composition of plasma membrane microdomains during the nodulation process.Transcriptomic data sets were mined from the SoyKB website.



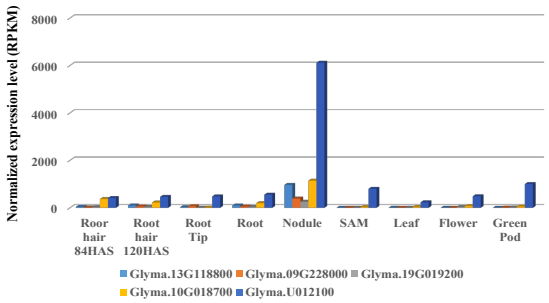
Gene model 9.0	Wm82.a2.v1
Glyma06g06930	Glyma.06g065600.1
Glyma04g06830	Glyma.04g064400.1

Supplemental Figure 5- 11 Characterization of Gm*FLOT2/4* genes based on syntenic relationships with MtFLOT4 (i.e., Mt3g106430) and their nodule-specific expression patterns. Using CoGe resources (Lyons and Freeling,2008), microsyntenic (A.) relationships with MtFLOT4 (Medtr3g106430) revealed 2 soybean orthologous genes (Glyma.04g064400 and Glyma.06G065600). Each panel is a visualization of chromosome region showing the gene models on positive and negative strands. The linked red and deep red blocks in the panel highlight microsyntenic regions between soybean and medicago based on gene function and orientation. The FLOT2/4 orthologs based on their microsyntenic relationships are highlighted with blue circles.
To highlight the conservation of the function of these orthologous genes, the soybean transcriptome atlas was mined from using the SoyKB website (Joshi et al., 2012; Joshi et al., 2014). Among the 2 soybean *FLOT2/4* genes, Glyma.06G065600 is highly expressed in nodules (B.). Oppositely, based on this transcriptomic analysis, Glyma.04g064400 could be considered as a pseudogene. The highly preferential expression of Glyma.06G065600 in nodules supports the conservation of the function between MtFLOT4 and Glyma.06G065600.

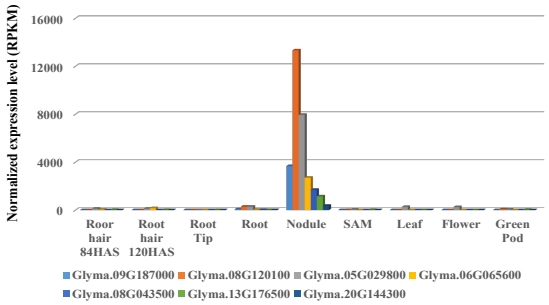


Supplemental Figure 5- 12 Macro (A.) and microsyntenic (B.) relationships between Medtr8g104870 and Gm*FWL3* (i.e., Glyma.08G043500). A. Macrosyntenic relationships are highlighted in green between soybean and medicago chromosomes. B. In addition, soybean and medicago genes showing microsyntenic relationships are linked together. These analyses were performed using CoGe resources.

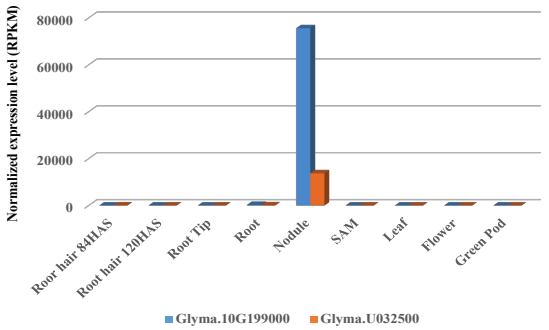
A.



B.



C.



D.

Soybean gene ID		Putative Function
Wm82.a2.v1	Phytozome V1.1	
Glyma.10G199000	Glyma10g34280	haemoglobin 2
Glyma.U032500	Glyma10g23790	uricase / urate oxidase / nodulin 35, putative
Glyma.09G187000	Glyma09g31910	PLAC8 family protein
Glyma.08G120100	Glyma08g12650	NOD26-like intrinsic protein 1;2
Glyma.05G029800	Glyma05g01360	SPFH/Band 7/PHB domain-containing membrane-associated protein family
Glyma.06G065600	Glyma06g06930	SPFH/Band 7/PHB domain-containing membrane-associated protein family
Glyma.08G043500	Glyma08g04830	PLANT CADMIUM RESISTANCE 2
Glyma.13G176500	Glyma13g24510	HSP20-like chaperones superfamily protein
Glyma.20G144300	Glyma20g28230	N/A
Glyma.13G118800	Glyma13g17910	PGP11 P-glycoprotein 11
Glyma.09G228000	Glyma09g36220	Protein of unknown function, DUF642
Glyma.19G019200	Glyma19g02270	H(+)-ATPase 11
Glyma.10G018700	Glyma10g02340	Adenine nucleotide alpha hydrolases-like
Glyma.U012100	Glyma11g17930	ATJ2,J2 DNAJ homologue 2

Supplemental Figure 5- 13 Expression levels of soybean genes encoding GmFWL1 protein partners and preferentially (≥3- and <10-fold change between the expression levels in nodules compared to the second most highly expressed genes; A. specifically (≥10- and <100-fold change; B. and very specifically (≥100- and <1000-fold change; C. expressed in nodules compared to other soybean cell types and organs [i.e. root hair cells collected 84 and 120 hours after seed sowing, root tip, root, Nodule, Shoot apical meristem (SAM), leaf, flower and green pods]. D. The putative function of the proteins encoded by these genes is indicted. Gene expression data sets were mined from the soybean transcriptome atlas (Libault et al., 2010c).

Supplemental Table 5- 1 List of the designed primers.

Co-immunoprecipitation

Primer	Sequence
3304/3305GmFWL1 Forward	cccggatccatggataccggtgaaggtaag
3304GmFWL1 Reverse	cccgagctcctagcgagtcatggcgggc
3305GmFWL1 Reverse	cccgagctcgcgagtcatggcgggcac

Split luciferase

Primer	Sequence
N/ClucGmFWL1 Forward	cccggtaccatggataccggtgaaggtaagt
NlucGmFWL1 Reverse	cccgtcgacgcgagtcatggcgggcaca
ClucGmFWL1 Reverse	cccgtcgacctagcgagtcatggcgggcaca
N/ClucGlyma05g01360 Forward	cccggtaccatggggaatctttttgttgtgtg
NlucGlyma05g01360 Reverse	cccgtcgacctgatgagaagcctgaagaagt
ClucGlyma05g01360 Reverse	cccgtcgacctactgatgagaagcctgaagaagt
N/ClucGlyma13g05120 Forward	cccggatcccatgggtcaagcactaggttgc
NlucGlyma13g05120 Reverse	cccgtcgacattctgtgaagtggtggcctga
ClucGlyma13g05120 Reverse	cccgtcgacttaattctgtgaagtggtggcctga
N/ClucGlyma19g02370 Forward	cccggatcccatgggtcaagcattcggttgct
NlucGlyma19g02370 Reverse	cccgtcgacattctgtgaggcagtggcctg
ClucGlyma19g02370 Reverse	cccgtcgacttaattctgtgaggcagtggcctg

Epifluorescent microscopy assay

Primer	Sequence
GmFWL1 cDNA AttB1-B2 Forward	aaaaagcaggctatatggataccggtgaaggtaagt
GmFWL1 cDNA AttB1-B2 Reverse	agaaagctgggttgcgagtcatggcgggcaca
GmFWL1 cDNA AttB4R-B3R Forward	aaagttgccatggataccggtgaaggtaagtca
GmFWL1 cDNA AttB4R-B3R Reverse	atacaaagttgtgcgagtcatggcgggcaca
GmFWL1 cDNA AttB3-B2 Forward	aataaagttgcggataccggtgaaggtaagt
GmFWL1 cDNA AttB3-B2 Reverse	agaaagctgggttgcgagtcatggcgggcaca
GFP AttB4RAttB3R Forward	tacaaagttgccatgagtaaaggagaagaactt
GFP AttB4RAttB3R Reverse	atacaaagttgtgtggtggtggtggtgtt
GFP AttB3AttB2 Forward	aataaagttgcgagtaaaggagaagaacttttc
GFP AttB3AttB2 Reverse	gaaagctgggtattagtggtggtggtggt
mCherry AttB4R-B3R Forward	tacaaagttgccatggtgagcaagggcgaggag
mCherry AttB3-B2 Forward	aataaagttgcggtgagcaagggcgaggaggat
mCherry AttB4R-B3R Reverse	atacaaagttgtcttgtacagctcgtccatgccg
mCherry AttB3-B2 Reverse	gaaagctgggtactacttgtacagctcgtccatgccg
GmFlot2/4 AttB3-B2 Forward	aataaagttgcgatgtacaaggtagcaaacgcc
GmFlot2/4 AttB3-B2 Reverse	gaaagctgggtatcaagaaccatt atcgggcag

GmFlot2/4 AttB4r-B3r Forward	tacaaagttgcgatgtacaaggtagcaaacgcc
GmFlot2/4 AttB4r-B3r Reverse	atacaaagttgtagaaccatt atcgggcag
GmFWL1promoter AttB1-B4 Forward	aaaaagcaggctttgtgtgcattaagttgtgagc
GmFWL1promoter AttB1-B4 Reverse	aaaagttgggtgtttgaagctctaatcagagacaaa
pSU AttB1-B4 Forward	aaaaaagcaggctaaaaaaacccctcacaaataca
pSU AttB1-B4 Reverse	aaaagttgggtgtgaaatatgactaacgaatatac
pCvMV AttB1-B4 Forward	aaaaagcaggctatccagaaggtaattatccaagatgt
pCvMV AttB1-B4 Reverse	aaaagttgggtgcaaacttacaaatttctctgaagtt

TEM

Primer	Sequence
HA AttB4R-B3R Forward	tacaaagttgccatgtacccgtacgacgt
HA AttB3-B2 Forward	aataaagttgcgtacccgtacgacgt
HA AttB5-B2 Forward	acaaaagttgcctacccgtacgacgt
HA AttB4R-B3R Reverse	atacaaagttgtagcgtagtccggaac
HA AttB3-B2 Reverse	gaaagctgggtactaagcgtagtccggaac
HA AttB5-B2 Reverse	gaaagctgggtactaagcgtagtccggaac

Gateway compatible primers

Primer	Sequence
AttB1 Forward	ggggacaagtttgtacaaaaaagcaggct
AttB4R Forward	ggggacaacttttctatacaaagttg
AttB3 Forward	ggggacaactttgtataataaagttg
AttB5 Forward	ggggacaactttgtatacaaaagttg
AttB4 Reverse	ggggacaactttgtatagaaaagttgggtg
AttB3R Reverse	ggggacaactttattatacaaagttgt
AttB2 Reverse	ggggaccactttgtacaagaaagctgggta

Creation of the pAKK1467B Gateway compatible

Primer	Sequence
CGT3304AttR1tNosfor	gggcctgcaggatcaaacaagtttgtacaaaaaagc
CGT3304AttR1tNosrev	gggcctgcaggagtaacatagatgacaccgcgc

Supplemental Table 5- 2 Characterization of 178 soybean proteins co-immunoprecipitated with GmFWL1. The proteins repetitively found interacting with GmFWL1 are highlighted in red.

Protein Family	Gene annotation		Predicted function	Protein mass	Number of predicted transmembrane domains
	Phytozome V1.1	Wm82.a2.v1			
FWL, FW2.2-like, CNR1	Glyma09g31910	Glyma.09G187000	GmFWL1, PLAC8 family protein	17 kDa	1
	Glyma08g04830	Glyma.08G043500	GmFWL3, PLANT CADMIUM RESISTANCE 2	19 kDa	0
Proton-ATPase	Glyma04g34370	Glyma.04G174800	H(+)-ATPase 11	105 kDa	8
	Glyma05g01460	Glyma.05G030700	H(+)-ATPase 11	105 kDa	8
	Glyma06g20200	Glyma.06G189900	H(+)-ATPase 11	105 kDa	8
	Glyma17g10420	Glyma.17G096100	H(+)-ATPase 11	105 kDa	8
	Glyma19g02270	Glyma.19G019200	H(+)-ATPase 11	97 kDa	8
	Glyma09g06250	Glyma.09G056300	H(+)-ATPase 2	105 kDa	10
	Glyma06g07990	Glyma.06G076100	H(+)-ATPase 5	105 kDa	10
	Glyma14g17360	Glyma.14G134100	H(+)-ATPase 5	105 kDa	10
Leu-rich repeat proteins	Glyma17g12880	Glyma.17G119800	Leucine-rich repeat protein kinase family protein	71 kDa	1
	Glyma06g31630	Glyma.06G226700	Leucine-rich repeat transmembrane protein kinase	89 kDa	1

Supplemental Table 5-2 cont.,

	Glyma13g34100	Glyma.13G266100	Leucine-rich repeat transmembrane protein kinase	110 kDa	2
	Glyma07g33950	Glyma.07G215900	Leucine-rich repeat transmembrane protein kinase	50 kDa	1
Phospholipase D	Glyma13g44170	Glyma.13G364900	phospholipase D alpha 1	93 kDa	0
	Glyma01g36680	Glyma.01G162100	phospholipase D delta	99 kDa	0
	Glyma11g08640	Glyma.11G081500	phospholipase D delta	98 kDa	0
Aquaporins	Glyma02g08110	Glyma.02G073600	PIP2;2,PIP2B plasma membrane intrinsic protein 2	30 kDa	6
	Glyma08g01860	Glyma.08G015300	plasma membrane intrinsic protein 1;4	31 kDa	6
	Glyma11g02530	Glyma.11G023200	plasma membrane intrinsic protein 1;4	31 kDa	6
	Glyma12g08040	Glyma.12G075400	plasma membrane intrinsic protein 2;4	31 kDa	6
	Glyma13g40100	Glyma.13G325900	plasma membrane intrinsic protein 2	64 kDa	6
	Glyma14g06680	Glyma.14G061500	plasma membrane intrinsic protein	31 kDa	6
	Glyma12g29510	Glyma.12G172500	plasma membrane intrinsic protein	31 kDa	6
	Glyma08g12650	Glyma.08G120100	NOD26-like intrinsic protein 1;2	29 kDa	6

Supplemental Table 5-2. cont.,

Receptor kinase	Glyma14g39550	Glyma.14G214700	receptor-like kinase 1	68 kDa	2
	Glyma02g41160	Glyma.02G244400	RKL1 (receptor-like kinase 1)	63 kDa	2
SPFH/Band 7/PHB: Prohibitin, Flottilin	Glyma05g01360	Glyma.05G029800	SPFH/Band 7/PHB domain-containing membrane-associated protein family/Flotillins	32 kDa	0
	Glyma06g06930	Glyma.06G065600	SPFH/Band 7/PHB domain-containing membrane-associated protein family/Flotillins	53 kDa	0
	Glyma13g05120	Glyma.13G065000	SPFH/Band 7/PHB domain-containing membrane-associated protein family/Flotillins	32 kDa	0
	Glyma16g32990	Glyma.16G204900	SPFH/Band 7/PHB domain-containing membrane-associated protein family/Flotillins	33 kDa	0
	Glyma17g10520	Glyma.17G097000	SPFH/Band 7/PHB domain-containing membrane-associated protein family/Flotillins	31 kDa	0
	Glyma19g02370	Glyma.19G020000	SPFH/Band 7/PHB domain-containing membrane-associated protein family/Flotillins	32 kDa	0

Supplemental Table 5-2. cont.,

Remorin	Glyma07g07970	Glyma.07G073100	Remorin family protein	22 kDa	0
	Glyma16g31290	Glyma.16G189500	Remorin family protein	22 kDa	0
Vesicle trafficking	Glyma13g31650	Glyma.13G243100	Translation protein SH3-like family	17 kDa	0
	Glyma01g40150	Glyma.01G193800	Translation protein SH3-like family protein	17 kDa	0
	Glyma03g36560	Glyma.03G207500	Translation protein SH3-like family protein	19 kDa	0
	Glyma13g31500	Glyma.13G241300	Clathrin light chain protein	35 kDa	0
	Glyma14g06340	Glyma.14G058300	Clathrin, heavy chain	193 kDa	0
	Glyma18g02960	Glyma.18G026000	Clathrin, heavy chain	193 kDa	0
	Glyma05g34540	Glyma.05G239300	DYNAMIN-like 1C	69 kDa	0
	Glyma08g45380	Glyma.08G340200	DYNAMIN-like 1E	69 kDa	0
	Glyma09g00430	Glyma.09G002100	dynamin-like 3	100 kDa	0
	Glyma12g37100	Glyma.12G242300	dynamin-like 3	100 kDa	0
	Glyma17g00480	Glyma.17G001500	dynamin-like 3	100 kDa	0
	Glyma13g32940	Glyma.13G256100	dynamin-related protein 3A	91 kDa	0
	Glyma05g36840	Glyma.05G217300	ADL1,ADL1A,AG68,DL1,DRP1A,RSW9 dynamin-like protein	68 kDa	0

Supplemental Table 5-2. cont.,

	Glyma13g26990	Glyma.13G199800	ANNAT8 annexin 8	36 kDa	0
	Glyma15g38040	Glyma.15G238200	ANNAT8 annexin 8	36 kDa	0
Vacuolar ATPase	Glyma07g00430	Glyma.07G002300	vacuolar ATP synthase subunit A	69 kDa	0
	Glyma02g07260	Glyma.02G065500	vacuolar ATP synthase subunit C (VATC) / V-ATPase C subunit / vacuolar proton pump C subunit (DET3)	43 kDa	0
	Glyma16g05470	Glyma.16G050200	vacuolar ATP synthase subunit C (VATC) / V-ATPase C subunit / vacuolar proton pump C subunit (DET3)	44 kDa	0
	Glyma16g26220	Glyma.16G146600	vacuolar ATP synthase subunit C (VATC) / V-ATPase C subunit / vacuolar proton pump C subunit (DET3)	39 kDa	0
	Glyma19g27360	Glyma.19G101000	vacuolar ATP synthase subunit C (VATC) / V-ATPase C subunit / vacuolar proton pump C subunit (DET3)	43 kDa	0
	Glyma08g23370	Glyma.08G218500	vacuolar ATP synthase subunit D (VATD) / V-ATPase D subunit / vacuolar proton pump D subunit (VATPD)	33 kDa	0
	Glyma02g06710	Glyma.02G059800	Vacuolar ATP synthase subunit H family protein	71 kDa	0
	Glyma16g25740	Glyma.16G142600	Vacuolar ATP synthase subunit H family protein	55 kDa	0

Chapter 6 Conclusion and Perspective

Portions of this conclusion chapter are from the addendum to "The Gm*FWL1* (*FW2-2-like*) nodulation gene encodes a plasma membrane microdomain-associated protein", which is published in Plant signaling & Behavior (DOI: 10.1080/15592324.2017.1365215), and allowed incorporation into this dissertation.

Zhenzhen Qiao and Marc Libault. (2017) Addendum: Function of plasma membrane microdomain-associated proteins during legume nodulation. Plant Signaling & Behavior.

Author contributions:

Dr. Marc Libault and I conceived and wrote the addendum published in Plant signaling & Behavior.

Root hair cell – A single plant cell type

Root hair cells maximize the surface of interaction between the root system and the rhizosphere to enhance nutrient and water uptakes. In legumes, the root hair cell is also the primary site of infection for the nitrogen-fixing bacterium, rhizobium. Based on these features, the root hair cell was selected to study the molecular mechanisms controlling plant cell elongation, and plant responses to abiotic and biotic stresses. Combining comparative functional analyses among different species, the conservation and divergence of the molecular mechanisms regulating root hair cell biology can be deciphered.

Having this objective in mind, we have developed the ultrasound aeroponic system to grow plants and then isolate root hair cells from different plant species. To date, 10 plant species (i.e. soybean, common bean, *Medicago*, *Arabidopsis*, sorghum, wheat, barley, rice, maize and millet) have been successfully cultivated in this system including dicotyledons and monocotyledons. Having access to diverse pools of root hair cells from these plant species allows accurate comparative genomic, transcriptomic, functional, and epigenomic analyses. Applying different environmental stresses on these plants, we also have an opportunity to clearly understand the molecular mechanisms controlling the expression of genes in root hair cells in response to biotic and abiotic stresses and the evolution of these responses between plant species. Similarly to the approach we used to identify conserved root hair *cis*-regulatory elements (see Chapter 3), we are proposing that additional regulatory elements controlling the expression of genes in response to environmental stresses could be characterized. Ultimately, the identification of conserved promoters and *cis-regulatory*

elements between plant species represents a unique set of molecular tools to precisely control the expression of plant transgenes in a specific cell type or in response to a specific biological process, leading to better understand the biological and molecular function of the transgenes. The access to the promoter/*cis*-regulatory elements described in this thesis (Chapter 3) will allow accurate functional genomic analyses at the level of the root hair cell in various plant species.

Functional conservation and divergence of duplicated genes

To maximize the outcome of functional genomic studies, plant biologists must be capable to transfer the biological knowledge gained from one model plant to other plant species, such as crops to enhance their agronomic properties. Such transfer is based on the assumption that homologous/orthologous genes are performing similar functions in different plant species. However, the multiple partial or whole duplication events of the plant genomes after speciation necessarily increase the pool of orthologous genes between two plant species creating a difficulty when identifying true functional homologs between plant species (Blanc and Wolfe, 2004b). For instance, *Medicago* and soybean genomes experienced one round of WGD 56.5 mya, then the soybean genome experienced another round of WGD 13 mya (Lavin et al., 2005; Schmutz et al., 2014). Although duplicated genes loss rates reached 74% and 58% respectively following the ancient and recent WGDs of the soybean genome (Schmutz et al. 2014), the remaining genes still represent an enrichment in the pool of genes compared to the *Medicago* genome. Therefore, an accurate evaluation of the evolution of duplicated genes is urgently needed.

Taking advantage of the release of soybean and *Medicago* genomes, we precisely delineated the paralogous and orthologous relationships existing between genes controlling the nodulation process (Chapter 4). For instance, we characterized the soybean genes orthologous to Mt*FLOT4*, Mt*HAP2-1* and Mt*NIN*, *Medicago* genes controlling the nodulation process. Similarly to Mt*FLOT4*, Mt*HAP2-1* and Mt*NIN*, these soybean orthologs are also preferentially expressed during the nodulation process, suggesting the functional conservation of these soybean and *Medicago* genes. Besides, we also observed the transcriptional divergence of soybean genes orthologous to *Medicago* genes controlling the nodulation process, notably those involved in nodule development. For example, Mt*Ann1*, encoding an annexin protein, has been found highly expressed in mature *Medicago* indeterminate nodules (Marx et al., 2016), but its soybean ortholog gene is broadly expressed in soybean root, shoot, flower, pod, and nodule tissues (Libault et al., 2010f). Nodulin 26, a soybean protein which controls ammonia transportation during nitrogen fixation and assimilation, was detected in the soybean symbiosome membrane, but its *Medicago* ortholog was not found in the *Medicago* symbiosome membrane (Hwang et al., 2010; Panter et al., 2000). The higher level of transcriptional divergence between soybean and *Medicago* genes controlling the late stage of nodulation could be related to the different developmental programs between determinate and indeterminate nodules in soybean and *Medicago*, respectively.

Besides the transcriptomic regulation, other molecular mechanisms exist to control legume nodule development. For instance, small RNA mi393 and mi164 regulate indeterminate but not determinate nodule development (Mao et al., 2013). Plant hormones are also important regulators of the nodulation process: while the inhibition

of auxin polar transportation triggers indeterminate nodule formation, the increase in polar auxin transportation as observed in Lotus, is essential to the development of determinate nodules. These observations suggest different roles of auxin transport during nodulation between legume species (Boot et al., 1999; Ferguson and Mathesius, 2014; Pacios-Bras et al., 2003). Therefore, distinct regulatory mechanisms participate in the nodulation process. However, as mentioned above, due to the anatomic and physiological complexity of nodules, it is not easy to decipher these mechanisms and their evolution, which is also a limitation presented in the project when analyzed the transcription conservation/divergence of genes controlling the late stage of nodule development. Hence, there is a need to separate the different cell types composing the nodule (i.e., the infected and uninfected cells of the nodule) to better appreciate the similarity and differences existing between these cells and across different legume species.

Function of plasma membrane microdomain-associated proteins during legume nodulation

In this thesis, we characterized the cellular function of the soybean GmFWL1 and GmFLOT2/4 proteins, two major regulators of soybean nodulation. Our co-immunoprecipitation assays on soybean nodules and microscopic observations of the protein localizations revealed that both proteins interact together and are plasma membrane microdomain-associated proteins (Chapter 5). Plasma membrane microdomains are plasma membrane sub-compartments characterized by their enrichment in sphingolipids and sterols and by a specific set of proteins (Pike, 2006).

They are involved in the recognition of signal molecules, in the transduction of these signals, and in the control of the endocytosis and exocytosis processes (Falk et al., 2004).

Co-immunoprecipitation assays performed on soybean nodules revealed 178 GmFWL1 protein partners including a large number of microdomain-associated proteins such as GmFLOT2/4. Among these 178 partners, 10 were repetitively characterized as GmFWL1 partners across the three independent biological replicates (Supplemental Table 5-2). Mining the soybean transcriptome atlas (Libault et al., 2010f), the genes encoding these GmFWL1 partners were highly and preferentially expressed in underground tissue, including 3 genes highly preferentially expressed in soybean nodules (Glyma.05G029800, Glyma.06G065600, and Glyma.13G065000). These three genes encode SPFH/Band 7/PHB domain–containing membrane-associated proteins, well-characterized plasma membrane microdomain-associated proteins. Interestingly, the *Arabidopsis thaliana* orthologs of Glyma.13G065000 and Glyma.05G029800, *HYPERSENSITIVE INDUCED REACTION 1* (At*HIR1*, AT1G69840) and *HYPERSENSITIVE INDUCED REACTION 4* (At*HIR4*, At5g62740), respectively, were previously characterized for their role in plant immunity (Faulkner, 2013; Qi et al., 2011). Notably, supporting its role as part of the plant defense system, At*HIR1* encodes a plasma membrane-associated protein capable to oligomerize upon pathogen perception (Qi et al., 2011).

Another GmFWL1 protein partner is the soybean GmFLOT2/4 (Glyma.06G065600) which is encoded by a gene orthologous to the *Medicago truncatula FLOT2* and *FLOT4* genes (Qiao et al., 2017). These two genes encode

regulators of the early events on the *M. truncatula* nodulation process (Haney and Long, 2010a). In *Medicago*, the "flotillin" family is composed by 11 members (e-value<2.3e-169 for 10 out the 11 member; Medtr3g065540 e-value is 1.8e-36 due to its shorter coding sequence). Among them, pairs of tandem duplicated flotillin genes were repetitively noticed. Mining the soybean genome, we identified only two flotillin genes, Glyma.06G065600 (e-value = 2.5e-256) and Glyma.04G064400 (e-value = 4.1e-21), which encode 482 and 97 amino acid proteins, respectively. Due to its short coding sequence and relatively low expression in soybean tissues, Glyma.04G064400 could be a potential pseudogene. In *Arabidopsis*, AtFLOT1 (At5g25250) has been characterized for its role in the endocytosis process independently of the clatharin mediation (Li et al., 2012). Similar function in plants is also supported by studies conducted on mammalian systems where flotillin proteins are playing an active role during endocytosis (Otto and Nichols, 2011). Hypothesizing that soybean flotillin might also control endocytosis, it is attempting to suggest that FWL1 and FLOT2/4 may form a protein complex to mediate the formation then elongation of the infection thread, a tubular structure allowing the infection of the plant by the symbiotic bacteria, and, ultimately, the endocytosis of the bacteroids into the infected plant nodule cells.

Beside FWL1 and flotillin proteins, remorins, which are also well-characterized plasma membrane microdomain-associated proteins, are also controlling the nodulation process. In *Medicago*, protein complementation assays revealed the interaction between SYMREM1 and Nod factor receptor like kinases (NFP, LYK3 and DMI2), suggesting a potential role of SYMREM1 in NF signal perception (Lefebvre et al., 2010a). Considering that in *Medicago*, MtFLOT4 and MtLYK3 co-localized in the root hair cell

tip in response to rhizobia infection, and that the soybean GmFLOT2/4 and GmFWL1 display a similar translocation at the tip of the root hair upon rhizobial inoculation, we hypothesize that FWL1, flotillin, and remorin work together with the NF receptors during the early stages of rhizobial infection.

The H^+-ATPase protein encoded by Glyma.09g056300 interacts with GmFWL1. Glyma.09g056300 is preferentially expressed in underground tissue (i.e., the root system and root hair cells). Its homolog in *Arabidopsis*, At*AHA1,* regulates the guard cell turgor pressure by interacting other plasma membrane-associated protein (Hashimoto-Sugimoto et al., 2013). The regulation of stomata aperture is the consequence of environmental changes, such as drought stress, and light perception (Higaki et al., 2013; Vasseur et al., 2011). In addition, in response to pathogens, At*AHA1* also regulates the jasmonic acid pathway to mediate microbial infection through the stomata (Zhou et al., 2015). Supporting the role of H^+-ATPase during nodulation, the medicago H^+-ATPase Mt*AHA5* has been characterized as a regulator of the very early stage of the nodulation process: at the time of the interaction between medicago plants and their symbiotic rhizobia (Nguyen et al., 2015).

Plasma membrane intrinsic proteins (PIP) are also important in controlling the nodulation process based on the interaction of the proteins encoded by the Glyma.8G015300, Glyma.02G073600 and Glyma.14G061500 genes and GmFWL1 (Supplemental Table 5-2). These proteins belong to the PIP1 and PIP2 subgroups. The PIP family proteins are involved in plant response to environmental stresses, such as water deficiency, salt, heavy metal and cold stresses (Smith‐Espinoza et al., 2003). For instance, GmPIP1;6 confers salt tolerance to soybean plants (Zhou et al., 2014). In

addition, PIPs are also important in controlling plant response to biotic stresses, such as plant-microbe symbiosis and plant immune response to pathogens (Aroca et al., 2007; Bárzana et al., 2014; Zou et al., 2005). Furthermore, in *Arabidopsis*, Li et al. (2011) reported that the PIP2;1 aquaporin co-localizes and co-migrates with FLOT1, supporting the location of PIP2 into plasma membrane microdomains (Li et al., 2011).

Three vacuolar ATPase proteins were also repetitively identified as GmFWL1 partners in nodules (Supplemental Table 5-2). These three proteins belong to different components of the ATPase complex: Glyma.08G218500, Glyma.08G224400, and Glyma.05G213800 encode a D, a A, and a E isoform 3 subunits, respectively. The vacuolar ATPases are proton pumps that responsible for acidification of intracellular compartments, including endosomes, lysosomes, phagosomes (Maxson and Grinstein, 2014). In *Arabidopsis*, the vacuolar H+-ATPases, A subunit, A1 are required for secretory and endocytic trafficking (Dettmer et al., 2006).

These ten soybean proteins repetitively interacting with GmFWL1 in nodules belong to 4 protein families. Among them, only flotillins have been well characterized for their role during legume nodulation (Haney and Long, 2010a) while H^+-ATPase, PIP and vacuolar ATPase are described as regulators of plant responses to biotic and abiotic stresses. Notably, most of them are channel proteins located in the plasma membrane allowing the regulation of plant cell homeostasis. In addition, it is important to note that flotillins, PIPs and vacuolars ATPases regulate the endocytosis process or membrane internalization, processes that relevant to the infection of plant cell by nitrogen-fixing symbiotic bacteria. Such observation suggests that plasma membrane microdomians might play a critical role in the recognition of the symbiotic bacteria,

then in the initiation and progression of the infection thread. Ultimately, microdomains could also regulate the final endocytosis-like process occurring in the nodule and resulting in the formation of the symbiosomes.

References

Abe, M., Takahashi, T., and Komeda, Y. (2001). Identification of a cis-regulatory element for L1 layer-specific gene expression, which is targeted by an L1-specific homeodomain protein. The Plant journal : for cell and molecular biology *26*, 487-494.

Adams, K.L. (2007). Evolution of duplicate gene expression in polyploid and hybrid plants. The Journal of heredity *98*, 136-141.

Alexandersson, E., Saalbach, G., Larsson, C., and Kjellbom, P. (2004). Arabidopsis plasma membrane proteomics identifies components of transport, signal transduction and membrane trafficking. Plant Cell Physiol *45*, 1543-1556.

Alvarado, V.Y., Tag, A., and Thomas, T.L. (2011). A cis regulatory element in the TAPNAC promoter directs tapetal gene expression. Plant Mol Biol *75*, 129-139.

Amor, B.B., Shaw, S.L., Oldroyd, G.E., Maillet, F., Penmetsa, R.V., Cook, D., Long, S.R., Denarie, J., and Gough, C. (2003). The NFP locus of Medicago truncatula controls an early step of Nod factor signal transduction upstream of a rapid calcium flux and root hair deformation. The Plant journal : for cell and molecular biology *34*, 495-506.

Ané, J.-M., Kiss, G.B., Riely, B.K., Penmetsa, R.V., Oldroyd, G.E., Ayax, C., Lévy, J., Debellé, F., Baek, J.-M., and Kalo, P. (2004). Medicago truncatula DMI1 required for bacterial and fungal symbioses in legumes. Science *303*, 1364-1367.

Ane, J.M., Levy, J., Thoquet, P., Kulikova, O., de Billy, F., Penmetsa, V., Kim, D.J., Debelle, F., Rosenberg, C., Cook, D.R*., et al.* (2002). Genetic and cytogenetic mapping of DMI1, DMI2, and DMI3 genes of Medicago truncatula involved in Nod factor transduction, nodulation, and mycorrhization. Mol Plant Microbe Interact *15*, 1108-1118.

Aroca, R., Porcel, R., and Ruiz‐Lozano, J.M. (2007). How does arbuscular mycorrhizal symbiosis regulate root hydraulic properties and plasma membrane aquaporins in Phaseolus vulgaris under drought, cold or salinity stresses? New Phytologist *173*, 808-816.

Arpat, A.B., Waugh, M., Sullivan, J.P., Gonzales, M., Frisch, D., Main, D., Wood, T., Leslie, A., Wing, R.A., and Wilkins, T.A. (2004a). Functional genomics of cell elongation in developing cotton fibers. Plant Mol Biol *54*, 911-929.

Arpat, A.B., Waugh, M., Sullivan, J.P., Gonzales, M., Frisch, D., Main, D., Wood, T., Leslie, A., Wing, R.A., and Wilkins, T.A. (2004b). Functional genomics of cell elongation in developing cotton fibers. Plant molecular biology *54*, 911-929.

Arrighi, J.-F., Barre, A., Amor, B.B., Bersoult, A., Soriano, L.C., Mirabella, R., de Carvalho-Niebel, F., Journet, E.-P., Ghérardi, M., and Huguet, T. (2006). The Medicago truncatula lysine motif-receptor-like kinase gene family includes NFP and new nodule-expressed genes. Plant physiology *142*, 265-279.

Babuke, T., Ruonala, M., Meister, M., Amaddii, M., Genzler, C., Esposito, A., and Tikkanen, R. (2009). Hetero-oligomerization of reggie-1/flotillin-2 and reggie-2/flotillin-1 is required for their endocytosis. Cellular signalling *21*, 1287-1297.

Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol *2*, 28-36.

Bárzana, G., Aroca, R., Bienert, G.P., Chaumont, F., and Ruiz-Lozano, J.M. (2014). New insights into the regulation of aquaporins by the arbuscular mycorrhizal symbiosis in maize plants under drought stress and possible implications for plant performance. Molecular Plant-Microbe Interactions *27*, 349-363.

Baudin, M., Laloum, T., Lepage, A., Ripodas, C., Ariel, F., Frances, L., Crespi, M., Gamas, P.C., Blanco, F.A., and Zanetti, M.E. (2015). A phylogenetically conserved group of NF-Y transcription factors interact to control nodulation in legumes. Plant physiology, pp. 01144.02015.

Becker, J.D., Boavida, L.C., Carneiro, J., Haury, M., and Feijo, J.A. (2003). Transcriptional profiling of Arabidopsis tissues reveals the unique characteristics of the pollen transcriptome. Plant physiology *133*, 713-725.

Becker, J.D., Takeda, S., Borges, F., Dolan, L., and Feijo, J.A. (2014a). Transcriptional profiling of Arabidopsis root hairs and pollen defines an apical cell growth signature. BMC Plant Biol *14*, 197.

Becker, J.D., Takeda, S., Borges, F., Dolan, L., and Feijó, J.A. (2014b). Transcriptional profiling of Arabidopsis root hairs and pollen defines an apical cell growth signature. BMC plant biology *14*, 197.

Beilstein, M.A., Nagalingum, N.S., Clements, M.D., Manchester, S.R., and Mathews, S. (2010). Dated molecular phylogenies indicate a Miocene origin for Arabidopsis thaliana. Proceedings of the National Academy of Sciences *107*, 18724-18728.

Benedito, V.A., Torres‐Jerez, I., Murray, J.D., Andriankaja, A., Allen, S., Kakar, K., Wandrey, M., Verdier, J., Zuber, H., and Ott, T. (2008). A gene expression atlas of the model legume Medicago truncatula. The Plant Journal *55*, 504-513.

Benfey, P.N., Linstead, P.J., Roberts, K., Schiefelbein, J.W., Hauser, M.T., and Aeschbacher, R.A. (1993). Root development in Arabidopsis: four mutants with dramatically altered root morphogenesis. Development (Cambridge, England) *119*, 57-70.

Benfey, P.N., Ren, L., and Chua, N.H. (1989). The CaMV 35S enhancer contains at least two domains which can confer different developmental and tissue-specific expression patterns. EMBO J *8*, 2195-2202.

Bernsel, A., Viklund, H., Hennerdal, A., and Elofsson, A. (2009). TOPCONS: consensus prediction of membrane protein topology. Nucleic Acids Res *37*, W465-468.

Birnbaum, K., Shasha, D.E., Wang, J.Y., Jung, J.W., Lambert, G.M., Galbraith, D.W., and Benfey, P.N. (2003). A gene expression map of the Arabidopsis root. Science *302*, 1956-1960.

Bisseling, T., and Ramos Escribano, J. (2003). A method for the isolation of root hairs from the model legume Medicago truncatula.

Blanc, G., Hokamp, K., and Wolfe, K.H. (2003). A recent polyploidy superimposed on older large-scale duplications in the Arabidopsis genome. Genome research *13*, 137-144.

Blanc, G., and Wolfe, K.H. (2004a). Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. Plant Cell *16*, 1679-1691.

Blanc, G., and Wolfe, K.H. (2004b). Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. The Plant Cell *16*, 1679-1691.

Boot, K.J., van Brussel, A.A., Tak, T., Spaink, H.P., and Kijne, J.W. (1999). Lipochitin oligosaccharides from Rhizobium leguminosarum bv. viciae reduce auxin transport

capacity in Vicia sativa subsp. nigra roots. Molecular plant-microbe interactions *12*, 839-844.

Boron, A.K., Van Orden, J., Nektarios Markakis, M., Mouille, G., Adriaensen, D., Verbelen, J.P., Hofte, H., and Vissenberg, K. (2014). Proline-rich protein-like PRPL1 controls elongation of root hairs in Arabidopsis thaliana. Journal of experimental botany *65*, 5485-5495.

Bowers, J.E., Chapman, B.A., Rong, J., and Paterson, A.H. (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature *422*, 433-438.

Brady, S.M., Orlando, D.A., Lee, J.-Y., Wang, J.Y., Koch, J., Dinneny, J.R., Mace, D., Ohler, U., and Benfey, P.N. (2007a). A high-resolution root spatiotemporal map reveals dominant expression patterns. Science *318*, 801-806.

Brady, S.M., Orlando, D.A., Lee, J.Y., Wang, J.Y., Koch, J., Dinneny, J.R., Mace, D., Ohler, U., and Benfey, P.N. (2007b). A high-resolution root spatiotemporal map reveals dominant expression patterns. Science *318*, 801-806.

Breakspear, A., Liu, C., Roy, S., Stacey, N., Rogers, C., Trick, M., Morieri, G., Mysore, K.S., Wen, J., Oldroyd, G.E*., et al.* (2014). The root hair "infectome" of Medicago truncatula uncovers changes in cell cycle genes and reveals a requirement for Auxin signaling in rhizobial infection. Plant Cell *26*, 4680-4701.

Brechenmacher, L., Lee, J., Sachdev, S., Song, Z., Nguyen, T.H., Joshi, T., Oehrle, N., Libault, M., Mooney, B., Xu, D*., et al.* (2009a). Establishment of a protein reference map for soybean root hair cells. Plant physiology *149*, 670-682.

Brechenmacher, L., Lee, J., Sachdev, S., Song, Z., Nguyen, T.H., Joshi, T., Oehrle, N., Libault, M., Mooney, B., Xu, D*., et al.* (2009b). Establishment of a protein reference map for soybean root hair cells. Plant Physiol *149*, 670-682.

Brechenmacher, L., Lee, J., Sachdev, S., Song, Z., Nguyen, T.H.N., Joshi, T., Oehrle, N., Libault, M., Mooney, B., Xu, D*., et al.* (2009c). Establishment of a Protein Reference Map for Soybean Root Hair Cells. Plant Physiology *149*, 670-682.

Brechenmacher, L., Lei, Z., Libault, M., Findley, S., Sugawara, M., Sadowsky, M.J., Sumner, L.W., and Stacey, G. (2010a). Soybean metabolites regulated in root hairs in

response to the symbiotic bacterium Bradyrhizobium japonicum. Plant Physiol *153*, 1808-1822.

Brechenmacher, L., Lei, Z., Libault, M., Findley, S., Sugawara, M., Sadowsky, M.J., Sumner, L.W., and Stacey, G. (2010b). Soybean metabolites regulated in root hairs in response to the symbiotic bacterium Bradyrhizobium japonicum. Plant Physiol *153*, 1808-1822.

Brechenmacher, L., Nguyen, T.H., Hixson, K., Libault, M., Aldrich, J., Pasa-Tolic, L., and Stacey, G. (2012a). Identification of soybean proteins from a single cell type: the root hair. Proteomics *12*, 3365-3373.

Brechenmacher, L., Nguyen, T.H.N., Hixson, K., Libault, M., Aldrich, J., Pasa-Tolic, L., and Stacey, G. (2012b). Identification of soybean proteins from a single cell type: The root hair. Proteomics *12*, 3365-3373.

Broughton, W.J., and Dilworth, M.J. (1971a). Control of leghaemoglobin synthesis in snake beans. The Biochemical journal *125*, 1075-1080.

Broughton, W.J., and Dilworth, M.J. (1971b). Control of leghaemoglobin synthesis in snake beans. Biochem J *125*, 1075-1080.

Browman, D.T., Hoegg, M.B., and Robbins, S.M. (2007). The SPFH domain-containing proteins: more than lipid raft markers. Trends in cell biology *17*, 394-402.

Bruex, A., Kainkaryam, R.M., Wieckowski, Y., Kang, Y.H., Bernhardt, C., Xia, Y., Zheng, X., Wang, J.Y., Lee, M.M., Benfey, P.*, et al.* (2012). A gene regulatory network for root epidermis cell differentiation in Arabidopsis. PLoS Genet *8*, e1002446.

Bucher*, M., Schroeer, B., Willmitzer, L., and Riesmeier, J. (1997). Two genes encoding extensin-like proteins are predominantly expressed in tomato root hair cells. Plant Mol Biol *35*, 497-508.

Capoen, W., Den Herder, J., Sun, J., Verplancke, C., De Keyser, A., De Rycke, R., Goormachtig, S., Oldroyd, G., and Holsters, M. (2009). Calcium spiking patterns and the role of the calcium/calmodulin-dependent kinase CCaMK in lateral root base nodulation of Sesbania rostrata. Plant Cell *21*, 1526-1540.

Catoira, R., Galera, C., de Billy, F., Penmetsa, R.V., Journet, E.-P., Maillet, F., Rosenberg, C., Cook, D., Gough, C., and Dénarié, J. (2000). Four Genes of Medicago truncatula Controlling Components of a Nod Factor Transduction Pathway. The Plant Cell Online *12*, 1647-1665.

Cerri, M.R., Frances, L., Laloum, T., Auriac, M.-C., Niebel, A., Oldroyd, G.E., Barker, D.G., Fournier, J., and de Carvalho-Niebel, F. (2012). Medicago truncatula ERN transcription factors: regulatory interplay with NSP1/NSP2 GRAS factors and expression dynamics throughout rhizobial infection. Plant physiology *160*, 2155-2172.

Chaudhary, B., Flagel, L., Stupar, R.M., Udall, J.A., Verma, N., Springer, N.M., and Wendel, J.F. (2009). Reciprocal silencing, transcriptional bias and functional divergence of homeologs in polyploid cotton (gossypium). Genetics *182*, 503-517.

Chen, D.-S., Liu, C.-W., Roy, S., Cousins, D., Stacey, N., and Murray, J.D. (2015). Identification of a core set of rhizobial infection genes using data from single cell-types. Frontiers in plant science *6*.

Chen, H., Zou, Y., Shang, Y., Lin, H., Wang, Y., Cai, R., Tang, X., and Zhou, J.M. (2008). Firefly luciferase complementation imaging assay for protein-protein interactions in plants. Plant Physiol *146*, 368-376.

Choudhury, A., and Lahiri, A. (2008). Arabidopsis thaliana regulatory element analyzer. Bioinformatics *24*, 2263-2264.

Chupeau, M.-C., Granier, F., Pichon, O., Renou, J.-P., Gaudin, V., and Chupeau, Y. (2013). Characterization of the Early Events Leading to Totipotency in an Arabidopsis Protoplast Liquid Culture by Temporal Transcript Profiling. The Plant Cell Online.

Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. Plant journal : for cell and molecular biology *16*, 735-743.

Collier, R., Fuchs, B., Walter, N., Kevin Lutke, W., and Taylor, C.G. (2005). Ex vitro composite plants: an inexpensive, rapid method for root biology. The Plant journal : for cell and molecular biology *43*, 449-457.

Cong, B., and Tanksley, S.D. (2006). FW2.2 and cell cycle control in developing tomato fruit: a possible example of gene co-option in the evolution of a novel organ. Plant Mol Biol *62*, 867-880.

Cui, Y., Barampuram, S., Stacey, M.G., Hancock, C.N., Findley, S., Mathieu, M., Zhang, Z., Parrott, W.A., and Stacey, G. (2013). Tnt1 retrotransposon mutagenesis: a tool for soybean functional genomics. Plant physiology *161*, 36-47.

Curtis, M.D., and Grossniklaus, U. (2003). A gateway cloning vector set for high-throughput functional analysis of genes in planta. Plant Physiol *133*, 462-469.

D'Hont, A., Denoeud, F., Aury, J.-M., Baurens, F.-C., Carreel, F., Garsmeur, O., Noel, B., Bocs, S., Droc, G., and Rouard, M. (2012). The banana (Musa acuminata) genome and the evolution of monocotyledonous plants. Nature *488*, 213-217.

Dahan, Y., Rosenfeld, R., Zadiranov, V., and Irihimovitch, V. (2010). A proposed conserved role for an avocado FW2.2-like gene as a negative regulator of fruit cell division. Planta *232*, 663-676.

Dare, A.P., Schaffer, R.J., Lin-Wang, K., Allan, A.C., and Hellens, R.P. (2008). Identification of a cis-regulatory element by transient analysis of co-ordinately regulated genes. Plant Methods *4*, 17.

De Franceschi, P., Stegmeir, T., Cabrera, A., van der Knaap, E., Rosyara, U.R., Sebolt, A.M., Dondini, L., Dirlewanger, E., Quero-Garcia, J., Campoy, J.A.*, et al.* (2013). Cell number regulator genes in provide candidate genes for the control of fruit size in sweet and sour cherry. Molecular breeding : new strategies in plant improvement *32*, 311-326.

De Smet, R., Adams, K.L., Vandepoele, K., Van Montagu, M.C., Maere, S., and Van de Peer, Y. (2013). Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. Proc Natl Acad Sci U S A *110*, 2898-2903.

Deal, R.B., and Henikoff, S. (2010a). A Simple Method for Gene Expression and Chromatin Profiling of Individual Cell Types within a Tissue. Developmental cell *18*, 1030-1040.

Deal, R.B., and Henikoff, S. (2010b). A simple method for gene expression and chromatin profiling of individual cell types within a tissue. Dev Cell *18*, 1030-1040.

Deal, R.B., and Henikoff, S. (2011). The INTACT method for cell type-specific gene expression and chromatin profiling in Arabidopsis thaliana. Nature protocols *6*, 56-68.

Dettmer, J., Hong-Hermesdorf, A., Stierhof, Y.-D., and Schumacher, K. (2006). Vacuolar H+-ATPase activity is required for endocytic and secretory trafficking in Arabidopsis. The Plant Cell *18*, 715-730.

Ding, B., Turgeon, R., and Parthasarathy, M.V. (1991). Routine Cryofixation of Plant Tissue by Propane Jet Freezing for Freeze Substitution. Journal of Electron Microscopy Technique *19*, 107-117.

Ding, J., Li, X., and Hu, H. (2012). Systematic prediction of cis-regulatory elements in the Chlamydomonas reinhardtii genome using comparative genomics. Plant Physiol *160*, 613-623.

Falk, J., Thoumine, O., Dequidt, C., Choquet, D., and Faivre-Sarrailh, C. (2004). NrCAM coupling to the cytoskeleton depends on multiple protein domains and partitioning into lipid rafts. Molecular biology of the cell *15*, 4695-4709.

Fan, L., Li, R., Pan, J., Ding, Z., and Lin, J. (2015). Endocytosis and its regulation in plants. Trends Plant Sci *20*, 388-397.

Faulkner, C. (2013). Receptor-mediated signaling at plasmodesmata. Frontiers in plant science *4*.

Ferguson, B.J., and Mathesius, U. (2014). Phytohormone regulation of legume-rhizobia interactions. Journal of chemical ecology *40*, 770-790.

Fernow, I., Icking, A., and Tikkanen, R. (2007). Reggie-1 and reggie-2 localize in non-caveolar rafts in epithelial cells: cellular localization is not dependent on the expression of caveolin proteins. European journal of cell biology *86*, 345-352.

Franklin-Tong, V.E. (1999). Signaling and the Modulation of Pollen Tube Growth. The Plant Cell Online *11*, 727-738.

Frary, A., Nesbitt, T.C., Grandillo, S., Knaap, E., Cong, B., Liu, J., Meller, J., Elber, R., Alpert, K.B., and Tanksley, S.D. (2000). fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. Science *289*, 85-88.

Frick, M., Bright, N.A., Riento, K., Bray, A., Merrified, C., and Nichols, B.J. (2007). Coassembly of flotillins induces formation of membrane microdomains, membrane curvature, and vesicle budding. Current biology : CB *17*, 1151-1156.

Gage, D.J. (2004). Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes. Microbiology and Molecular Biology Reviews *68*, 280-300.

Galbraith, D.W., and Birnbaum, K. (2006). Global studies of cell type-specific gene expression in plants. Annu Rev Plant Biol *57*, 451-475.

Gao, Z., Zhao, R., and Ruan, J. (2013). A genome-wide cis-regulatory element discovery method based on promoter sequences and gene co-expression networks. BMC Genomics *14 Suppl 1*, S4.

Garsmeur, O., Schnable, J.C., Almeida, A., Jourda, C., D'Hont, A., and Freeling, M. (2014). Two evolutionarily distinct classes of paleopolyploidy. Mol Biol Evol *31*, 448-454.

Govindarajulu, M., Kim, S.Y., Libault, M., Berg, R.H., Tanaka, K., Stacey, G., and Taylor, C.G. (2009). GS52 ecto-apyrase plays a critical role during soybean nodulation. Plant Physiol *149*, 994-1004.

Guevara-Garcia, A., Lopez-Ochoa, L., Lopez-Bucio, J., Simpson, J., and Herrera-Estrella, L. (1998). A 42 bp fragment of the pmas1' promoter containing an ocs-like element confers a developmental, wound- and chemically inducible expression pattern. Plant Mol Biol *38*, 743-753.

Guo, M., Rupe, M.A., Dieter, J.A., Zou, J., Spielbauer, D., Duncan, K.E., Howard, R.J., Hou, Z., and Simmons, C.R. (2010). Cell Number Regulator1 affects plant and organ size in maize: implications for crop yield enhancement and heterosis. Plant Cell *22*, 1057-1073.

Haberer, G., Hindemitt, T., Meyers, B.C., and Mayer, K.F. (2004). Transcriptional similarities, dissimilarities, and conservation of cis-elements in duplicated genes of Arabidopsis. Plant Physiol *136*, 3009-3022.

Haney, C.H., and Long, S.R. (2010a). Plant flotillins are required for infection by nitrogen-fixing bacteria. Proceedings of the National Academy of Sciences *107*, 478-483.

Haney, C.H., and Long, S.R. (2010b). Plant flotillins are required for infection by nitrogen-fixing bacteria. Proc Natl Acad Sci U S A *107*, 478-483.

Haney, C.H., Riely, B.K., Tricoli, D.M., Cook, D.R., Ehrhardt, D.W., and Long, S.R. (2011). Symbiotic rhizobia bacteria trigger a change in localization and dynamics of the Medicago truncatula receptor kinase LYK3. Plant Cell *23*, 2774-2787.

Hashimoto-Sugimoto, M., Higaki, T., Yaeno, T., Nagami, A., Irie, M., Fujimi, M., Miyamoto, M., Akita, K., Negi, J., and Shirasu, K. (2013). A Munc13-like protein in Arabidopsis mediates H+-ATPase translocation that is essential for stomatal responses. Nature communications *4*, 2215.

Higaki, T., Hashimoto-Sugimoto, M., Akita, K., Iba, K., and Hasezawa, S. (2013). Dynamics and environmental responses of PATROL1 in Arabidopsis subsidiary cells. Plant and Cell Physiology *55*, 773-780.

Higgins, J., Magusin, A., Trick, M., Fraser, F., and Bancroft, I. (2012). Use of mRNA-seq to discriminate contributions to the transcriptome from the constituent genomes of the polyploid crop species Brassica napus. BMC Genomics *13*, 247.

Hossain, M.S., Joshi, T., and Stacey, G. (2015). System approaches to study root hairs as a single cell plant model: current status and future perspectives. Front Plant Sci *6*, 363.

Hoth, S., Niedermeier, M., Feuerstein, A., Hornig, J., and Sauer, N. (2010). An ABA-responsive element in the AtSUC1 promoter is involved in the regulation of AtSUC1 expression. Planta *232*, 911-923.

Hwang, J.H., Ellingson, S.R., and Roberts, D.M. (2010). Ammonia permeability of the soybean nodulin 26 channel. FEBS letters *584*, 4339-4343.

Inaba, T., Nagano, Y., Reid, J.B., and Sasaki, Y. (2000). DE1, a 12-base pair cis-regulatory element sufficient to confer dark-inducible and light down-regulated expression to a minimal promoter in pea. J Biol Chem *275*, 19723-19727.

Inaba, T., Nagano, Y., Sakakibara, T., and Sasaki, Y. (1999). Identification of a cis-regulatory element involved in phytochrome down-regulated expression of the pea small GTPase gene pra2. Plant Physiol *120*, 491-500.

Indrasumunar, A., Searle, I., Lin, M.H., Kereszt, A., Men, A., Carroll, B.J., and Gresshoff, P.M. (2011). Nodulation factor receptor kinase 1α controls nodule organ number in soybean (Glycine max L. Merr). The Plant Journal *65*, 39-50.

Ithal, N., Recknor, J., Nettleton, D., Maier, T., Baum, T.J., and Mitchum, M.G. (2007). Developmental Transcript Profiling of Cyst Nematode Feeding Cells in Soybean Roots. Molecular Plant-Microbe Interactions *20*, 510-525.

Ito, S., Ohtake, N., Sueyoshi, K., and Ohyama, T. (2007). Characteristics of initial growth of hypernodulation soybean mutants, NOD1-3, NOD2-4 and NOD3-7, affected by inoculation of bradyrhizobia and nitrate supply. Soil Science & Plant Nutrition *53*, 66-71.

Jarsch, I.K., Konrad, S.S., Stratil, T.F., Urbanus, S.L., Szymanski, W., Braun, P., Braun, K.H., and Ott, T. (2014). Plasma Membranes Are Subcompartmentalized into a Plethora of Coexisting and Diverse Microdomains in Arabidopsis and Nicotiana benthamiana. Plant Cell *26*, 1698-1711.

Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H., and Soltis, P.S. (2011). Ancestral polyploidy in seed plants and angiosperms. Nature *473*, 97-100.

Jones, K.M., Kobayashi, H., Davies, B.W., Taga, M.E., and Walker, G.C. (2007). How rhizobial symbionts invade plants: the Sinorhizobium–Medicago model. Nature Reviews Microbiology *5*, 619-633.

Jones, M.A., Raymond, M.J., and Smirnoff, N. (2006). Analysis of the root-hair morphogenesis transcriptome reveals the molecular identity of six genes with roles in root-hair development in Arabidopsis. The Plant journal : for cell and molecular biology *45*, 83-100.

Kathryn, M.J., Hajime, K., Bryan, W.D., Michiko, E.T., and Graham, C.W. (2007). How rhizobial symbionts invade plants: the Sinorhizobium–Medicago model. Nature Reviews Microbiology *5*, 619-633.

Kijne, J.W. (1992). The Rhizobium Infection Process Jan W. Kijne. Biological nitrogen fixation, 1349.

Kim, D.W., Lee, S.H., Choi, S.-B., Won, S.-K., Heo, Y.-K., Cho, M., Park, Y.-I., and Cho, H.-T. (2006a). Functional conservation of a root hair cell-specific cis-element in angiosperms with different root hair distribution patterns. The Plant Cell *18*, 2958-2970.

Kim, D.W., Lee, S.H., Choi, S.B., Won, S.K., Heo, Y.K., Cho, M., Park, Y.I., and Cho, H.T. (2006b). Functional conservation of a root hair cell-specific cis-element in angiosperms with different root hair distribution patterns. Plant Cell *18*, 2958-2970.

Kim, W.C., Ko, J.H., and Han, K.H. (2012). Identification of a cis-acting regulatory motif recognized by MYB46, a master transcriptional regulator of secondary wall biosynthesis. Plant Mol Biol *78*, 489-501.

Kistner, C., and Parniske, M. (2002). Evolution of signal transduction in intracellular symbiosis. Trends in plant science *7*, 511-518.

Klink, V., Alkharouf, N., MacDonald, M., and Matthews, B. (2005). Laser Capture Microdissection (LCM) and Expression Analyses of Glycine max (Soybean) Syncytium Containing Root Regions Formed by the Plant Pathogen Heterodera glycines (Soybean Cyst Nematode). Plant Mol Biol *59*, 965-979.

Kosuta, S., Held, M., Hossain, M.S., Morieri, G., MacGillivary, A., Johansen, C., Antolín‐Llovera, M., Parniske, M., Oldroyd, G.E., and Downie, A.J. (2011). Lotus japonicus symRK‐14 uncouples the cortical and epidermal symbiotic program. The Plant Journal *67*, 929-940.

Kurata, N., and Yamazaki, Y. (2006). Oryzabase. An Integrated Biological and Genome Information Database for Rice. Plant Physiology *140*, 12-17.

Laloum, T., Baudin, M., Frances, L., Lepage, A., Billault‐Penneteau, B., Cerri, M.R., Ariel, F., Jardinaud, M.F., Gamas, P., and Carvalho‐Niebel, F. (2014). Two CCAAT‐box‐binding transcription factors redundantly regulate early steps of the legume‐rhizobia endosymbiosis. The Plant Journal *79*, 757-768.

Lan, P., Li, W., Lin, W.D., Santi, S., and Schmidt, W. (2013). Mapping gene activity of Arabidopsis root hairs. Genome biology *14*, R67.

Langham, R.J., Walsh, J., Dunn, M., Ko, C., Goff, S.A., and Freeling, M. (2004). Genomic duplication, fractionation and the origin of regulatory novelty. Genetics *166*, 935-945.

Larrainzar, E., Riely, B.K., Kim, S.C., Carrasquilla-Garcia, N., Yu, H.J., Hwang, H.J., Oh, M., Kim, G.B., Surendrarao, A.K., Chasman, D.*, et al.* (2015). Deep Sequencing of

the Medicago truncatula Root Transcriptome Reveals a Massive and Early Interaction between Nodulation Factor and Ethylene Signals. Plant Physiol *169*, 233-265.

Lavin, M., Herendeen, P.S., and Wojciechowski, M.F. (2005). Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary. Syst Biol *54*, 575-594.

Lee, T.-H., Tang, H., Wang, X., and Paterson, A.H. (2013). PGDD: a database of gene and genome duplication in plants. Nucleic acids research *41*, D1152-D1158.

Lefebvre, B., Furt, F., Hartmann, M.A., Michaelson, L.V., Carde, J.P., Sargueil-Boiron, F., Rossignol, M., Napier, J.A., Cullimore, J., Bessoule, J.J*., et al.* (2007). Characterization of lipid rafts from Medicago truncatula root plasma membranes: a proteomic study reveals the presence of a raft-associated redox system. Plant Physiol *144*, 402-418.

Lefebvre, B., Timmers, T., Mbengue, M., Moreau, S., Hervé, C., Tóth, K., Bittencourt-Silvestre, J., Klaus, D., Deslandes, L., and Godiard, L. (2010a). A remorin protein interacts with symbiotic receptors and regulates bacterial infection. Proceedings of the National Academy of Sciences *107*, 2343-2348.

Lefebvre, B., Timmers, T., Mbengue, M., Moreau, S., Herve, C., Toth, K., Bittencourt-Silvestre, J., Klaus, D., Deslandes, L., Godiard, L*., et al.* (2010b). A remorin protein interacts with symbiotic receptors and regulates bacterial infection. Proc Natl Acad Sci U S A *107*, 2343-2348.

Lévy, J., Bres, C., Geurts, R., Chalhoub, B., Kulikova, O., Duc, G., Journet, E.-P., Ané, J.-M., Lauber, E., and Bisseling, T. (2004). A putative Ca2+ and calmodulin-dependent protein kinase required for bacterial and fungal symbioses. Science *303*, 1361-1364.

Li, R., Liu, P., Wan, Y., Chen, T., Wang, Q., Mettbach, U., Baluška, F., Šamaj, J., Fang, X., and Lucas, W.J. (2012). A membrane microdomain-associated protein, Arabidopsis Flot1, is involved in a clathrin-independent endocytic pathway and is required for seedling development. The Plant Cell *24*, 2105-2122.

Li, W., and Lan, P. (2015). Re-analysis of RNA-seq transcriptome data reveals new aspects of gene activity in Arabidopsis root hairs. Front Plant Sci *6*, 421.

Li, X., Wang, X., Yang, Y., Li, R., He, Q., Fang, X., Luu, D.-T., Maurel, C., and Lin, J. (2011). Single-molecule analysis of PIP2; 1 dynamics and partitioning reveals multiple

modes of Arabidopsis plasma membrane aquaporin regulation. The Plant Cell *23*, 3780-3797.

Li, Y., Zhu, Y., Liu, Y., Shu, Y., Meng, F., Lu, Y., Liu, B., Bai, X., and Guo, D. (2008). Cis-regulatory element based gene finding: an application in Arabidopsis thaliana. Genome Inform *21*, 177-187.

Li, Z., and He, C. (2015). Physalis floridana Cell Number Regulator1 encodes a cell membrane-anchored modulator of cell cycle and negatively controls fruit size. Journal of experimental botany *66*, 257-270.

Libault, M., Brechenmacher, L., Cheng, J., Xu, D., and Stacey, G. (2010a). Root hair systems biology. Trends Plant Sci *15*, 641-650.

Libault, M., Brechenmacher, L., Cheng, J., Xu, D., and Stacey, G. (2010b). Root hair systems biology. Trends Plant Sci *15*, 641-650.

Libault, M., and Chen, S. (2016). Editorial: Plant Single Cell Type Systems Biology. Front Plant Sci *7*, 35.

Libault, M., Farmer, A., Brechenmacher, L., Drnevich, J., Langley, R.J., Bilgin, D.D., Radwan, O., Neece, D.J., Clough, S.J., and May, G.D. (2010c). Complete transcriptome of the soybean root hair cell, a single-cell model, and its alteration in response to Bradyrhizobium japonicum infection. Plant physiology *152*, 541-552.

Libault, M., Farmer, A., Brechenmacher, L., Drnevich, J., Langley, R.J., Bilgin, D.D., Radwan, O., Neece, D.J., Clough, S.J., May, G.D*., et al.* (2010d). Complete Transcriptome of the Soybean Root Hair Cell, a Single-Cell Model, and Its Alteration in Response to Bradyrhizobium japonicum Infection. Plant Physiology *152*, 541-552.

Libault, M., Farmer, A., Joshi, T., Takahashi, K., Langley, R.J., Franklin, L.D., He, J., Xu, D., May, G., and Stacey, G. (2010e). An integrated transcriptome atlas of the crop model Glycine max, and its use in comparative analyses in plants. Plant J *63*, 86-99.

Libault, M., Farmer, A., Joshi, T., Takahashi, K., Langley, R.J., Franklin, L.D., He, J., Xu, D., May, G., and Stacey, G. (2010f). An integrated transcriptome atlas of the crop model Glycine max, and its use in comparative analyses in plants. The Plant Journal *63*, 86-99.

Libault, M., Farmer, A., Joshi, T., Takahashi, K., Langley, R.J., Franklin, L.D., He, J., Xu, D., May, G., and Stacey, G. (2010g). An integrated transcriptome atlas of the crop model Glycine max, and its use in comparative analyses in plants. The Plant journal : for cell and molecular biology *63*, 86-99.

Libault, M., Joshi, T., Takahashi, K., Hurley-Sommer, A., Puricelli, K., Blake, S., Finger, R.E., Taylor, C.G., Xu, D., Nguyen, H.T.*, et al.* (2009a). Large-scale analysis of putative soybean regulatory gene expression identifies a Myb gene involved in soybean nodule development. Plant Physiol *151*, 1207-1220.

Libault, M., Joshi, T., Takahashi, K., Hurley-Sommer, A., Puricelli, K., Blake, S., Finger, R.E., Taylor, C.G., Xu, D., Nguyen, H.T.*, et al.* (2009b). Large-scale analysis of putative soybean regulatory gene expression identifies a Myb gene involved in soybean nodule development. Plant Physiol *151*, 1207-1220.

Libault, M., Joshi, T., Takahashi, K., Hurley-Sommer, A., Puricelli, K., Blake, S., Finger, R.E., Taylor, C.G., Xu, D., Nguyen, H.T.*, et al.* (2009c). Large-scale analysis of putative soybean regulatory gene expression identifies a Myb gene involved in soybean nodule development. Plant Physiol *151*, 1207-1220.

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

Libault, M., Zhang, X.C., Govindarajulu, M., Qiu, J., Ong, Y.T., Brechenmacher, L., Berg, R.H., Hurley-Sommer, A., Taylor, C.G., and Stacey, G. (2010h). A member of the highly conserved FWL (tomato FW2.2-like) gene family is essential for soybean nodule organogenesis. The Plant journal : for cell and molecular biology *62*, 852-864.

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

Lin, Q., Ohashi, Y., Kato, M., Tsuge, T., Gu, H., Qu, L.J., and Aoyama, T. (2015). GLABRA2 Directly Suppresses Basic Helix-Loop-Helix Transcription Factor Genes with Diverse Functions in Root Hair Development. Plant Cell *27*, 2894-2906.

Lota, F., Wegmuller, S., Buer, B., Sato, S., Brautigam, A., Hanf, B., and Bucher, M. (2013). The cis-acting CTTC-P1BS module is indicative for gene function of LjVTI12, a Qb-SNARE protein gene that is required for arbuscule formation in Lotus japonicus. The Plant journal : for cell and molecular biology *74*, 280-293.

Lu, F., Lipka, A.E., Glaubitz, J., Elshire, R., Cherney, J.H., Casler, M.D., Buckler, E.S., and Costich, D.E. (2013). Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. PLoS Genet *9*, e1003215.

Lynch, M., and Conery, J.S. (2000). The evolutionary fate and consequences of duplicate genes. Science *290*, 1151-1155.

Lyons, E., and Freeling, M. (2008). How to usefully compare homologous plant genes and chromosomes as DNA sequences. The Plant journal : for cell and molecular biology *53*, 661-673.

Lyons, E., Pedersen, B., Kane, J., Alam, M., Ming, R., Tang, H., Wang, X., Bowers, J., Paterson, A., and Lisch, D. (2008a). Finding and comparing syntenic regions among Arabidopsis and the outgroups papaya, poplar, and grape: CoGe with rosids. Plant physiology *148*, 1772-1781.

Lyons, E., Pedersen, B., Kane, J., Alam, M., Ming, R., Tang, H., Wang, X., Bowers, J., Paterson, A., Lisch, D*., et al.* (2008b). Finding and comparing syntenic regions among Arabidopsis and the outgroups papaya, poplar, and grape: CoGe with rosids. Plant Physiol *148*, 1772-1781.

Madsen, E.B., Madsen, L.H., Radutoiu, S., Olbryt, M., Rakwalska, M., Szczyglowski, K., Sato, S., Kaneko, T., Tabata, S., and Sandal, N. (2003a). A receptor kinase gene of the LysM type is involved in legumeperception of rhizobial signals. Nature *425*, 637-640.

Madsen, E.B., Madsen, L.H., Radutoiu, S., Olbryt, M., Rakwalska, M., Szczyglowski, K., Sato, S., Kaneko, T., Tabata, S., Sandal, N*., et al.* (2003b). A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals. Nature *425*, 637-640.

Mao, G., Turner, M., Yu, O., and Subramanian, S. (2013). miR393 and miR164 influence indeterminate but not determinate nodule development. Plant signaling & behavior *8*, e26753.

Marsh, J.F., Rakocevic, A., Mitra, R.M., Brocard, L., Sun, J., Eschstruth, A., Long, S.R., Schultze, M., Ratet, P., and Oldroyd, G.E. (2007). Medicago truncatula NIN is essential for rhizobial-independent nodule organogenesis induced by autoactive calcium/calmodulin-dependent protein kinase. Plant physiology *144*, 324-335.

Marx, H., Minogue, C.E., Jayaraman, D., Richards, A.L., Kwiecien, N.W., Siahpirani, A.F., Rajasekar, S., Maeda, J., Garcia, K., and Del Valle-Echevarria, A.R. (2016). A proteomic atlas of the legume Medicago truncatula and its nitrogen-fixing endosymbiont Sinorhizobium meliloti. Nature biotechnology *34*, 1198-1205.

Matsuura, H., Takenami, S., Kubo, Y., Ueda, K., Ueda, A., Yamaguchi, M., Hirata, K., Demura, T., Kanaya, S., and Kato, K. (2013). A computational and experimental approach reveals that the 5'-proximal region of the 5'-UTR has a Cis-regulatory signature responsible for heat stress-regulated mRNA translation in Arabidopsis. Plant Cell Physiol *54*, 474-483.

Maxson, M.E., and Grinstein, S. (2014). The vacuolar-type H+-ATPase at a glance–more than a proton pump (The Company of Biologists Ltd).

Messinese, E., Mun, J.-H., Yeun, L.H., Jayaraman, D., Rougé, P., Barre, A., Lougnon, G., Schornack, S., Bono, J.-J., and Cook, D.R. (2007). A novel nuclear protein interacts with the symbiotic DMI3 calcium-and calmodulin-dependent protein kinase of Medicago truncatula. Molecular Plant-Microbe Interactions *20*, 912-921.

Mishra, B.S., Singh, M., Aggrawal, P., and Laxmi, A. (2009). Glucose and Auxin Signaling Interaction in Controlling <italic>Arabidopsis thaliana</italic> Seedlings Root Growth and Development. PLoS ONE *4*, e4502.

Miwa, H., Sun, J., Oldroyd, G.E., and Downie, J.A. (2006). Analysis of calcium spiking using a cameleon calcium sensor reveals that nodulation gene expression is regulated by calcium spike number and the developmental status of the cell. The Plant journal : for cell and molecular biology *48*, 883-894.

Miyahara, A., Richens, J., Starker, C., Morieri, G., Smith, L., Long, S., Downie, J.A., and Oldroyd, G.E. (2010). Conservation in function of a SCAR/WAVE component during infection thread and root hair growth in Medicago truncatula. Molecular plant-microbe interactions *23*, 1553-1562.

Moore, M.J., Bell, C.D., Soltis, P.S., and Soltis, D.E. (2007). Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. Proceedings of the National Academy of Sciences *104*, 19363-19368.

Morel, J., Claverol, S., Mongrand, S., Furt, F., Fromentin, J., Bessoule, J.J., Blein, J.P., and Simon-Plas, F. (2006). Proteomics of plant detergent-resistant membranes. Mol Cell Proteomics *5*, 1396-1411.

Myburg, A.A., Grattapaglia, D., Tuskan, G.A., Hellsten, U., Hayes, R.D., Grimwood, J., Jenkins, J., Lindquist, E., Tice, H., and Bauer, D. (2014). The genome of Eucalyptus grandis. Nature *510*, 356-362.

Nagano, Y., Inaba, T., Furuhashi, H., and Sasaki, Y. (2001). Trihelix DNA-binding protein with specificities for two distinct cis-elements: both important for light down-regulated and dark-inducible gene expression in higher plants. J Biol Chem *276*, 22238-22243.

Nesbitt, T.C., and Tanksley, S.D. (2001). fw2.2 directly affects the size of developing tomato fruit, with secondary effects on fruit number and photosynthate distribution. Plant Physiol *127*, 575-583.

Neumann-Giesen, C., Falkenbach, B., Beicht, P., Claasen, S., Luers, G., Stuermer, C.A., Herzog, V., and Tikkanen, R. (2004). Membrane and raft association of reggie-1/flotillin-2: role of myristoylation, palmitoylation and oligomerization and induction of filopodia by overexpression. Biochem J *378*, 509-518.

Nguyen, T.H., Brechenmacher, L., Aldrich, J.T., Clauss, T.R., Gritsenko, M.A., Hixson, K.K., Libault, M., Tanaka, K., Yang, F., Yao, Q.*, et al.* (2012a). Quantitative phosphoproteomic analysis of soybean root hairs inoculated with Bradyrhizobium japonicum. Mol Cell Proteomics *11*, 1140-1155.

Nguyen, T.H., Brechenmacher, L., Aldrich, J.T., Clauss, T.R., Gritsenko, M.A., Hixson, K.K., Libault, M., Tanaka, K., Yang, F., Yao, Q.*, et al.* (2012b). Quantitative Phosphoproteomic Analysis of Soybean Root Hairs Inoculated with Bradyrhizobium japonicum. Mol Cell Proteomics *11*, 1140-1155.

Nguyen, T.H.N., Brechenmacher, L., Aldrich, J.T., Clauss, T.R., Gritsenko, M.A., Hixson, K.K., Libault, M., Tanaka, K., Yang, F., Yao, Q.*, et al.* (2012c). Quantitative phosphoproteomic analysis of soybean root hairs inoculated with Bradyrhizobium japonicum. Mol Cell Proteomics *11*, 1140-1155, 1116 pp.

Nguyen, T.T., Volkening, J.D., Rose, C.M., Venkateshwaran, M., Westphall, M.S., Coon, J.J., Ané, J.-M., and Sussman, M.R. (2015). Potential regulatory phosphorylation sites in a Medicago truncatula plasma membrane proton pump implicated during early symbiotic signaling in roots. FEBS letters *589*, 2186-2193.

O'Rourke, J.A., Iniguez, L.P., Fu, F., Bucciarelli, B., Miller, S.S., Jackson, S.A., McClean, P.E., Li, J., Dai, X., Zhao, P.X.*, et al.* (2014). An RNA-Seq based gene expression atlas of the common bean. BMC Genomics *15*, 866.

O'Rourke, J.A., Iniguez, L.P., Fu, F., Bucciarelli, B., Miller, S.S., Jackson, S.A., McClean, P.E., Li, J., Dai, X., and Zhao, P.X. (2014). An RNA-Seq based gene expression atlas of the common bean. BMC genomics *15*, 1.

Oldroyd, G.E. (2013). Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. Nat Rev Microbiol *11*, 252-263.

Oldroyd, G.E., Murray, J.D., Poole, P.S., and Downie, J.A. (2011a). The rules of engagement in the legume-rhizobial symbiosis. Annu Rev Genet *45*, 119-144.

Oldroyd, G.E., Murray, J.D., Poole, P.S., and Downie, J.A. (2011b). The rules of engagement in the legume-rhizobial symbiosis. Annual review of genetics *45*, 119-144.

Omasits, U., Ahrens, C.H., Muller, S., and Wollscheid, B. (2014). Protter: interactive protein feature visualization and integration with experimental proteomic data. Bioinformatics *30*, 884-886.

Ondřej, V., Kitner, M., Doležalová, I., Nádvorník, P., Navrátilová, B., and Lebeda, A. (2009). Chromatin structural rearrangement during dedifferentiation of protoplasts of Cucumis sativus L. Mol Cells *27*, 443-447.

Otto, G.P., and Nichols, B.J. (2011). The roles of flotillin microdomains–endocytosis and beyond. J Cell Sci *124*, 3933-3940.

Pacios-Bras, C., Schlaman, H.R., Boot, K., Admiraal, P., Langerak, J.M., Stougaard, J., and Spaink, H.P. (2003). Auxin distribution in Lotus japonicus during root nodule development. Plant Mol Biol *52*, 1169-1180.

Padmalatha, K.V., Dhandapani, G., Kanakachari, M., Kumar, S., Dass, A., Patil, D.P., Rajamani, V., Kumar, K., Pathak, R., Rawat, B.*, et al.* (2012). Genome-wide transcriptomic analysis of cotton under drought stress reveal significant down-regulation of genes and pathways involved in fibre elongation and up-regulation of defense responsive genes. Plant Mol Biol *78*, 223-246.

Panchy, N., Lehti-Shiu, M., and Shiu, S.-H. (2016). Evolution of gene duplication in plants. Plant physiology *171*, 2294-2316.

Panter, S., Thomson, R., De Bruxelles, G., Laver, D., Trevaskis, B., and Udvardi, M. (2000). Identification with proteomics of novel proteins associated with the

peribacteroid membrane of soybean root nodules. Molecular plant-microbe interactions *13*, 325-333.

Perry, J.A., Wang, T.L., Welham, T.J., Gardner, S., Pike, J.M., Yoshida, S., and Parniske, M. (2003). A TILLING Reverse Genetics Tool and a Web-Accessible Collection of Mutants of the Legume Lotus japonicus. Plant Physiology *131*, 866-871.

Petersson, S.V., Johansson, A.I., Kowalczyk, M., Makoveychuk, A., Wang, J.Y., Moritz, T., Grebe, M., Benfey, P.N., Sandberg, G., and Ljung, K. (2009). An Auxin Gradient and Maximum in the Arabidopsis Root Apex Shown by High-Resolution Cell-Specific Analysis of IAA Distribution and Synthesis. The Plant Cell Online *21*, 1659-1668.

Petit, J.M., van Wuytswinkel, O., Briat, J.F., and Lobreaux, S. (2001). Characterization of an iron-dependent regulatory sequence involved in the transcriptional control of AtFer1 and ZmFer1 plant ferritin genes by iron. J Biol Chem *276*, 5584-5590.

Pike, L.J. (2006). Rafts defined: a report on the Keystone Symposium on Lipid Rafts and Cell Function. Journal of lipid research *47*, 1597-1598.

Pingault, L., Choulet, F., Alberti, A., Glover, N., Wincker, P., Feuillet, C., and Paux, E. (2015). Deep transcriptome sequencing provides new insights into the structural and functional organization of the wheat genome. Genome biology *16*, 29.

Pislariu, C.I., Murray, J.D., Wen, J., Cosson, V., Muni, R.R.D., Wang, M., Benedito, V.A., Andriankaja, A., Cheng, X., and Jerez, I.T. (2012). A Medicago truncatula tobacco retrotransposon insertion mutant collection with defects in nodule development and symbiotic nitrogen fixation. Plant physiology *159*, 1686-1699.

Pla, M., Vilardell, J., Guiltinan, M.J., Marcotte, W.R., Niogret, M.F., Quatrano, R.S., and Pages, M. (1993). The cis-regulatory element CCACGTGG is involved in ABA and water-stress responses of the maize gene rab28. Plant Mol Biol *21*, 259-266.

Pontier, D., Balague, C., Bezombes-Marion, I., Tronchet, M., Deslandes, L., and Roby, D. (2001). Identification of a novel pathogen-responsive element in the promoter of the tobacco gene HSR203J, a molecular marker of the hypersensitive response. The Plant journal : for cell and molecular biology *26*, 495-507.

Qi, Y., Tsuda, K., Nguyen, L.V., Wang, X., Lin, J., Murphy, A.S., Glazebrook, J., Thordal-Christensen, H., and Katagiri, F. (2011). Physical association of Arabidopsis

hypersensitive induced reaction proteins (HIRs) with the immune receptor RPS2. Journal of Biological Chemistry *286*, 31297-31307.

Qiao, Z., Brechenmacher, L., Smith, B., Strout, G.W., Mangin, W., Taylor, C., Russell, S.D., Stacey, G., and Libault, M. (2017). The GmFWL1 (FW2‐2‐like) nodulation gene encodes a plasma membrane microdomain‐associated protein. Plant, Cell & Environment.

Qiao, Z., and Libault, M. (2013a). Unleashing the potential of the root hair cell as a single plant cell type model in root systems biology. Frontiers in plant science *4*, 484.

Qiao, Z., and Libault, M. (2013b). Unleashing the potential of the root hair cell as a single plant cell type model in root systems biology. Front Plant Sci *4*, 484.

Radutoiu, S., Madsen, L.H., Madsen, E.B., Felle, H.H., Umehara, Y., Grønlund, M., Sato, S., Nakamura, Y., Tabata, S., and Sandal, N. (2003). Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. Nature *425*, 585-592.

Raffaele, S., Bayer, E., Lafarge, D., Cluzet, S., German Retana, S., Boubekeur, T., Leborgne-Castel, N., Carde, J.P., Lherminier, J., Noirot, E.*, et al.* (2009). Remorin, a solanaceae protein resident in membrane rafts and plasmodesmata, impairs potato virus X movement. Plant Cell *21*, 1541-1555.

Ramakers, C., Ruijter, J.M., Deprez, R.H.L., and Moorman, A.F.M. (2003). Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. Neurosci Lett *339*, 62-66.

Renny-Byfield, S., and Wendel, J.F. (2014). Doubling down on genomes: polyploidy and crop plants. American journal of botany *101*, 1711-1725.

Rivera-Milla, E., Stuermer, C.A., and Malaga-Trillo, E. (2006). Ancient origin of reggie (flotillin), reggie-like, and other lipid-raft proteins: convergent evolution of the SPFH domain. Cellular and molecular life sciences : CMLS *63*, 343-357.

Roehm, M., and Werner, D. (1987). Isolation of root hairs from seedlings of Pisum sativum. Identification of root hair specific proteins by in situ labeling. Physiol Plant *69*, 129-136.

Rogg, L.E., Lasswell, J., and Bartel, B. (2001). A Gain-of-Function Mutation in IAA28 Suppresses Lateral Root Development. The Plant Cell Online *13*, 465-480.

Roulin, A., Auer, P.L., Libault, M., Schlueter, J., Farmer, A., May, G., Stacey, G., Doerge, R.W., and Jackson, S.A. (2013). The fate of duplicated genes in a polyploid plant genome. The Plant journal : for cell and molecular biology *73*, 143-153.

Roux, B., Rodde, N., Jardinaud, M.F., Timmers, T., Sauviac, L., Cottret, L., Carrere, S., Sallet, E., Courcelle, E., Moreau, S.*, et al.* (2014). An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. The Plant journal : for cell and molecular biology *77*, 817-837.

Ruan, Y.-L., Llewellyn, D.J., and Furbank, R.T. (2001). The Control of Single-Celled Cotton Fiber Elongation by Developmentally Reversible Gating of Plasmodesmata and Coordinated Expression of Sucrose and K+ Transporters and Expansin. The Plant Cell Online *13*, 47-60.

Safrany, J., Haasz, V., Mate, Z., Ciolfi, A., Feher, B., Oravecz, A., Stec, A., Dallmann, G., Morelli, G., Ulm, R.*, et al.* (2008). Identification of a novel cis-regulatory element for UV-B-induced transcription in Arabidopsis. The Plant journal : for cell and molecular biology *54*, 402-414.

Santi, S., and Schmidt, W. (2008). Laser microdissection-assisted analysis of the functional fate of iron deficiency-induced root hairs in cucumber. Journal of experimental botany *59*, 697-704.

Satheesh, V., Jagannadham, P.T., Chidambaranathan, P., Jain, P.K., and Srinivasan, R. (2014). NAC transcription factor genes: genome-wide identification, phylogenetic, motif and cis-regulatory element analysis in pigeonpea (Cajanus cajan (L.) Millsp.). Mol Biol Rep *41*, 7763-7773.

Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., Kato, T., Nakao, M., Sasamoto, S., Watanabe, A., Ono, A., and Kawashima, K. (2008a). Genome structure of the legume, Lotus japonicus. DNA research *15*, 227-239.

Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., Kato, T., Nakao, M., Sasamoto, S., Watanabe, A., Ono, A., Kawashima, K.*, et al.* (2008b). Genome structure of the legume, Lotus japonicus. DNA Res *15*, 227-239.

Schauser, L., Handberg, K., Sandal, N., Stiller, J., Thykjaer, T., Pajuelo, E., Nielsen, A., and Stougaard, J. (1998). Symbiotic mutants deficient in nodule establishment identified after T-DNA transformation of Lotus japonicus. Molecular & general genetics : MGG *259*, 414-423.

Schiefelbein, J., Kwak, S.-H., Wieckowski, Y., Barron, C., and Bruex, A. (2009). The gene regulatory network for root epidermal cell-type pattern formation in Arabidopsis. Journal of experimental botany *60*, 1515-1521.

Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D.L., Song, Q., Thelen, J.J., and Cheng, J. (2010a). Genome sequence of the palaeopolyploid soybean. nature *463*, 178-183.

Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D.L., Song, Q., Thelen, J.J., Cheng, J.*, et al.* (2010b). Genome sequence of the palaeopolyploid soybean. Nature *463*, 178-183.

Schmutz, J., McClean, P.E., Mamidi, S., Wu, G.A., Cannon, S.B., Grimwood, J., Jenkins, J., Shu, S., Song, Q., Chavarro, C.*, et al.* (2014). A reference genome for common bean and genome-wide analysis of dual domestications. Nat Genet *46*, 707-713.

Shiao, T.-l., and Doran, P.M. (2000). Root hairiness: effect on fluid flow and oxygen transfer in hairy root cultures. J Biotechnol *83*, 199-210.

Sieberer, B.J., Chabaud, M., Timmers, A.C., Monin, A., Fournier, J., and Barker, D.G. (2009a). A nuclear-targeted cameleon demonstrates intranuclear Ca2+ spiking in Medicago truncatula root hairs in response to rhizobial nodulation factors. Plant Physiol *151*, 1197-1206.

Sieberer, B.J., Chabaud, M., Timmers, A.C., Monin, A., Fournier, J., and Barker, D.G. (2009b). A nuclear-targeted cameleon demonstrates intranuclear Ca2+ spiking in Medicago truncatula root hairs in response to rhizobial nodulation factors. Plant Physiol *151*, 1197-1206.

Singh, S., Katzer, K., Lambert, J., Cerri, M., and Parniske, M. (2014). CYCLOPS, a DNA-binding transcriptional activator, orchestrates symbiotic root nodule development. Cell host & microbe *15*, 139-152.

Smit, P., Limpens, E., Geurts, R., Fedorova, E., Dolgikh, E., Gough, C., and Bisseling, T. (2007a). Medicago LYK3, an entry receptor in rhizobial nodulation factor signaling. Plant physiology *145*, 183-191.

Smit, P., Limpens, E., Geurts, R., Fedorova, E., Dolgikh, E., Gough, C., and Bisseling, T. (2007b). Medicago LYK3, an entry receptor in rhizobial nodulation factor signaling. Plant Physiol *145*, 183-191.

Smit, P., Raedts, J., Portyanko, V., Debellé, F., Gough, C., Bisseling, T., and Geurts, R. (2005). NSP1 of the GRAS protein family is essential for rhizobial Nod factor-induced transcription. Science *308*, 1789-1791.

Smith‐Espinoza, C., Richter, A., Salamini, F., and Bartels, D. (2003). Dissecting the response to dehydration and salt (NaCl) in the resurrection plant Craterostigma plantagineum. Plant, Cell & Environment *26*, 1307-1315.

Solis, G.P., Hoegg, M., Munderloh, C., Schrock, Y., Malaga-Trillo, E., Rivera-Milla, E., and Stuermer, C.A. (2007). Reggie/flotillin proteins are organized into stable tetramers in membrane microdomains. Biochem J *403*, 313-322.

Soltis, D.E., Albert, V.A., Leebens-Mack, J., Bell, C.D., Paterson, A.H., Zheng, C., Sankoff, D., Wall, P.K., and Soltis, P.S. (2009). Polyploidy and angiosperm diversification. American journal of botany *96*, 336-348.

Soltis, D.E., Visger, C.J., and Soltis, P.S. (2014). The polyploidy revolution then… and now: Stebbins revisited. American Journal of Botany *101*, 1057-1078.

Soyano, T., Kouchi, H., Hirota, A., and Hayashi, M. (2013). Nodule inception directly targets NF-Y subunit genes to regulate essential processes of root nodule development in Lotus japonicus. PLoS genetics *9*, e1003352.

Swatek, K.N., Lee, C.B., and Thelen, J.J. (2014). Purification of protein complexes and characterization of protein-protein interactions. Methods Mol Biol *1062*, 609-628.

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

Tadege, M., Wen, J., He, J., Tu, H., Kwak, Y., Eschstruth, A., Cayrel, A., Endre, G., Zhao, P.X., Chabaud, M.*, et al.* (2008b). Large-scale insertional mutagenesis using the Tnt1 retrotransposon in the model legume Medicago truncatula. The Plant journal : for cell and molecular biology *54*, 335-347.

Takahashi, D., Kawamura, Y., and Uemura, M. (2013a). Changes of detergent-resistant plasma membrane proteins in oat and rye during cold acclimation: association with differential freezing tolerance. Journal of proteome research *12*, 4998-5011.

Takahashi, D., Kawamura, Y., and Uemura, M. (2013b). Detergent-resistant plasma membrane proteome to elucidate microdomain functions in plant cells. Front Plant Sci *4*, 27.

Takahashi, D., Kawamura, Y., Yamashita, T., and Uemura, M. (2012). Detergent-resistant plasma membrane proteome in oat and rye: similarities and dissimilarities between two monocotyledonous plants. Journal of proteome research *11*, 1654-1665.

Takeda, S., Sugimoto, K., Otsuki, H., and Hirochika, H. (1999). A 13-bp cis-regulatory element in the LTR promoter of the tobacco retrotransposon Tto1 is involved in responsiveness to tissue culture, wounding, methyl jasmonate and fungal elicitors. The Plant journal : for cell and molecular biology *18*, 383-393.

Takehisa, H., Sato, Y., Igarashi, M., Abiko, T., Antonio, B.A., Kamatsuki, K., Minami, H., Namiki, N., Inukai, Y., Nakazono, M.*, et al.* (2012). Genome-wide transcriptome dissection of the rice root system: implications for developmental and physiological functions. The Plant journal : for cell and molecular biology *69*, 126-140.

Tanner, W., Malinsky, J., and Opekarova, M. (2011). In plant and animal cells, detergent-resistant membranes do not define functional membrane rafts. Plant Cell *23*, 1191-1193.

Tessadori, F., Chupeau, M.-C., Chupeau, Y., Knip, M., Germann, S., van Driel, R., Fransz, P., and Gaudin, V. (2007). Large-scale dissociation and sequential reassembly of pericentric heterochromatin in dedifferentiated Arabidopsis cells. Journal of Cell Science *120*, 1200-1208.

Tiley, G.P., Ane, C., and Burleigh, J.G. (2016). Evaluating and Characterizing Ancient Whole-Genome Duplications in Plants with Gene Count Data. Genome Biol Evol *8*, 1023-1037.

Titapiwatanakun, B., Blakeslee, J.J., Bandyopadhyay, A., Yang, H., Mravec, J., Sauer, M., Cheng, Y., Adamec, J., Nagashima, A., Geisler, M.*, et al.* (2009). ABCB19/PGP19 stabilises PIN1 in membrane microdomains in Arabidopsis. The Plant journal : for cell and molecular biology *57*, 27-44.

Toth, K., Stratil, T.F., Madsen, E.B., Ye, J., Popp, C., Antolin-Llovera, M., Grossmann, C., Jensen, O.N., Schussler, A., Parniske, M.*, et al.* (2012). Functional domain analysis of the Remorin protein LjSYMREM1 in Lotus japonicus. PLoS One *7*, e30817.

Tsuwamoto, R., and Harada, T. (2010). Identification of a cis-regulatory element that acts in companion cell-specific expression of AtMT2B promoter through the use of Brassica vasculature and gene-gun-mediated transient assay. Plant Cell Physiol *51*, 80-90.

VandenBosch, K.A., Bradley, D.J., Knox, J.P., Perotto, S., Butcher, G.W., and Brewin, N.J. (1989). Common components of the infection thread matrix and the intercellular space identified by immunocytochemical analysis of pea nodules and uninfected roots. The EMBO journal *8*, 335.

Vandepoele, K., Quimbaya, M., Casneuf, T., De Veylder, L., and Van de Peer, Y. (2009). Unraveling transcriptional control in Arabidopsis using cis-regulatory elements and coexpression networks. Plant Physiol *150*, 535-546.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome biology *3*, RESEARCH0034.

Vasseur, F., Pantin, F., and Vile, D. (2011). Changes in light intensity reveal a major role for carbon balance in Arabidopsis responses to high temperature. Plant, cell & environment *34*, 1563-1576.

Velasco, R., Zharkikh, A., Affourtit, J., Dhingra, A., Cestaro, A., Kalyanaraman, A., Fontana, P., Bhatnagar, S.K., Troggio, M., and Pruss, D. (2010). The genome of the domesticated apple (Malus [times] domestica Borkh.). Nature genetics *42*, 833-839.

Verdier, J., Torres-Jerez, I., Wang, M., Andriankaja, A., Allen, S.N., He, J., Tang, Y., Murray, J.D., and Udvardi, M.K. (2013). Establishment of the Lotus japonicus Gene Expression Atlas (LjGEA) and its use to explore legume seed maturation. The Plant journal : for cell and molecular biology *74*, 351-362.

Vijayakumar, P., Datta, S., and Dolan, L. (2016). ROOT HAIR DEFECTIVE SIX-LIKE4 (RSL4) promotes root hair elongation by transcriptionally regulating the expression of genes required for cell growth. New Phytol *212*, 944-953.

Wagner, S. (2012). Biological nitrogen fixation. Nature Education Knowledge *3*, 15.

Wan, J., Torres, M., Ganapathy, A., Thelen, J., DaGue, B.B., Mooney, B., Xu, D., and Stacey, G. (2005a). Proteomic analysis of soybean root hairs after infection by Bradyrhizobium japonicum. Mol Plant Microbe Interact *18*, 458-467.

Wan, J., Torres, M., Ganapathy, A., Thelen, J., DaGue, B.B., Mooney, B., Xu, D., and Stacey, G. (2005b). Proteomic Analysis of Soybean Root Hairs After Infection by Bradyrhizobium japonicum. Molecular Plant-Microbe Interactions *18*, 458-467.

Wang, Q.Q., Liu, F., Chen, X.S., Ma, X.J., Zeng, H.Q., and Yang, Z.M. (2010a). Transcriptome profiling of early developing cotton fiber by deep-sequencing reveals significantly differential expression of genes in a fuzzless/lintless mutant. Genomics *96*, 369-376.

Wang, Z., Hobson, N., Galindo, L., Zhu, S., Shi, D., McDill, J., Yang, L., Hawkins, S., Neutelings, G., and Datla, R. (2012). The genome of flax (Linum usitatissimum) assembled de novo from short shotgun sequence reads. The Plant Journal *72*, 461-473.

Wang, Z., Libault, M., Joshi, T., Valliyodan, B., Nguyen, H.T., Xu, D., Stacey, G., and Cheng, J. (2010b). SoyDB: a knowledge database of soybean transcription factors. BMC plant biology *10*, 14.

Wendel, J.F. (2000). Genome evolution in polyploids. In Plant molecular evolution (Springer), pp. 225-249.

Wendel, J.F., Jackson, S.A., Meyers, B.C., and Wing, R.A. (2016). Evolution of plant genome architecture. Genome biology *17*, 37.

Weterings, K., Schrauwen, J., Wullems, G., and Twell, D. (1995). Functional dissection of the promoter of the pollen-specific gene NTP303 reveals a novel pollen-specific, and conserved cis-regulatory element. The Plant journal : for cell and molecular biology *8*, 55-63.

Williams, E.J., and Bowles, D.J. (2004). Coexpression of neighboring genes in the genome of Arabidopsis thaliana. Genome Res *14*, 1060-1067.

Won, S.K., Lee, Y.J., Lee, H.Y., Heo, Y.K., Cho, M., and Cho, H.T. (2009). Cis-element- and transcriptome-based screening of root hair-specific genes and their functional characterization in Arabidopsis. Plant Physiol *150*, 1459-1473.

Xu, J., Xiong, W., Cao, B., Yan, T., Luo, T., Fan, T., and Luo, M. (2013). Molecular characterization and functional analysis of "fruit-weight 2.2-like" gene family in rice. Planta *238*, 643-655.

Yan, Z., Hossain, M.S., Valdes-Lopez, O., Hoang, N.T., Zhai, J., Wang, J., Libault, M., Brechenmacher, L., Findley, S., Joshi, T.*, et al.* (2015). Identification and functional characterization of soybean root hair microRNAs expressed in response to Bradyrhizobium japonicum infection. Plant Biotechnol J.

Yang, W.-C., de Blank, C., Meskiene, I., Hirt, H., Bakker, J., van Kammen, A., Franssen, H., and Bisseling, T. (1994). Rhizobium nod factors reactivate the cell cycle during infection and nodule primordium formation, but the cycle is only completed in primordium formation. The Plant Cell Online *6*, 1415-1426.

Yang, Y.W., Lai, K.N., Tai, P.Y., and Li, W.H. (1999). Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between Brassica and other angiosperm lineages. J Mol Evol *48*, 597-604.

Yano, K., Tansengco, M.L., Hio, T., Higashi, K., Murooka, Y., Imaizumi-Anraku, H., Kawaguchi, M., and Hayashi, M. (2006). New nodulation mutants responsible for infection thread development in Lotus japonicus. Molecular plant-microbe interactions *19*, 801-810.

Yano, K., Yoshida, S., Müller, J., Singh, S., Banba, M., Vickers, K., Markmann, K., White, C., Schuller, B., and Sato, S. (2008). CYCLOPS, a mediator of symbiotic intracellular accommodation. Proceedings of the National Academy of Sciences *105*, 20540-20545.

Yilmaz, A., Nishiyama, M.Y., Jr., Fuentes, B.G., Souza, G.M., Janies, D., Gray, J., and Grotewold, E. (2009). GRASSIUS: a platform for comparative regulatory genomics across the grasses. Plant Physiol *149*, 171-180.

Yokota, K., Fukai, E., Madsen, L.H., Jurkiewicz, A., Rueda, P., Radutoiu, S., Held, M., Hossain, M.S., Szczyglowski, K., and Morieri, G. (2009). Rearrangement of actin cytoskeleton mediates invasion of Lotus japonicus roots by Mesorhizobium loti. The Plant Cell *21*, 267-284.

Young, N.D., Debellé, F., Oldroyd, G.E., Geurts, R., Cannon, S.B., Udvardi, M.K., Benedito, V.A., Mayer, K.F., Gouzy, J., and Schoof, H. (2011a). The Medicago genome provides insight into the evolution of rhizobial symbioses. Nature *480*, 520-524.

Young, N.D., Debelle, F., Oldroyd, G.E., Geurts, R., Cannon, S.B., Udvardi, M.K., Benedito, V.A., Mayer, K.F., Gouzy, J., Schoof, H.*, et al.* (2011b). The Medicago genome provides insight into the evolution of rhizobial symbioses. Nature *480*, 520-524.

Zanetti, M.E., Blanco, F.A., Beker, M.P., Battaglia, M., and Aguilar, O.M. (2010). AC subunit of the plant nuclear factor NF-Y required for rhizobial infection and nodule development affects partner selection in the common bean–Rhizobium etli symbiosis. The Plant Cell Online *22*, 4142-4157.

Zhan, S., Horrocks, J., and Lukens, L.N. (2006). Islands of co-expressed neighbouring genes in Arabidopsis thaliana suggest higher-order chromosome domains. The Plant journal : for cell and molecular biology *45*, 347-357.

Zhang, C., Gong, F., Lambert, G., and Galbraith, D. (2005a). Cell type-specific characterization of nuclear DNA contents within complex tissues and organs. Plant Methods *1*, 7.

Zhang, W., Ruan, J., Ho, T.H., You, Y., Yu, T., and Quatrano, R.S. (2005b). Cis-regulatory element based targeted gene finding: genome-wide identification of abscisic acid- and abiotic stress-responsive genes in Arabidopsis thaliana. Bioinformatics *21*, 3074-3081.

Zhao, J., Morozova, N., Williams, L., Libs, L., Avivi, Y., and Grafi, G. (2001). Two Phases of Chromatin Decondensation during Dedifferentiation of Plant Cells: DISTINCTION BETWEEN COMPETENCE FOR CELL FATE SWITCH AND A COMMITMENT FOR S PHASE. Journal of Biological Chemistry *276*, 22772-22778.

Zhou, L., Wang, C., Liu, R., Han, Q., Vandeleur, R.K., Du, J., Tyerman, S., and Shou, H. (2014). Constitutive overexpression of soybean plasma membrane intrinsic protein GmPIP1; 6 confers salt tolerance. BMC plant biology *14*, 181.

Zhou, Z., Wu, Y., Yang, Y., Du, M., Zhang, X., Guo, Y., Li, C., and Zhou, J.-M. (2015). An Arabidopsis plasma membrane proton ATPase modulates JA signaling and is exploited by the Pseudomonas syringae effector protein AvrB for stomatal invasion. The Plant Cell *27*, 2032-2041.

Zou, J., Rodriguez-Zas, S., Aldea, M., Li, M., Zhu, J., Gonzalez, D.O., Vodkin, L.O., DeLucia, E., and Clough, S.J. (2005). Expression profiling soybean response to Pseudomonas syringae reveals new defense-related genes and rapid HR-specific downregulation of photosynthesis. Molecular Plant-Microbe Interactions *18*, 1161-1174.