EVOLUTION OF SIGNALING IN PLANT SYMBIOSES

EDITED BY: Jeanne Marie Harris, Ulrike Mathesius, Katharina Pawlowski and Uta Paszkowski

frontiers Research Topics

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

EVOLUTION OF SIGNALING IN PLANT SYMBIOSES

Topic Editors:
Jeanne Marie Harris, University of Vermont, United States
Ulrike Mathesius, Australian National University, Australia
Katharina Pawlowski, Stockholm University, Sweden
Uta Paszkowski, University of Cambridge, United Kingdom

Table of Contents

Editorial: Evolution of Signaling in Plant Symbioses

Jeanne Marie Harris[1]*, Katharina Pawlowski[2] and Ulrike Mathesius[3]

[1] Department of Plant Biology, University of Vermont, Burlington, VT, United States, [2] Department of Ecology, Environment and Plant Sciences, Stockholm University, Stockholm, Sweden, [3] Division of Plant Sciences, Research School of Biology, Australian National University, Canberra, ACT, Australia

Editorial on the Research Topic

Evolution of Signaling in Plant Symbioses

Plants are surrounded by microbes, but only a small number of microbes have evolved an intimate, endosymbiotic association, in which they live inside host cells. Root symbioses are important sources of nutrition for plants and microbes alike, with over 80% of all terrestrial plants forming intracellular symbioses with arbuscular mycorrhizal fungi. This Research Topic explores the evolutionary links between different plant root endosymbioses, focusing specifically on the evolution of signaling in four: the very ancient arbuscular mycorrhizae, the more recent ericoid mycorrhizae, and root nodules formed with nitrogen-fixing soil bacteria, either with *Frankia* (in actinorhizal nodulation) or rhizobia (in rhizobium-legume nodulation) (Genre and Russo, 2016). In each case, symbiosis is initiated after signal exchange between the two partners that ultimately leads to the plant hosting the microbes inside plant cells, requiring changes in plant host development and physiology (Geurts et al., 2016; MacLean et al., 2017). During establishment of the endosymbiosis, continued signal exchange between host and microbe functions to fine-tune the interaction.

Most of the plant microbiome is found on the surface of the plant, or between plant cells. Access to the exclusive niche inside cells begins with perception of microbial signals, initiating a cascade of events within the plant host allowing infection and leading to intracellular accommodation of the endosymbiont. The arbuscular mycorrhizal symbiosis is associated with the colonization of land by plants, thus recognition of mycorrhizal fungi is ancient and widespread throughout the plant kingdom. Curiously, the plant signaling pathway triggered by mycorrhizal fungi is also used by bacterial newcomers to the plant endosymbiotic world, rhizobia and *Frankia* strains. Rhizobial signal molecules are almost identical to mycorrhizal signal molecules: short chains of chitin with a lipid tail, lipochitooligosaccharides (LCOs). The *Frankia* and ericoid mycorrhizal signal molecules are unknown, but the signal of most *Frankia* strains is almost certainly not an LCO Cissoko et al..

Recent key studies have overturned our understanding of the evolution of root endosymbiosis and put this special issue in context: in particular the conclusion that the most likely scenario is a single gain of root nodule symbiosis in a clade of Rosids, followed by multiple losses of the symbiosis (Griesmann et al., 2018; van Velzen et al., 2018). This likely involved the symbiont *Frankia* as the initial symbiont of actinorhizal species, with the later emergence of rhizobia in legume symbioses (van Velzen et al., 2019). Secondly, the common use of orthologous nodulation genes across actinorhizal and legume species underlines the similarity between signaling pathways in legume and actinorhizal symbioses, which has also, at least partly, evolved from mycorrhizal symbiosis signaling (Markmann and Parniske, 2009). Contributions to this special issue have provided new insights into aspects of the evolution of root endosymbiosis by examining the

similarities at all stages of symbiosis, starting with signal perception, signal transduction, control of defense responses and eventually maintenance of the symbiosis.

Both LCOs and chito-oligosaccharides (COs) are perceived by LysM receptor kinases. The history of this gene family reveals repeated gene duplications, neofunctionalization and losses. Over this is laid the shift in function from perception of short COs to mycorrhizal signals and then to rhizobial Nod factors, analyzed in this issue by Gough et al. Further, they explore the relationship between symbiotic and immune signaling. This topic is examined in more depth by Liao et al., who demonstrate subfunctionalization after gene duplication within the LysM receptor family, specifically focusing on the evolution of the tomato chitin receptor CERK1, where chitin and mycorrhizal signal recognition is partitioned among four paralogs. LysM receptors require a co-receptor to transmit signals to the nucleus, and in plant endosymbioses this function is taken by SYMRK (SYMBIOSIS RECEPTOR KINASE). Li et al. examine this next step in the pathway to ask how SYMRK transitioned from a key position in the mycorrhizal signaling pathway to a shared function in mycorrhization and nodulation. These studies clearly highlight the evolution of signal perception from an ancient chitin and mycorrhiza perception system to the more recent perception of rhizobia by legumes.

Understanding of the *Frankia*/actinorhizal root nodule symbiosis has lagged behind the study of other root endosymbioses: the signal molecule is not an LCO, many *Frankia* species are unculturable, most plant hosts are woody perennials, and culturable *Frankia* strains (a filamentous bacterium) are difficult to grow and transform. However, actinorhizal root nodules and rhizobium-induced root nodules share an evolutionary origin in that all hosts belong to a single clade within the Rosids. Taking advantage of these parallels, comparative transcriptomics and developmental studies have provided great insights into actinorhizal nodule development and function. In legumes, flavonoids signal the rhizobial partner to induce the expression of nodulation genes and synthesis of the Nod factor signal molecule, an LCO. To explore possible new structures and roles of flavonoids in actinorhizal symbioses, Gifford et al. demonstrated synthesis of flavonoids, suggesting a functional role during actinorhizal symbiosis. This will form the basis of identifying flavonoids that may play similar signaling functions between actinorhizal hosts and *Frankia* as has been described for legume-rhizobia interactions. The diversity of flavonoids provides an additional layer of specificity to the rhizobium-legume interaction that is determined by Nod factor structure, ensuring that only specific host/microbe pairs proceed to infection.

An evolutionary assessment of the various mechanisms conferring specificity is presented by Wang et al. who also highlight recent advances in our understanding of how immune responses have been co-opted to delimit successful symbiotic responses. Even successful infections must be limited, so that they do not excessively drain plant resources. Plant peptides of the CLE (CLAVATA3/EMBRYO SURROUNDING REGION-RELATED) family have long been known to control nodule numbers as part of an internal regulatory mechanism termed autoregulation of nodulation. In their review of mechanisms of autoregulation in plant-microbe interactions more broadly, Wang et al. present the first evidence that one of the components of the autoregulation of nodulation pathway, the peptide receptor CLV2 (CLAVATA2), which also controls shoot apical meristem formation, is additionally required for autoregulation of mycorrhization in the non-legume tomato. This finding further underscores the parallels between signaling between plant host and endosymbiont in the mycorrhizal symbiosis and the rhizobium-legume symbiosis.

Once inside plant cells, endosymbionts and their plant host continue to signal each other to coordinate metabolism and development. Kereszt et al. present a broad overview of the diverse functions and evolution of plant signaling peptides in symbiotic processes in legumes and non-legumes with an emphasis on families of nodule-specific cysteine-rich (NCR) peptides (over 700 in *Medicago truncatula*!) with the ability to control bacteroid differentiation and elongation. Salgado et al. find that small defensin peptides produced by *Alnus glutinosa*, also produced by other actinorhizal plants, play parallel roles to NCR peptides in nodule symbiosis, suggesting convergent evolution or a common origin of microsymbiont control mechanisms.

In addition to the mycorrhizal origins of the plant-microbe interactions shared by all these root endosymbioses, root nodule formation involves the formation of a new organ. Mycorrhizae in general do not form a new organ, so nodule organs must have a different origin, meaning that the evolution of root nodule symbioses required both the co-option of the mycorrhizal signal response pathway as well as a second evolutionary event leading to the formation of a new organ. For decades, the model for the second event was that the ancestor of the fabids (Fabales, Fagales, Rosales and Cucurbitales) acquired a predisposition for nodulation, followed by multiple gains of the ability to nodulate (Doyle, 2016). However, as mentioned above, recent phylogenomic studies demonstrate a single origin of nodulation, followed by multiple losses, thus overturning this long-standing paradigm (Griesmann et al., 2018; van Velzen et al., 2018). This raised the importance of comparative studies including non-legume root nodule symbioses in order to distinguish between common vs. lineage-specific traits essential for nodule development and function. In this issue, five studies deal with root nodule symbioses of the Rosales. Van Zeijl et al. describe the analysis of CRISPR-Cas9 mediated knockouts of four putative nodulation-related genes in *Parasponia andersonii* (Cannabaceae, Rosales), the only non-legume able to enter a root nodule symbiosis with rhizobia, uncovering a novel role for the ethylene signaling component EIN2 in intracellular infection of *P. andersonii* nodules. Billault-Penneteau et al. describe the first step to examine another root nodule-forming member of the Rosales, establishing the sister species *Dryas octopetala* (ectomycorrhizal, non-nodulating) and *Dryas drummondii* (arbuscular mycorrhizal, nodulating) as model systems. Salgado et al. compare the nodule transcriptomes of two actinorhizal species that interact with closely related microsymbionts, *Ceanothus thyrsiflorus* (Rhamnaceae, Rosales) and *Datisca glomerata* (Datiscaceae, Cucurbitales). As mentioned

above, Gifford et al. analyze the role of flavonoids in the symbiosis of *D. glomerata*. Last but not least, Cissoko et al., using the best-examined actinorhizal model system, *Casuarina glauca* (Casuarinaceae, Fagales) show the similarity between Root Hair Deforming Factor and the *NIN* activating factor, with *NIN* encoding a central transcription factor essential for root nodule symbioses. This provides an important step toward the identification of the signal factor of Fagales-infective *Frankia*.

Continuing this approach, Battenberg et al. perform a comparative transcriptomic analysis of the same two actinorhizal nodulators analyzed by Salgado et al., *Ceanothus thyrsiflorus* (Rhamnaceae, Rosales) and *Datisca glomerata* (Datiscaceae, Cucurbitales), providing new support for the two-step process of nodule evolution, in which an initial event at the base of the Nitrogen-Fixing Clade is followed by additional selection pressure at the base of each host lineage. This extends the model of a single origin of nodulation (Griesmann et al., 2018; van Velzen et al., 2018), revealing additional insight into the origins of nodulation (Parniske, 2018).

Finally, the more recent ericoid and ectomycorrhizal symbioses, which evolved from fungal lineages distinct from those that form arbuscular mycorrhizae (Wang and Qiu, 2006), are revealing novel ways in which fungi can signal the plant.

The ericoid mycorrhizal symbiosis in particular, is unusual in that it also is able to enter cells to form an endosymbiosis but the signaling between fungus and plant is completely unknown. Casarrubia et al. identify a small, secreted protein, OmSSP1, from an ericoid mycorrhizal fungus, that shares features with hydrophobin effectors produced by ectomycorrhizal fungi, but appears to function in ericoid infection or colonization of the *Vaccinium myrtillus* root, thus identifying one of the first molecular signals between endosymbiont and host in this symbiosis.

By comparing the conserved and divergent processes by which these microbial and plant partners interact to form endosymbioses, these research articles, perspectives and reviews, provide insight into the evolutionary history of symbiotic signaling and pave the way for a deeper understanding of the commonalities and the unique features among these endosymbiotic interactions of plant roots.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

Doyle, J. J. (2016). Chasing unicorns: nodulation origins and the paradox of novelty. *Am. J. Bot.* 103, 1865–1868. doi: 10.3732/ajb.1600260

Genre, A., and Russo, G. (2016). Does a common pathway transduce symbiotic signals in plant-microbe interactions? *Front. Plant Sci.* 7:96. doi: 10.3389/fpls.2016.00096

Geurts, R., Xiao, T. T., and Reinhold-Hurek, B. (2016). What does it take to evolve a nitrogen-fixing endosymbiosis? *Trends Plant Sci.* 21, 199–208. doi: 10.1016/j.tplants.2016.01.012

Griesmann, M., Chang, Y., Liu, X., Song, Y., Haberer, G., Crook, M. B., et al. (2018). Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. *Science* 361:eaat1743. doi: 10.1126/science.aat1743

MacLean, A. M., Bravo, A., and Harrison, M. J. (2017). Plant signaling and metabolic pathways enabling arbuscular mycorrhizal symbiosis. *Plant Cell* 29, 2319–2335. doi: 10.1105/tpc.17.00555

Markmann, K., and Parniske, M. (2009). Evolution of root endosymbiosis with bacteria: how novel are nodules? *Trends Plant Sci.* 14, 77–86. doi: 10.1016/j.tplants.2008.11.009

Parniske, M. (2018). Uptake of bacteria into living plant cells, the unifying and distinct feature of the nitrogen-fixing root nodule symbiosis. *Curr. Opin. Plant Biol.* 44, 164–174. doi: 10.1016/j.pbi.2018.05.016

van Velzen, R., Doyle, J. J., and Geurts, R. (2019). A resurrected scenario: single gain and massive loss of nitrogen-fixing nodulation. *Trends Plant Sci.* 24, 49–57. doi: 10.1016/j.tplants.2018.10.005

van Velzen, R., Holmer, R., Bu, F., Rutten, L., van Zeijl, A., Liu, W., et al. (2018). Comparative genomics of the nonlegume Parasponia reveals insights into evolution of nitrogen-fixing rhizobium symbioses. *Proc. Natl. Acad. Sci. U.S.A.* 115, E4700–E4709. doi: 10.1073/pnas.1721395115

Wang, B., and Qiu, Y. L. (2006). Phylogenetic distribution and evolution of mycorrhizas in land plants. *Mycorrhiza* 16, 299–363. doi: 10.1007/s00572-005-0033-6

CRISPR/Cas9-Mediated Mutagenesis of Four Putative Symbiosis Genes of the Tropical Tree *Parasponia andersonii* Reveals Novel Phenotypes

Arjan van Zeijl, Titis A. K. Wardhani, Maryam Seifi Kalhor, Luuk Rutten, Fengjiao Bu, Marijke Hartog, Sidney Linders, Elena E. Fedorova[†], Ton Bisseling, Wouter Kohlen and Rene Geurts*

Laboratory of Molecular Biology, Department of Plant Sciences, Wageningen University & Research, Wageningen, Netherlands

Parasponia represents five fast-growing tropical tree species in the Cannabaceae and is the only plant lineage besides legumes that can establish nitrogen-fixing nodules with rhizobium. Comparative analyses between legumes and *Parasponia* allows identification of conserved genetic networks controlling this symbiosis. However, such studies are hampered due to the absence of powerful reverse genetic tools for *Parasponia*. Here, we present a fast and efficient protocol for *Agrobacterium tumefaciens*-mediated transformation and CRISPR/Cas9 mutagenesis of *Parasponia andersonii*. Using this protocol, knockout mutants are obtained within 3 months. Due to efficient micro-propagation, bi-allelic mutants can be studied in the T_0 generation, allowing phenotypic evaluation within 6 months after transformation. We mutated four genes – *PanHK4*, *PanEIN2*, *PanNSP1*, and *PanNSP2* – that control cytokinin, ethylene, or strigolactone hormonal networks and that in legumes commit essential symbiotic functions. Knockout mutants in *Panhk4* and *Panein2* displayed developmental phenotypes, namely reduced procambium activity in *Panhk4* and disturbed sex differentiation in *Panein2* mutants. The symbiotic phenotypes of *Panhk4* and *Panein2* mutant lines differ from those in legumes. In contrast, *PanNSP1* and *PanNSP2* are essential for nodule formation, a phenotype similar as reported for legumes. This indicates a conserved role for these GRAS-type transcriptional regulators in rhizobium symbiosis, illustrating the value of *Parasponia* trees as a research model for reverse genetic studies.

Keywords: *Parasponia andersonii*, rhizobium, nodule, symbiosis, CRISPR/Cas9, stable transformation

INTRODUCTION

Parasponia are tropical tree species belonging to the Cannabis family (Cannabaceae) and are known as the only non-legume plants that can establish a nitrogen-fixing endosymbiosis with rhizobium (Clason, 1936; Trinick, 1973; Akkermans et al., 1978). The *Parasponia* genus consists of five species indigenous to the Malay Archipelago and Papua New Guinea, where they grow on the

slopes of volcanic mountains (Clason, 1936; Soepadmo, 1974; Becking, 1992). *Parasponia* spp. are typical fast-growing pioneer plants, capable of covering nitrogen-poor eroded soils in a relatively short time span (Becking, 1992). Under suitable greenhouse conditions, young *Parasponia* trees can grow at speeds exceeding 45 centimeters per month, and fix up to 850 kg N ha^{-1} year^{-1} in association with rhizobium (Trinick, 1980, 1981; Trinick and Hadobas, 1989). As *Parasponia* is the only non-legume that can establish rhizobium symbiosis, it may represent a valuable model to study the core genetic networks underlying this symbiosis (Geurts et al., 2012, 2016; Behm et al., 2014).

Like legumes, *Parasponia* develops specialized root nodular organs to host the rhizobium partner. Nodules provide the rhizobium bacteria with suitable environmental conditions to convert atmospheric nitrogen into ammonium. The Cannabaceae and legume family (Fabaceae) diverged about a 100 million years ago (Wang et al., 2009), underlining that the rhizobium symbiosis in legumes and *Parasponia* evolved largely independent (Li et al., 2015). This is reflected in the distinct nodule-types found in both lineages (Behm et al., 2014). Legume nodules possess a large central zone of infected cells, which is surrounded by peripheral vascular bundles. In contrast, *Parasponia* nodules have a central vascular bundle and infected cells in the peripheral zone, giving these nodules a lateral root-like appearance. Nevertheless, initial comparative studies revealed that both symbioses are founded on conserved signaling networks. In legumes as well as *Parasponia*, root nodule formation is induced upon recognition of rhizobial secreted lipo-chitooligosaccharide (LCO) signals (Marvel et al., 1987; Op den Camp et al., 2011; Granqvist et al., 2015). Research on model legumes, like *Medicago truncatula* and *Lotus japonicus*, showed that the perception of these symbiotic signals requires a signaling cascade that has been co-opted from the much older endomycorrhizal symbiosis (Geurts et al., 2012; Oldroyd, 2013). In legumes, activation of the LCO signaling network results in a massive transcriptional reprogramming, requiring among others the GRAS-type transcriptional regulators NODULATION SIGNALLING PATHWAY 1 (NSP1) and NSP2 and the cytokinin receptor MtCRE1/LjLHK1 (Kaló et al., 2005; Smit et al., 2005; Gonzalez-Rizzo et al., 2006; Heckmann et al., 2006; Murray et al., 2007; Tirichine et al., 2007; Plet et al., 2011). Subsequent nodule formation is tightly controlled by regulatory feedback loops, including negative regulation by ethylene signaling (Penmetsa et al., 2008; Miyata et al., 2013; van Zeijl et al., 2015b).

A reference quality genome sequence for *Parasponia andersonii* and draft genome sequences of two additional *Parasponia* species have been generated (van Velzen et al., 2017). Mining these genomes uncovered ∼1,800 putative symbiosis genes, of which 100s are close homologs of legume symbiosis genes (van Velzen et al., 2017). Initial reverse genetic studies in *P. andersonii*, using a transient *Agrobacterium rhizogenes*-based root transformation system, revealed that at least two genes – *NOD FACTOR PERCEPTION 1* (*PanNFP1*) and *CALCIUM AND CALMODULIN-DEPENDENT PROTEIN KINASE* (*PanCCaMK*) – commit conserved functions in the *Parasponia*

and legume LCO signaling pathways (Op den Camp et al., 2011). We argue that a more comprehensive comparative analysis between legumes and *Parasponia* will allow identification of conserved genetic networks that are essential to establish symbiosis with rhizobium. However, to use *Parasponia* as an effective research model – alongside the legume models *M. truncatula* and *L. japonicus* – efficient transformation and genome editing tools are required.

Here, we exploit an efficient *in vitro* micro-propagation system available for *P. andersonii* to establish stable transformation and CRISPR/Cas9-mediated mutagenesis for this species (Davey et al., 1993; Webster et al., 1995; Cao et al., 2012). We show that using *Agrobacterium tumefaciens*-mediated transformation, stable transgenic lines of *P. andersonii* can be obtained in ∼3–4 months. Additionally, we show that *P. andersonii* is amenable to targeted mutagenesis using the CRISPR/Cas9 system. As ∼40% of the resulting T$_0$ lines harbor bi-allelic mutations, these can be phenotyped upon *in vitro* propagation. As proof of concept, we mutated four genes in *P. andersonii* that in legumes control hormonal pathways as well as commit symbiotic functions. These include: the GRAS-type transcriptional regulators *NSP1* and *NSP2* that are essential for nodule organogenesis (Kaló et al., 2005; Smit et al., 2005; Heckmann et al., 2006) and control strigolactone biosynthesis by mediating *DWARF27* (*D27*) expression (Liu et al., 2011; van Zeijl et al., 2015a); the cytokinin receptor *HISTIDINE KINASE 4* (*HK4*) that in legumes is essential for nodule organogenesis (Gonzalez-Rizzo et al., 2006; Murray et al., 2007; Plet et al., 2011); and the ethylene signaling hub *ETHYLENE INSENSITIVE 2* (*EIN2*) that is a negative regulator of nodulation in legumes (Penmetsa and Cook, 1997; Penmetsa et al., 2008; Miyata et al., 2013).

MATERIALS AND METHODS

Plant Materials and Growth Conditions

All experiments were conducted using *P. andersonii* WU1 or offspring thereof (Op den Camp et al., 2011; van Velzen et al., 2017). *P. andersonii* trees were grown in a conditioned greenhouse at 28°C, 85% humidity and a 16/8 h day/night regime. For *in vitro* culturing, *P. andersonii* was grown in an Elbanton growth cabinet at 28°C, 16/8 h day/night. Growth of young *P. andersonii* plantlets for nodulation assays or qRT-PCR analysis was performed in 1 L crystal-clear polypropelene containers equipped with a gas exchange filter (OS140BOX, Duchefa Biochemie, Netherlands). Pots were half-filled with agraperlite (Maasmond-Westland, Netherlands) and watered with modified EKM medium [3 mM MES ($C_6H_{13}NO_4$) pH 6.6, 2.08 mM MgSO$_4$, 0.88 mM KH$_2$PO$_4$, 2.07 mM K$_2$HPO$_4$, 1.45 mM CaCl$_2$, 0.70 mM Na$_2$SO$_4$, 0.375 mM NH$_4$NO$_3$, 15 μM Fe-citrate, 6.6 μM MnSO$_4$, 1.5 μM ZnSO$_4$, 1.6 μM CuSO$_4$, 4 μM H$_3$BO$_3$, 4.1 μM Na$_2$MoO$_4$] (Becking, 1983) and placed in a climate room set at 28°C, 16/8 h day/night. For nodulation assays, EKM medium was inoculated with *Mesorhizobium plurifarium* BOR2 (OD$_{600}$ = 0.025) (van Velzen et al., 2017).

Vectors and Constructs

For CRISPR/Cas9-mediated mutagenesis, binary transformation constructs were created using Golden Gate assembly (Engler et al., 2009). For an overview of all Golden Gate clones used in this study, see Supplementary Table 1. sgRNAs were designed based on the principles described in Doench et al. (2014) and PCR amplified using specific forward primers and a universal reverse primer (Supplementary Table 2), using Addgene plasmid # 46966 as template (Nekrasov et al., 2013). These were cloned behind the AtU6p small RNA promoter and inserted behind the neomycin phosphotransferase II gene (*NPTII*) and an *Arabidopsis thaliana* codon-optimized variant of Cas9 (Fauser et al., 2014) fused to an N-terminal nuclear localization signal and driven by the 35S promoter (Supplementary Table 1). As negative control, a binary vector was created containing only the NPTII- and NLS-Cas9-encoding sequences (Supplementary Table 1). To setup *P. andersonii* stable transformation, vector pKGWFS7-RR was used (Karimi et al., 2002).

Phylogenetic Reconstruction

Protein sequences of *Glycine max* (Wm82.a2.v1) (Schmutz et al., 2010), *M. truncatula* (Mt4.0v1) (Young et al., 2011; Tang et al., 2014) and *Populus trichocarpa* (v3.0) (Tuskan et al., 2006) were obtained through Phytozome 10[1]. Protein sequences of *P. andersonii* (PanWU01x14_asm01_ann01) and *Trema orientalis* (TorRG33x02_asm01_ann01) were obtained from www.parasponia.org (van Velzen et al., 2017). These sequences were mined using sequences from *A. thaliana* (TAIR10[2]) (Lamesch et al., 2012) and *M. truncatula*. Protein sequences were aligned using MAFFT v7.017 (Katoh et al., 2002) implemented in Geneious 8.1.9 (Biomatters, New Zealand), using default parameter settings. Approximately-maximum-likelihood phylogenetic trees were constructed using FastTree (Price et al., 2009) implemented in Geneious 8.1.9. Mid-point rooting was applied for better tree visualization using FigTree v1.4.2[3].

Plant Transformation

Stable transformation of *P. andersonii* was performed using *A. tumefaciens* strain AGL1 (Lazo et al., 1991). *A. tumefaciens* was grown for 2 days on agar-solidified LB medium containing appropriate antibiotics. For each *P. andersonii* transformation, two Petri dishes (Ø 9 cm) of *A. tumefaciens* were used. Bacteria were scraped from plate and resuspended in 25 ml of infiltration medium [SH10 (Supplementary Table 3), 20 mg/l acetosyringone (Sigma, United States), 0.001% (v/v) Silwet L-77[4]]. *P. andersonii* tissue explants used for transformation were harvested from mature trees grown under greenhouse conditions and sterilized in 2% commercial bleach for 15 min. Tissue explants were cut at both ends inside the *A. tumefaciens* suspension, creating fresh wound surfaces, and kept inside the suspension for about 20 min. Subsequently, excess liquid was removed from tissue explants using sterilized filter paper and explants were placed on co-cultivation medium [Root-inducing

medium (Supplementary Table 3), 20 mg/l acetosyringone (Sigma, United States)]. Plates were incubated for 2 days at 21°C in darkness. After 2 days, tissue explants were washed three times using SH10 (Supplementary Table 3) and subsequently dried using filter paper. Tissue explants were placed on root-inducing medium containing 50 mg/l kanamycin and 300 mg/l cefotaxime and incubated at 28°C, 16/8 h day/night. Nine days after transformation, tissue explants were transferred to propagation medium (Supplementary Table 3) containing 50 mg/l kanamycin and 300 mg/l cefotaxime. Plates were refreshed every other week. When regenerative calli reached ∼2 mm in size they were separated from tissue explants to stimulate shoot formation. A single shoot was selected per tissue explant. These shoots were propagated on propagation medium (Supplementary Table 3), as previously described (Cao et al., 2012). Rooted plantlets were generated by placing individual shoots on root-inducing medium (Supplementary Table 3) (Cao et al., 2012).

Characterization of Transgenic Lines

For T-DNA copy number estimates based on qPCR analysis, genomic DNA was isolated using the DNeasy Plant Mini Kit (Qiagen, Germany). qPCR was set up in a 10 μl reaction system with 2x iQ SYBR Green Super-mix (Bio-Rad, United States) and 5 ng template DNA. The experimental setup and procedure were executed on a CFX Connect optical cycler, according to the manufacturer's protocol (Bio-Rad, United States). T-DNA copy number was estimated using two primer pairs amplifying part of the T-DNA and two primer pairs amplifying single copy *P. andersonii* genes (*PanAGT1* and PanWU01x14_asm01_ann01_338920) that were selected based on a study by Duarte et al. (2010). Primer sequences are listed in Supplementary Table 2. Data analysis was performed using CFX Manager 3.0 software (Bio-Rad, United States). For T-DNA copy number estimates based on Southern blotting, genomic DNA was separately digested with *XbaI*, *HindIII*, and *EcoRI*. Blots were hybridized with a 516 bp α-32P-labeled probe corresponding to part of the *NPTII* gene that was amplified using primers nptII_Fw and nptII_Rv listed in Supplementary Table 2.

Genotyping of transgenic lines was performed using the Phire Plant Direct PCR Kit (Thermo Scientific, United States) and gene specific primers listed in Supplementary Table 2. Ploidy estimates of transgenic lines were determined by FACS as described by van Velzen et al. (2017).

To determine ethylene sensitivity of *Panein2* mutants, tips of young branches of 4 months-old trees were covered with 1 L plastic bags and injected with 1 ml of pure ethylene gas. After 3 days, bags were removed and leaf abscission examined. Total number of leaves on treated branches varied from 6 to 18.

Microtome Sectioning

Stem cross-sections were made from the primary stem, 5 cm below the apical meristem, of 2 month-old trees. Shoot tissue was fixed in 5% glutaraldehyde and embedded in Technovit 7100 (Heraeus-Kulzer, Germany), according to the manufacturer's protocol. Semi-thin (7 μm) sections were cut using a microtome (Reichert-Jung, Leica Microsystems, Netherlands) and stained with 0.05% Toluidine Blue O. Images were taken using a Leica

---

[1] http://phytozome.jgi.doe.gov/

[2] www.arabidopsis.org

[3] http://tree.bio.ed.ac.uk/software/figtree/

DM5500B microscope equipped with a DFC425C camera (Leica Microsystems, Germany). Average procambium cell number was quantified by averaging the number of cells within 25–40 cell files for each of the biological replicates.

Nodule tissue fixation and embedding was performed as previously described (Fedorova et al., 1999). Semi-thin (0.6 μm) sections were cut using a Leica Ultracut microtome (Leica Microsystems, Germany) and photographed as described above.

RNA Isolation and qRT-PCR Analysis

RNA was isolated from snap-frozen root tips (∼2–3 cm) as described by van Velzen et al. (2017). cDNA was prepared from 1 μg of total RNA using the i-script cDNA synthesis kit (Bio-Rad, United States), following the manufacturer's instructions. RT-qPCR was set up as described above. Normalization was performed based on two stably expressed reference genes [*UNKNOWN 2* (*PanUNK2*) and *ELONGATION FACTOR 1α* (*PanEF1α*)], chosen based on previous study (Czechowski et al., 2005; Bansal et al., 2015). All primer sequences are listed in Supplementary Table 2.

Statistical Analysis

Statistical differences were determined based on one-way ANOVA and Tukey *post hoc* tests. Statistical analyses were performed using IBM SPSS Statistics 23.0 (IBM, United States).

RESULTS

Agrobacterium tumefaciens-Mediated Transformation of *Parasponia*

To establish a protocol for stable transformation of *P. andersonii*, we first determined the most optimal conditions for regeneration of non-transgenic tissue. We compared regeneration efficiencies of nine tissue explant types in combination with 11 different media, including the propagation and root-inducing media previously used for *P. andersonii* (Supplementary Tables 4, 5) (Op den Camp et al., 2011; Cao et al., 2012). This revealed that young stem pieces and petioles placed on original propagation medium regenerate plantlets most efficiently (Supplementary Table 4). Next, we questioned whether stem pieces and petioles could be transformed efficiently using *A. tumefaciens*. To this end, we used *A. tumefaciens* AGL1 carrying a binary transformation vector containing in its T-DNA the kanamycin resistance gene *NPTII* and the red fluorescent protein *DsRED1*. Co-cultivation of *A. tumefaciens* and *P. andersonii* stem or petiole explants was conducted in darkness for 2 days at 21°C to promote T-DNA transfer (Cao et al., 2012). Afterward, tissue explants were placed on selective medium and incubated at 28°C in the light. These latter conditions are most favorable for *P. andersonii* regeneration (Cao et al., 2012). From day 8 onwards, DsRED1-fluorescent cells could be observed near the wound surface indicating a successful transfer of the T-DNA.

Recent research on *A. thaliana* showed that acquisition of pluripotency requires activation of a root developmental
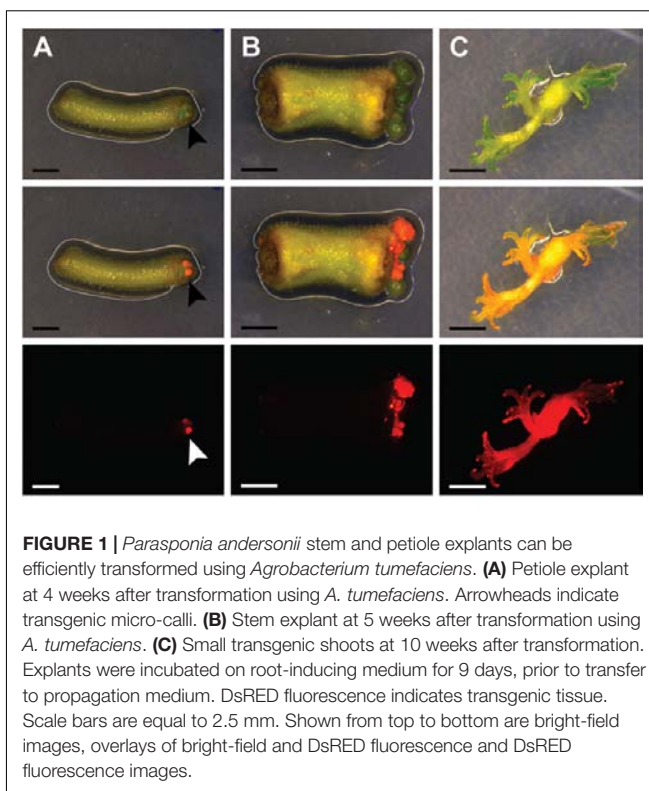
FIGURE 1 | *Parasponia andersonii* stem and petiole explants can be efficiently transformed using *Agrobacterium tumefaciens*. **(A)** Petiole explant at 4 weeks after transformation using *A. tumefaciens*. Arrowheads indicate transgenic micro-calli. **(B)** Stem explant at 5 weeks after transformation using *A. tumefaciens*. **(C)** Small transgenic shoots at 10 weeks after transformation. Explants were incubated on root-inducing medium for 9 days, prior to transfer to propagation medium. DsRED fluorescence indicates transgenic tissue. Scale bars are equal to 2.5 mm. Shown from top to bottom are bright-field images, overlays of bright-field and DsRED fluorescence and DsRED fluorescence images.

program (Kareem et al., 2015). We tested whether an initial culturing period on root-inducing medium further improves the transformation efficiency of *P. andersonii*. This showed to be the case (Supplementary Table 6). About half of the explants formed regenerative calli at 4 weeks after co-cultivation (**Figure 1A**). When 2 mm in size, transgenic calli were separated from tissue explants, which stimulated shoot formation (**Figures 1B,C**). Two to three months after the start of transformation, a single shoot was selected from each explant to ensure that the transgenic lines represent independent transformation events. These shoots can be genotyped and vegetatively propagated (Supplementary Figure 1). The latter allows clonal multiplication of individual transgenic lines in a period of ∼4–6 weeks, which means that phenotyping assays could be initiated at ∼4 months after the start of transformation.

To characterize the resulting transgenic *P. andersonii* lines at the molecular level, we selected – based on red fluorescence – 20 independent transformants for further analyses. PCR reactions using primers amplifying a sequence near the right T-DNA border indicated complete T-DNA integration in 19 out of 20 lines (Supplementary Table 7). To determine whether the transformation procedure might affect ploidy level of the regenerated transgenic lines, we estimated genome size based on flow cytometry. This showed no effect of the transformation procedure on the genome size of transgenic lines (Supplementary Table 7). To estimate the number of T-DNA integrations, we used quantitative RT-PCR (qRT-PCR) as well as Southern blotting. This showed an overall

low T-DNA copy number, varying between one and three integrations per line (Supplementary Table 7). We selected three transgenic lines with a single T-DNA integration to examine T-DNA stability. In greenhouse-grown trees as well as *in vitro* propagated material, DsRED1 fluorescence could still be observed at 6–12 months after transgenic lines were selected (Supplementary Figures 1, 2). This indicates that *trans*-genes remain stably integrated into the *P. andersonii* genome and actively transcribed, even after multiple rounds of vegetative propagation. Taken together, the protocol described above allows generating *A. tumefaciens*-transformed *P. andersonii* plantlets within 3 months, which can be phenotyped upon vegetative propagation.

Parasponia Is Amenable to CRISPR/Cas9-Mediated Mutagenesis

To test whether CRISPR/Cas9 could be used for targeted mutagenesis in *P. andersonii*, we aimed at mutating the *P. andersonii* putative orthologs of *EIN2*, *MtCRE1/LjLHK1*, *NSP1,* and *NSP2*. These genes were selected, because they control legume root nodule formation as well as commit essential non-symbiotic functions in hormone homeostasis. Putative orthologs of all four genes were previously identified from the *P. andersonii* genome and named *PanEIN2*, *PanHK4*, *PanNSP1*, and *PanNSP2*, respectively (van Velzen et al., 2017). Phylogenetic reconstruction based on protein sequences confirmed that these represent the most likely orthologs of legume symbiotic genes (Supplementary Figures 3–6). To mutate *PanEIN2*, *PanHK4*, *PanNSP1* and *PanNSP2*, three single guide RNAs (sgRNAs) targeting *PanHK4* and *PanNSP2* and single sgRNAs targeting *PanEIN2* and *PanNSP1* were placed under an *A. thaliana AtU6* small RNA promoter (Nekrasov et al., 2013). These were cloned into a binary transformation vector containing the *NPTII* kanamycin resistance gene as well as a Cas9-encoding sequence fused to an N-terminal nuclear-localization signal and driven by the CaMV 35S promoter (Engler et al., 2014; Fauser et al., 2014). The resulting constructs as well a control construct containing only the NPTII- and Cas9-encoding sequences were transformed to *P. andersonii* using the method described above. For all constructs, transgenic shoots were obtained, although in case of the construct targeting *PanHK4* regeneration took considerably longer (up to 6 months). Genotyping of regenerated shoots showed that >85% contained the *Cas9* gene, indicating successful

transformation. Potential mutations at any of the target sites were identified through PCR amplification and subsequent sequencing of the PCR product. This revealed mutations at the target site in about half of the transgenic shoots examined, of which the majority were bi-allelic (**Table 1**). Most mutations represent small insertions and deletions but also larger deletions and inversions were identified, some of which occur in between two target sites (Supplementary Figures 7–10). In case of *PanHK4*, most mutants contained small in-frame deletions of 3 or 6 bp that most likely do not disrupt protein function. In fact, only two bi-allelic knockout mutants could be identified (Supplementary Figure 8). For the remaining three constructs, multiple bi-allelic knockout mutants were identified of which three individuals were selected for further studies (for an overview of mutant alleles see Supplementary Figures 7–10).

For phenotypic evaluation, *P. andersonii* T_0 transgenic lines are propagated vegetatively. Therefore, we first evaluated whether any of the mutant lines might be chimeric. To this end, tissue samples were taken from at least three different positions and genotyped for the corresponding target mutation. For each of the mutant lines, except *Pannsp2–9*, the same mutations were retrieved, suggesting that genome editing occurred soon after T-DNA integration. In case of *Pannsp2–9*, chimeric mutations were detected at the first of three target sites (Supplementary Figure 10C). However, the nature of the mutations at the second and third target site prevent that gene function could be restored in this line. Therefore, all 11 mutants are suitable for phenotypic evaluation. This proves that CRISPR/Cas9 can be used to efficiently mutagenize *P. andersonii* in the T_0 generation.

Non-symbiotic Phenotypes in *Parasponia ein2, hk4, nsp1,* and *nsp2* Mutant Lines

To characterize the resulting *Panein2*, *Panhk4*, *Pannsp1* and *Pannsp2* mutant lines, we studied their non-symbiotic phenotypes. *PanEIN2* putatively encodes a central component of the ethylene signaling pathway and therefore *Panein2* mutants are expected to be ethylene insensitive. One phenotype triggered in response to ethylene treatment is abscission of leaves and flowers, as shown in amongst others common bean (*Phaseolus vulgaris*), cotton (*Gossypium hirsutum*), and citrus (*Citrus clementina*) (Jackson and Osborne, 1970; Brown, 1997;

TABLE 1 | Mutation frequency in CRISPR/Cas9 transgenic lines.

Target gene	No. of sgRNA's	No. of lines	Non-mutated[a]	Mutated		
				Bi-allelic	Heterozygous	Unknown[b]
NSP1	1	29	15 (51.7%)	11 (37.9%)	0	3 (10.3%)
NSP2	3	29	13 (44.8%)	10 (34.5%)	3 (10.3%)	3 (10.3%)
EIN2	1	9	1 (11.1%)	6 (66.7%)	2 (22.2%)	0
HK4	3	26	13 (50.0%)	12 (46.2%)	1 (3.8%)	0
Total		**93**	**42 (45.2%)**	**39 (41.9%)**	**6 (6.5%)**	**6 (%)**

[a]This includes an unknown number of individuals that are not transgenic. [b]Sequencing of the PCR product indicates that plants are mutated, but the exact mutation and zygosity were not determined.
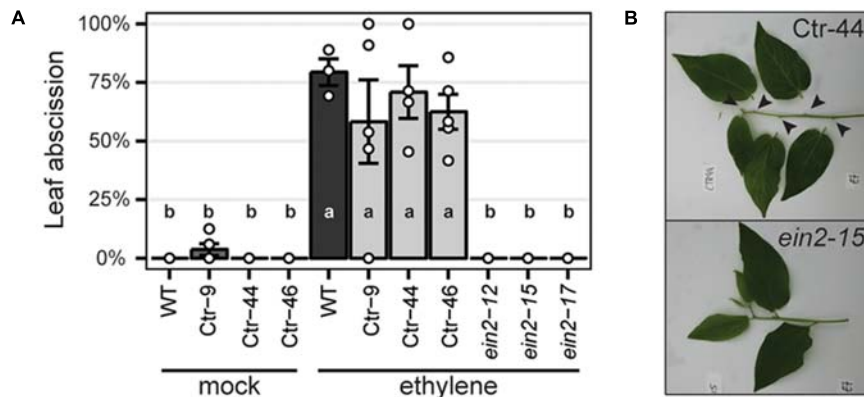
FIGURE 2 | *Panein2* mutants are insensitive to ethylene treatment. **(A)** Percentage of abscised leaves after mock or ethylene treatment. Data represent means of 3–5 biological replicates ± SEM. Dots represent measurement values of biological repeats. Different letters indicate statistical significance ($p < 0.05$) as determined by ANOVA in combination with Tukey *post hoc* test. **(B)** Representative images showing abscission of leaves on a transgenic control line (Ctr-44), but not on a *Panein2* mutant. Abscission points are indicated by arrowheads.

Agustí et al., 2009). We exploited this phenotype to assess ethylene sensitivity of *Panein2* mutants. To this end, the tips of young shoot branches of greenhouse grown trees were exposed to ethylene gas. Within 3 days, ethylene triggered abscission of ∼65% of treated leaves on wild-type (WT) *P. andersonii* as well as control transgenic lines (**Figure 2**). In contrast, leaf abscission was not observed on *Panein2* mutant trees (**Figure 2**). This demonstrates that *Panein2* mutants are indeed ethylene insensitive.

Inspection of *Panein2* mutant trees revealed an additional non-symbiotic phenotype. These trees form bisexual flowers containing both male and female reproductive organs (**Figures 3A–C**). In contrast, WT *P. andersonii* trees form unisexual flowers that contain either stamens or carpels (Becking, 1992) (**Figures 3D,E**). This suggests that ethylene is involved in the regulation of *Parasponia* sex type.

Cytokinins are important regulators of cambial activity, as shown in *A. thaliana* and poplar (*Populus tremula* x *tremuloides*) (Matsumoto-Kitano et al., 2008; Nieminen et al., 2008; Bhalerao and Fischer, 2017). To determine whether reduced cytokinin sensitivity in *Panhk4* mutant lines affects the activity of the procambium, we sectioned young primary stems, 5 cm below the apical meristem. This showed that procambium activity is reduced in *Panhk4* mutant lines compared to transgenic controls (**Figure 4**). Therefore, we conclude that PanHK4-mediated cytokinin signaling is required for regulation of *P. andersonii* secondary growth.

Expression studies in *M. truncatula* previously identified a set of genes downregulated in roots of *Mtnsp1* and *Mtnsp2* mutants (Liu et al., 2011). Among these are *DWARF27* (*MtD27*; Medtr1g471050) and *MORE AXILLARY BRANCHING 1* (*MtMAX1*; Medtr3g104560) that are putatively involved in strigolactone biosynthesis (Liu et al., 2011; Cardoso et al., 2014; Zhang et al., 2014; van Zeijl et al., 2015a). We identified putative *P. andersonii* orthologs of these genes (Supplementary Figures 11, 12) and compared their expression levels in young root segments of three *Pannsp1*, *Pannsp2* and control plants by qRT-PCR. This
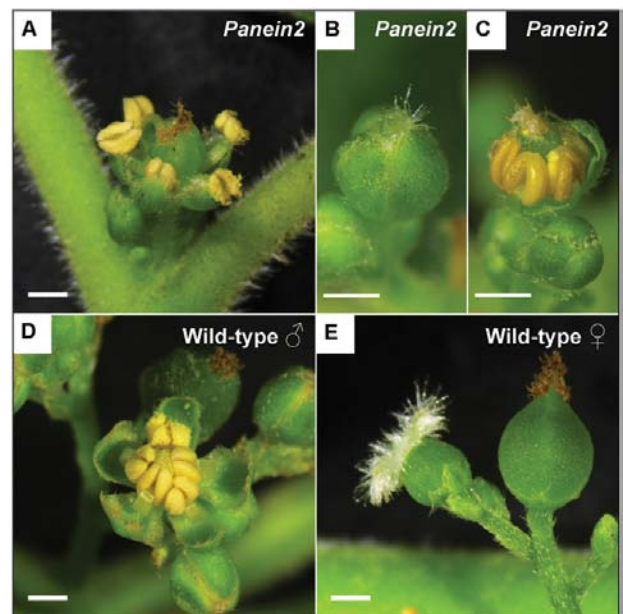
FIGURE 3 | *Panein2* mutants form bisexual flowers. **(A)** Mature *Panein2* flower. Note the presence of both stamen and a carpel inside *Panein2* flowers. **(B)** Immature *Panein2* flower. Note the presence of stigmata, indicating presence of a carpel inside the flower. **(C)** Immature *Panein2* flower of which sepals have been removed to show the presence of stamen. **(D)** Mature wild-type (WT) male flower. **(E)** Mature WT female flowers. Left: young female flower. Right: older female flower. Scale bars are equal to 1 mm.

showed that expression of *PanD27* and *PanMAX1* is reduced in roots of *Pannsp1* and *Pannsp2* mutant lines (**Figure 5**). We noted that *Pannsp1* mutant lines differ in the level of *PanD27* and *PanMAX1* expression. Both genes have an intermediate expression level in *Pannsp1–6* and *Pannsp1–13*, compared to *Pannsp1–39* and *Pannsp2* mutants (**Figure 5**). The three *Pannsp1* mutant lines differ from each other in the type of mutations that
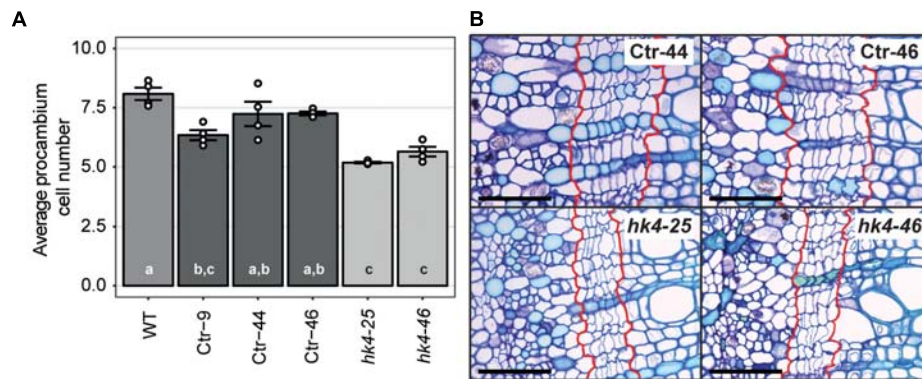
FIGURE 4 | *Panhk4* mutants display reduced procambium activity. **(A)** Average number of procambium cells in WT, transgenic control (Ctr) and *Panhk4* mutant stems. Data represent means of four biological replicates ± SEM. Dots represent measurement values of biological repeats. Different letters indicate statistical significance ($p < 0.05$) as determined by ANOVA in combination with Tukey *post hoc* test. **(B)** Representative images of stem cross-sections of control transgenic (Ctr) and *Panhk4* mutant lines. The location of the procambium is indicated by red lines. Scale bars represent 50 μm.

were created. *Pannsp1–6* and *Pannsp1–13* contain a 1 bp insertion and 5 bp deletion close to the 5′-end of the coding region, respectively. These mutations are immediately followed by a second in-frame ATG that in WT PanNSP1 encodes a methionine at position 16. In contrast, *Pannsp1–39* contains a large 232 bp deletion that removes this in-frame ATG (see Supplementary Figure 9). Together, this suggests that *Pannsp1–6* and *Pannsp1–13* might represent weak alleles. Overall, these data suggest that regulation of *D27* and *MAX1* expression by NSP1 and NSP2 is conserved between *M. truncatula* and *P. andersonii*.

Taken together, we showed that EIN2, HK4, NSP1, and NSP2 in *P. andersonii* commit non-symbiotic functions in hormonal homeostasis. These functions are in line with what is described for other plant species, suggesting that the generated *P. andersonii* lines represent true mutants.

Nodulation Phenotypes of *Parasponia Panein2* and *Panhk4* Mutants Differ From Their Legume Counterparts

To determine whether *PanEIN2*, *PanHK4*, *PanNSP1,* and *PanNSP2* perform similar functions during nodule formation as their legume orthologs, *P. andersonii* mutant plantlets were inoculated with *Mesorhizobium plurifarium* BOR2 (van Velzen et al., 2017). Nodulation phenotypes were examined 1 month after inoculation.

The strong *Pannsp1–39* mutant allele and all three *Pannsp2* mutant lines are unable to form root nodules (**Figure 6A** and Supplementary Figure 13). This is similar as described for *M. truncatula, L. japonicus,* and *Pisum sativum nsp1* and *nsp2* mutants (Kaló et al., 2005; Smit et al., 2005; Heckmann et al., 2006; Shtark et al., 2016). In contrast, the weak *Pannsp1* alleles *Pannsp1–6* and *Pannsp1–13* could be nodulated similar as WT or control transgenic plants (**Figure 6A** and Supplementary Figure 13), suggesting that residual PanNSP1 activity is sufficient to support root nodule formation. Overall, these data show that NSP1 and NSP2 functioning is essential for root nodule formation in *Parasponia*.
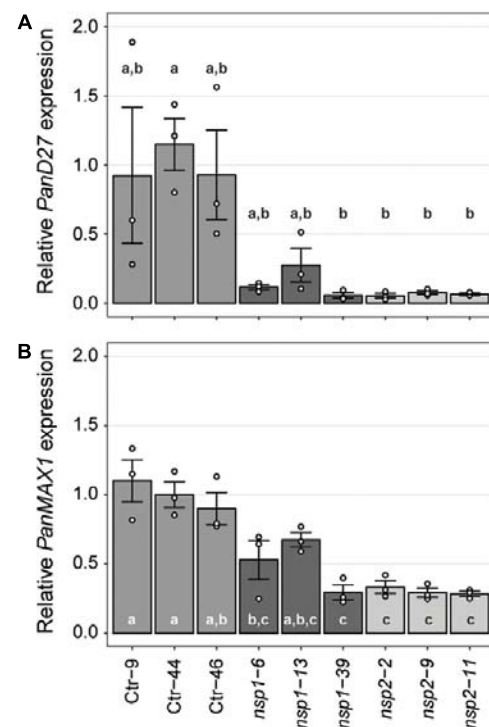
FIGURE 5 | Expression of *PanD27* and *PanMAX1* is reduced in *Pannsp1* and *Pannsp2* mutant roots. Relative expression of *PanD27* **(A)** and *PanMAX1* **(B)** in roots of transgenic control (Ctr) and *Pannsp1* and *Pannsp2* mutant lines. Data represent means of three biological replicates ± SEM. Dots represent measurement values of biological repeats. Different letters indicate statistical significance ($p < 0.05$) as determined by ANOVA in combination with Tukey *post hoc* test.

Analysis of the nodulation phenotype of *P. andersonii Panhk4* mutants showed that PanHK4 is not required for root nodule formation. Both *Panhk4* mutant lines formed a similar amount of nodules as WT and transgenic controls (**Figure 6B**
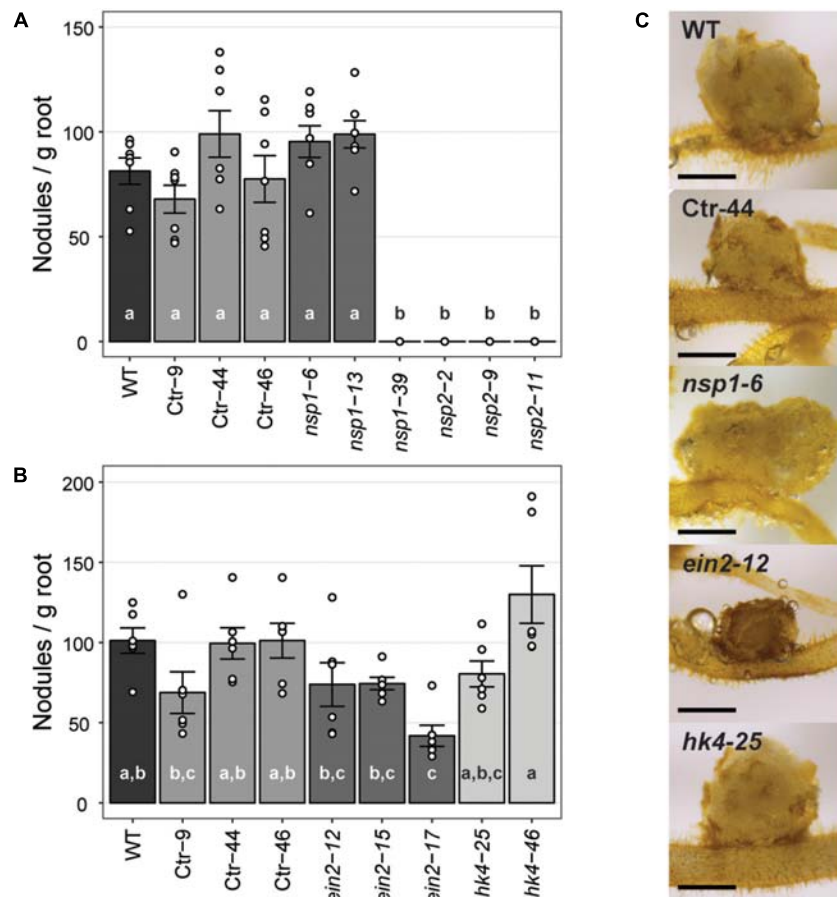
FIGURE 6 | Nodule formation on *P. andersonii* CRISPR/Cas9 mutant lines. **(A)** Nodule number per gram root fresh weight on WT, transgenic control (Ctr) and *Pannsp1* and *Pannsp2* mutant lines. Nodule number was determined at 1 month after inoculation with *Mesorhizobium plurifarium* BOR2. **(B)** Nodule number per gram root fresh weight on WT, transgenic control (Ctr) and *Panein2* and *Panhk4* mutant lines. Nodule number was determined at 1 month after inoculation with *Mesorhizobium plurifarium* BOR2. **(C)** Representative images of 1 month-old nodules. Note the dark color of *Panein2* nodules. Scale bars are equal to 0.5 mm. Data in **(A,B)** represent means of 5–7 biological replicates ± SEM. Dots represent measurement values of biological repeats. Different letters indicate statistical significance (*p* < 0.05) as determined by ANOVA in combination with Tukey *post hoc* test. Data on nodule number and root weight are shown in Supplementary Figure 13.

and Supplementary Figure 13). This is different from the corresponding legume mutants – *M. truncatula Mtcre1* and *L. japonicus Ljlhk1* – that are generally not forming root nodules (Murray et al., 2007; Plet et al., 2011).

The phenotype of *P. andersonii Panein2* mutants also differs from that of legume mutants. *M. truncatula ein2* mutants – as well as *L. japonicus* plants in which both EIN2-encoding genes have been silenced – form more nodules than WT, which are clustered in distinct zones along the root (Penmetsa and Cook, 1997; Miyata et al., 2013). *Panein2* mutants do not form such nodule clusters and nodule number is not higher than WT (**Figure 6B** and Supplementary Figure 13). However, nodules formed on *Panein2* mutant plants are smaller and dark colored when compared to nodules of control plants (**Figure 6C**). This suggests impaired nodule development in *P. andersonii ein2* mutants.

To determine the cytoarchitecture of *Panein2*, *Panhk4,* and *Pannsp1–6/Pannsp1–13* mutant nodules, we sectioned ∼10

nodules for each mutant line and studied these by light microscopy. Wild-type *P. andersonii* nodules harbor an apical meristem, followed by several cell layers that contain infection threads (**Figure 7A**) (Op den Camp et al., 2012). Below this infection zone, 2–3 cell layers are present that display vacuolar fragmentation and increase in size compared to non-infected cells (**Figure 7B**). These cells are followed by cells that are filled with fixation threads (**Figures 7B,C**). The general cytoarchitecture of *Panhk4* and *Pannsp1–6/Pannsp1–13* mutant nodules does not differ from that of WT or transgenic control nodules (**Figures 7A,D,E**), suggesting that these are functional. In contrast, in *Panein2* mutant nodules intracellular infection is hampered (**Figures 7F–I**). Most (>75%) *Panein2* mutant nodules harbor only infection threads as well as large apoplastic colonies (**Figure 7I**). Some mutant nodules, harbor cells that contain fixation threads. However, even in the best nodules, fixation thread formation is severely delayed and many cells in the fixation zone still show vacuolar fragmentation (**Figure 7I**). This shows
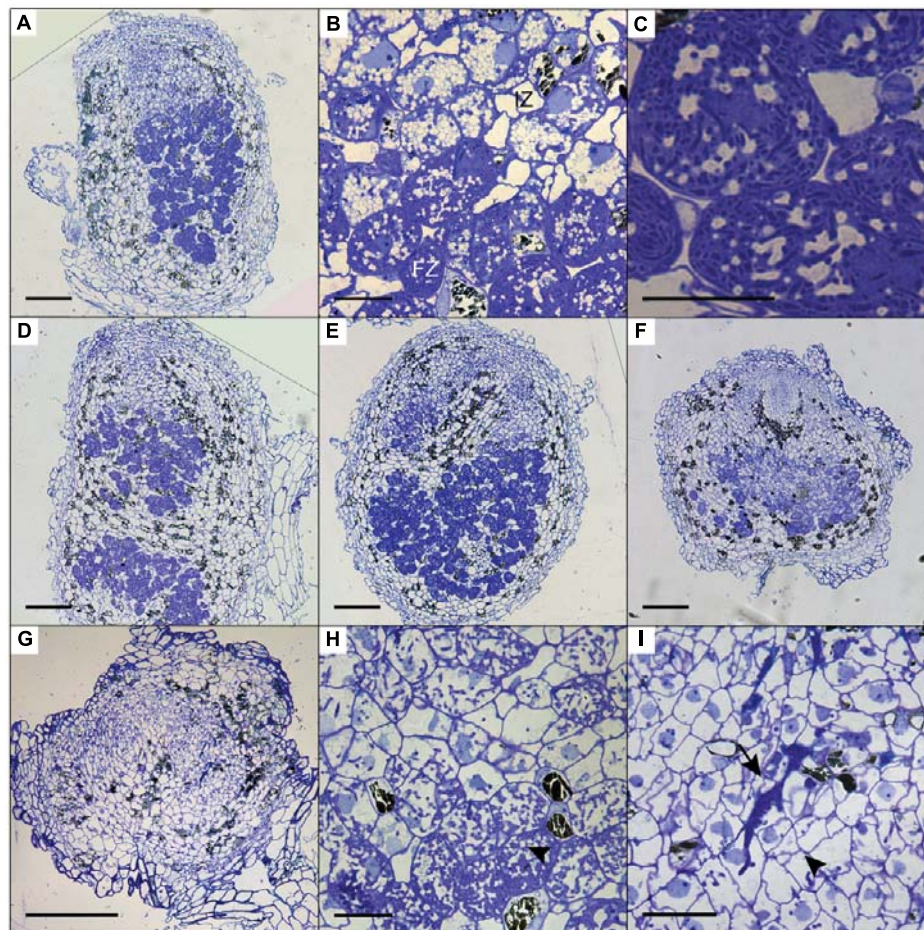
FIGURE 7 | Cytoarchitecture of CRISPR/Cas9 mutant nodules. **(A)** Longitudinal nodule sections of 1 month-old nodule formed on transgenic control line Ctr-9. **(B)** Zoom in on cells in the infection (IZ) and fixation zone (FZ) of the transgenic control nodule shown in **(A)**. Note the presence of small fragmented vacuoles in infected cells in the infection zone. **(C)** Zoom in on a cell in the fixation zone of a transgenic control nodule showing the presence of fixation threads. **(D)** Longitudinal nodule sections of 1 month-old nodule formed on *Pannsp1–13*. **(E)** Longitudinal nodule sections of 1 month-old nodule formed on *Panhk4–25*. **(F)** Longitudinal nodule sections of 1 month-old nodule formed on *Panein2–15*. **(G)** Longitudinal nodule sections of 1 month-old nodule formed on *Panein2–17*. **(H)** Zoom in on cells in the basal part of the *Panein2–15* nodule shown in **(F)**. Indicated by an arrowhead are cells containing fixation threads. **(I)** Zoom in of the *Panein2–17* nodule shown in **(G)**. Indicated by an arrowhead are infection threads. Indicated by an arrow are large apoplastic colonies. Scale bars in **(A,D–G)** are equal to 150 and 25 μm in **(B,C,H,I)**.

that ethylene signaling is required for efficient fixation thread formation in *P. andersonii* nodules.

Taken together, these data reveal symbiotic mutant phenotypes for *nsp1, nsp2* and *ein2*, whereas no effect on nodule formation was found by knocking out *hk4* in *P. andersonii*. Interestingly, we uncovered a novel role for the ethylene signaling component EIN2 in intracellular infection of *P. andersonii* nodules.

DISCUSSION

Comparative studies between legumes and the Cannabaceae tree *Parasponia* can provide insights into 'core' genetic networks underlying rhizobium symbiosis (van Velzen et al., 2017). To facilitate such studies, we aimed to establish a reverse genetics

platform for *P. andersonii* based on CRISPR/Cas9 genome editing. We show that using *A. tumefaciens* transformation, *P. andersonii* stable transgenic lines can be obtained in 3–4 months. In combination with CRISPR/Cas9 mutagenesis, this allows efficient generation of bi-allelic knockout mutants. As a proof-of-concept, we mutated four genes that commit essential symbiotic functions in legumes as well as control different hormonal networks. Characterization of the resulting lines revealed both symbiotic as well as non-symbiotic mutant phenotypes. Therefore, we conclude that stable *A. tumefaciens*-mediated transformation in combination with CRISPR/Cas9 genome editing can be efficiently used for reverse genetic analysis in *P. andersonii*.

Plant transformation efficiency is the main bottleneck in plant genome editing (Altpeter et al., 2016; Ledford, 2016). Especially regeneration of an entire transgenic plant out of a

single transformed cell remains difficult for most plant species. We took advantage of an efficient micro-propagation system available for *Parasponia* spp. to establish a protocol for stable transformation (Davey et al., 1993; Webster et al., 1995; Cao et al., 2012). About 8–12 weeks after cocultivation with *A. tumefaciens*, ∼50% of explants develop transgenic shoots. This relatively high efficiency is, in part, obtained through an initial 9-day culturing period on root-inducing medium, before incubation on standard propagation medium. This adaptation in the protocol was inspired by a recent study that showed that regeneration of plant cells consists of two distinctive steps (Kareem et al., 2015). Regenerative competence is established through activation of a root developmental program, followed by activation of shoot promoting factors that are required to complete shoot regeneration (Kareem et al., 2015). The latter explains why transfer to propagation medium is required to regenerate *P. andersonii* transgenic shoots. However, this promoting effect of rooting medium on regeneration of transgenic shoots might differ between different explant types, as noted for *P. andersonii* stems and petioles (Supplementary Table 3).

An advantage of the *Parasponia* system is that T$_0$ transgenic knockout mutants can be clonally propagated through *in vitro* micro-propagation (Davey et al., 1993; Webster et al., 1995; Cao et al., 2012). This allows a large number of rooted plantlets to be generated in a relatively short time span. As a result, phenotypic characterization can be initiated already at 4 months after the start of the transformation. However, a disadvantage of clonal propagation in combination with CRISPR/Cas9 mutagenesis is the possibility of obtaining chimeric mutants. Among the mutant lines we created, we identified one line (out of 11) that was chimeric for one out of three CRISPR target sites (Supplementary Figure 10C). Most mutant lines were genetically homogeneous, suggesting that mutations are induced soon after T-DNA integration. This is consistent with results in poplar, which also revealed a low percentage of chimeric mutants (Fan et al., 2015). Since chimeras are observed occasionally, thorough genotypic analysis will be required when phenotyping is performed in the T$_0$ generation. Besides vegetative propagation, *Parasponia* trees can also be propagated generatively. Under suitable greenhouse conditions, *Parasponia* trees flower within ∼6–9 months and are self-compatible (Becking, 1992). However, *Parasponia* trees can be monoecious or diecious and female flowers are wind pollinated (Soepadmo, 1974). This complicates selfing of trees and the production of pure seed badges. Additionally, *Parasponia* trees are fast growing and occupy a substantial amount of space in a tropical greenhouse (28°C, ∼100% relative humidity), making generative propagation of multiple mutant lines logistically somewhat challenging. An alternative to generative propagation is *in vitro* maintenance of transgenic lines. Additionally, the fast and efficient transformation procedure presented here will allow recreation of a particular mutant in less than 6 months.

Among the mutants we created, *Panhk4* and *Panein2* showed symbiotic phenotypes that differ from corresponding legume mutants. *P. andersonii Panhk4* mutants form nodules with a WT cytoarchitecture, indicating that these nodules are most likely functional. Analysis of stem cross-sections showed that *Panhk4* mutants possess a reduced procambial activity. Similar

phenotypes are observed in homologous mutants in *A. thaliana* (Mahonen et al., 2006a,b). Procambium activity is slightly reduced in the orthologs receptor mutant *arabidopsis histidine kinase 4* (*ahk4*), whereas it is completely abolished in the *ahk2 ahk3 ahk4* triple mutant (Mahonen et al., 2006a,b). The comparable phenotypes in cambium activity upon mutating histidine kinases suggest that *PanHK4* encodes a functional cytokinin receptor. *M. truncatula* and *L. japonicus* mutants in the cytokinin receptors orthologs to *PanHK4* are characterized as nodulation deficient (Murray et al., 2007; Plet et al., 2011). However, these mutants occasionally form nodules (Plet et al., 2011; Held et al., 2014; Boivin et al., 2016). This suggests redundant functioning of additional cytokinin receptors in both legume species. The *P. andersonii* genome also encodes two additional cytokinin receptors: *PanHK2* and *PanHK3* (van Velzen et al., 2017) (Supplementary Figure 4). Therefore, redundant functioning of one of these receptors cannot be excluded. In legumes, cell divisions associated with nodule development are initiated in the root cortex in response to epidermal perception of rhizobial signals (Timmers et al., 1999; Xiao et al., 2014). Cytokinin appears important for activation of this cortical organogenesis program (Vernie et al., 2015; Gamas et al., 2017). In *Parasponia*, cell divisions associated with nodule development are first observed in the epidermis, the cell layer that is in direct contact with the rhizobium bacteria (Lancelle and Torrey, 1984; Geurts et al., 2016). This difference in mitotically-responding tissues could create different dependencies on cytokinin signaling between legumes and *Parasponia*. However, whether this explains the absence of a symbiotic phenotype of *Panhk4* mutants requires further experimentation.

Panein2 mutants are ethylene insensitive, as indicated by the absence of leaf abscission following ethylene treatment. Additionally, we noticed a disturbed sex differentiation in *Panein2* flowers. Functioning of ethylene in flower sex differentiation is known in cucurbit species, like cucumber (*Cucumis sativus*) and melon (*Cucumis melo*) (Rudich et al., 1972; Yin and Quinn, 1995; Tanurdzic and Banks, 2004). Molecular genetic studies revealed that flower bud-specific expression of ACC synthase (ACS) genes, which are essential for biosynthesis of the ethylene precursor ACC, inhibits stamen development (Boualem et al., 2008, 2015). In line with these findings in cucurbits, we hypothesize that EIN2-mediated ethylene signaling commits a similar function in sex differentiation in *Parasponia* species.

In symbiotic context, *EIN2* knockout mutations result in different phenotypes between *Parasponia* and legumes. In legumes, ethylene negatively regulates rhizobial infection and root nodule formation (Penmetsa and Cook, 1997; Penmetsa et al., 2008; Miyata et al., 2013). This is illustrated by the phenotype of the *M. truncatula ein2* mutant (named *sickle*) that forms extensive epidermal infection threads and clusters of small nodules (Penmetsa and Cook, 1997; Xiao et al., 2014). *P. andersonii ein2* mutants also form smaller nodules than WT. However, in contrast to the *Mtein2* mutant, these nodules are regularly spaced on the root system. This suggests that in *Parasponia* ethylene signaling is not involved in regulating nodule number. Additionally, also the infection phenotype of

Panein2 mutants differs from that in legumes. In *M. truncatula* and *L. japonicus*, interference with ethylene signaling increases the number of epidermal infection threads but does not affect intracellular colonization of nodule cells (Penmetsa and Cook, 1997; Nukui et al., 2004; Lohar et al., 2009). In contrast, in *P. andersonii Panein2* mutants, intracellular colonization is hampered. Inside nodules, large apoplastic colonies are observed and fixation thread formation is severely reduced or even absent. This suggests that in *Parasponia* a functional ethylene signaling pathway is required for efficient intracellular infection of nodule cells.

Mutagenesis of the *NSP2* ortholog of *P. andersonii* indicated a conserved symbiotic role for this GRAS-type transcriptional regulator. In legumes, NSP2 works in concert with NSP1 to control root nodule formation (Hirsch et al., 2009). Mutagenesis of the *NSP1* ortholog of *P. andersonii* resulted in contrasting nodulation phenotypes. Two mutant lines, *Pannsp1–6* and *Pannsp1–13*, form nodules with a WT cytoarchitecture, whereas mutant line *Pannsp1–39* is unable to form nodules (**Figures 6**, **7**). However, all three mutants are affected in transcriptional regulation of strigolactone biosynthesis genes *PanD27* and *PanMAX1* (**Figure 5**). The three *Pannsp1* mutant lines differ from each other in the type of mutations that were created. *Pannsp1–6* and *Pannsp1–13* contain small deletions that are immediately followed by a second in-frame ATG that in WT PanNSP1 encodes a methionine at position 16. In contrast, *Pannsp1–39* contains a larger deletion that removes this in-frame ATG (see Supplementary Figure 9). Several reports have shown that alternative start codons are occasionally used to initiate transcription (Chabregas et al., 2003; Thatcher et al., 2007; Bazykin and Kochetov, 2011). Therefore, *Pannsp1–6* and *Pannsp1–13* most probably represent weak alleles that still possess residual PanNSP1 function. Such residual levels of PanNSP1 are affecting the expression of strigolactone biosynthesis genes, but are still sufficient to allow nodule formation. Therefore, we argue that the *P. andersonii Pannsp1–39* line carries a knockout mutation, indicating that in *P. andersonii* both NSP1 and NSP2 are essential for rhizobium root nodule formation.

Taken together, we showed that *P. andersonii* can be efficiently transformed using *A. tumefaciens* and is amenable to targeted mutagenesis using CRISPR/Cas9. This protocol takes only marginally more time than the transient *A. rhizogenes* transformation system that is generally used to study root nodule formation (e.g., Boisson-Dernier et al., 2001; Kumagai and Kouchi, 2003; Limpens et al., 2004; Op den Camp et al., 2011; Cao et al., 2012) but has several advantages. One of these is the absence of the *A. rhizogenes root inducing locus* (*rol*) that interferes with hormone homeostasis (Nilsson and Olsson, 1997). The protocol we developed will allow studies on *P. andersonii* symbiosis genes to determine to what extent legumes and *Parasponia* use a

similar mechanism to establish a nitrogen-fixing symbiosis with rhizobium.

DATA AVAILABILITY STATEMENT

All datasets analyzed for this study are included in the manuscript and the supplementary files. Gene identifiers for all *P. andersonii* genes used in this study can be found in Supplementary Table 8. Sequences can be downloaded from www.parasponia.org.

AUTHOR CONTRIBUTIONS

Conceptualization, AvZ and RG; Methodology, AvZ, MH, SL, and WK; Investigation, AvZ, TW, MSK, LR, FB, MH, SL, EF, and WK; Formal analysis, AvZ, TW, and EF; Visualization, AvZ; Writing – original draft, AvZ; Writing – review and editing, AvZ and RG; Funding acquisition, TB and RG; Supervision, RG.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2018.00284/full#supplementary-material

REFERENCES

Agustí, J., Merelo, P., Cercós, M., Tadeo, F. R., and Talón, M. (2009). Comparative transcriptional survey between laser-microdissected cells from

laminar abscission zone and petiolar cortical tissue during ethylene-promoted abscission in citrus leaves. *BMC Plant Biol.* 9:127. doi: 10.1186/1471-2229-9-127

Akkermans, A. D. L., Abdulkadir, S., and Trinick, M. J. (1978). Nitrogen-fixing root nodules in Ulmaceae. *Nature* 274:190. doi: 10.1038/274190c0

Altpeter, F., Springer, N. M., Bartley, L. E., Blechl, A. E., Brutnell, T. P., Citovsky, V., et al. (2016). Advancing crop transformation in the era of genome editing. *Plant Cell* 28, 1510–1520. doi: 10.1105/tpc.16.00196

Bansal, R., Mittapelly, P., Cassone, B. J., Mamidala, P., Redinbaugh, M. G., and Michel, A. (2015). Recommended reference genes for quantitative PCR analysis in soybean have variable stabilities during diverse biotic stresses. *PLoS One* 10:e0134890. doi: 10.1371/journal.pone.0134890

Bazykin, G. A., and Kochetov, A. V. (2011). Alternative translation start sites are conserved in eukaryotic genomes. *Nucleic Acids Res.* 39, 567–577. doi: 10.1093/nar/gkq806

Becking, J. H. (1983). The *Parasponia parviflora*-rhizobium symbiosis-isotopic nitrogen-fixation, hydrogen evolution and nitrogen-fixation efficiency, and oxygen relations. *Plant Soil* 75, 343–360. doi: 10.1007/BF02369970

Becking, J. H. (1992). "The Rhizobium symbiosis of the nonlegume *Parasponia*," in *Biological Nitrogen Fixation*, eds G. S. Stacey, H. J. Evans, and R. H. Burris (New York, NY: Routledge), 497–559.

Behm, J. E., Geurts, R., and Kiers, E. T. (2014). *Parasponia*: a novel system for studying mutualism stability. *Trends Plant Sci.* 19, 757–763. doi: 10.1016/j.tplants.2014.08.007

Bhalerao, R. P., and Fischer, U. (2017). Environmental and hormonal control of cambial stem cell dynamics. *J. Exp. Bot.* 68, 79–87. doi: 10.1093/jxb/erw466

Boisson-Dernier, A., Chabaud, M., Garcia, F., Bécard, G., Rosenberg, C., and Barker, D. G. (2001). *Agrobacterium rhizogenes*-transformed roots of *Medicago truncatula* for the study of nitrogen-fixing and endomycorrhizal symbiotic associations. *Mol. Plant Microbe Interact.* 14, 695–700. doi: 10.1094/MPMI.2001.14.6.695

Boivin, S., Kazmierczak, T., Brault, M., Wen, J. Q., Gamas, P., Mysore, K. S., et al. (2016). Different cytokinin histidine kinase receptors regulate nodule initiation as well as later nodule developmental stages in *Medicago truncatula*. *Plant Cell Environ.* 39, 2198–2209. doi: 10.1111/pce.12779

Boualem, A., Fergany, M., Fernandez, R., Troadec, C., Martin, A., Morin, H., et al. (2008). A conserved mutation in an ethylene biosynthesis enzyme leads to andromonoecy in melons. *Science* 321, 836–838. doi: 10.1126/science.1159023

Boualem, A., Troadec, C., Camps, C., Lemhemdi, A., Morin, H., Sari, M. A., et al. (2015). A cucurbit androecy gene reveals how unisexual flowers develop and dioecy emerges. *Science* 350, 688–691. doi: 10.1126/science.aac8370

Brown, K. M. (1997). Ethylene and abscission. *Physiol. Plant.* 100, 567–576. doi: 10.1111/j.1399-3054.1997.tb03062.x

Cao, Q., Op den Camp, R., Seifi Kalhor, M., Bisseling, T., and Geurts, R. (2012). Efficiency of *Agrobacterium rhizogenes*–mediated root transformation of *Parasponia* and *Trema* is temperature dependent. *Plant Growth Regul.* 68, 459–465. doi: 10.1007/s10725-012-9734-y

Cardoso, C., Zhang, Y. X., Jamil, M., Hepworth, J., Charnikhova, T., Dimkpa, S. O. N., et al. (2014). Natural variation of rice strigolactone biosynthesis is associated with the deletion of two *MAX1* orthologs. *Proc. Natl. Acad. Sci. U.S.A.* 111, 2379–2384. doi: 10.1073/pnas.1317360111

Chabregas, S. M., Luche, D. D., Van Sluys, M. A., Menck, C. F. M., and Silva-Filho, M. C. (2003). Differential usage of two in-frame translational start codons regulates subcellular localization of *Arabidopsis thaliana* THI1. *J. Cell Sci.* 116, 285–291. doi: 10.1242/jcs.00228

Clason, E. W. (1936). The vegetation of the upper-Badak region of mount Kelut (east java). *Bull. Jard. Bot. Buitenz. III* 13, 509–518. doi: 10.1073/pnas.1714977115

Czechowski, T., Stitt, M., Altmann, T., Udvardi, M. K., and Scheible, W. R. (2005). Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. *Plant Physiol.* 139, 5–17. doi: 10.1104/pp.105.063743

Davey, M. R., Webster, G., Manders, G., Ringrose, F. L., Power, J. B., and Cocking, E. C. (1993). Effective nodulation of micro-propagated shoots of the non-legume *Parasponia andersonii* by *Bradyrhizobium*. *J. Exp. Bot.* 44, 863–867. doi: 10.1093/jxb/44.5.863

Doench, J. G., Hartenian, E., Graham, D. B., Tothova, Z., Hegde, M., Smith, I., et al. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat. Biotechnol.* 32, 1262–1267. doi: 10.1038/nbt.3026

Duarte, J. M., Wall, P. K., Edger, P. P., Landherr, L. L., Ma, H., Pires, J. C., et al. (2010). Identification of shared single copy nuclear genes in *Arabidopsis*,

Populus, Vitis and *Oryza* and their phylogenetic utility across various taxonomic levels. *BMC Evol. Biol.* 10:61. doi: 10.1186/1471-2148-10-61

Engler, C., Gruetzner, R., Kandzia, R., and Marillonnet, S. (2009). Golden Gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS One* 4:e5553. doi: 10.1371/journal.pone.0005553

Engler, C., Youles, M., Gruetzner, R., Ehnert, T. M., Werner, S., Jones, J. D. G., et al. (2014). A Golden Gate modular cloning toolbox for plants. *ACS Synth. Biol.* 3, 839–843. doi: 10.1021/sb4001504

Fan, D., Liu, T. T., Li, C. F., Jiao, B., Li, S., Hou, Y. S., et al. (2015). Efficient CRISPR/Cas9-mediated targeted mutagenesis in *Populus* in the first generation. *Sci. Rep.* 5:12217. doi: 10.1038/srep12217

Fauser, F., Schiml, S., and Puchta, H. (2014). Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in *Arabidopsis thaliana*. *Plant J.* 79, 348–359. doi: 10.1111/tpj.12554

Fedorova, E., Thomson, R., Whitehead, L. F., Maudoux, O., Udvardi, M. K., and Day, D. A. (1999). Localization of H^+-ATPases in soybean root nodules. *Planta* 209, 25–32. doi: 10.1007/s004250050603

Gamas, P., Brault, M., Jardinaud, M.-F., and Frugier, F. (2017). Cytokinins in symbiotic nodulation: when, where, what for? *Trends Plant Sci.* 22, 792–802. doi: 10.1016/j.tplants.2017.06.012

Geurts, R., Lillo, A., and Bisseling, T. (2012). Exploiting an ancient signalling machinery to enjoy a nitrogen fixing symbiosis. *Curr. Opin. Plant Biol.* 15, 438–443. doi: 10.1016/j.pbi.2012.04.004

Geurts, R., Xiao, T. T., and Reinhold-Hurek, B. (2016). What does it take to evolve a nitrogen-fixing endosymbiosis? *Trends Plant Sci.* 21, 199–208. doi: 10.1016/j.tplants.2016.01.012

Gonzalez-Rizzo, S., Crespi, M., and Frugier, F. (2006). The *Medicago truncatula* CRE1 cytokinin receptor regulates lateral root development and early symbiotic interaction with *Sinorhizobium meliloti*. *Plant Cell* 18, 2680–2693. doi: 10.1105/tpc.106.043778

Granqvist, E., Sun, J., Op den Camp, R., Pujic, P., Hill, L., Normand, P., et al. (2015). Bacterial-induced calcium oscillations are common to nitrogen-fixing associations of nodulating legumes and non-legumes. *New Phytol.* 207, 551–558. doi: 10.1111/nph.13464

Heckmann, A. B., Lombardo, F., Miwa, H., Perry, J. A., Bunnewell, S., Parniske, M., et al. (2006). *Lotus japonicus* nodulation requires two GRAS domain regulators, one of which is functionally conserved in a non-legume. *Plant Physiol.* 142, 1739–1750. doi: 10.1104/pp.106.089508

Held, M., Hou, H., Miri, M., Huynh, C., Ross, L., Hossain, M. S., et al. (2014). *Lotus japonicus* cytokinin receptors work partially redundantly to mediate nodule formation. *Plant Cell* 26, 678–694. doi: 10.1105/tpc.113.119362

Hirsch, S., Kim, J., Muñoz, A., Heckmann, A. B., Downie, J. A., and Oldroyd, G. E. D. (2009). GRAS proteins form a DNA binding complex to induce gene expression during nodulation signaling in *Medicago truncatula*. *Plant Cell* 21, 545–557. doi: 10.1105/tpc.108.064501

Jackson, M. B., and Osborne, D. J. (1970). Ethylene, the natural regulator of leaf abscission. *Nature* 225, 1019–1022. doi: 10.1038/2251019a0

Kaló, P., Gleason, C., Edwards, A., Marsh, J., Mitra, R. M., Hirsch, S., et al. (2005). Nodulation signaling in legumes requires NSP2, a member of the GRAS family of transcriptional regulators. *Science* 308, 1786–1789. doi: 10.1126/science.1110951

Kareem, A., Durgaprasad, K., Sugimoto, K., Du, Y., Pulianmackal, A. J., Trivedi, Z. B., et al. (2015). PLETHORA genes control regeneration by a two-step mechanism. *Curr. Biol.* 25, 1017–1030. doi: 10.1016/j.cub.2015.02.022

Karimi, M., Inze, D., and Depicker, A. (2002). GATEWAY™ vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci.* 7, 193–195. doi: 10.1016/S1360-1385(02)02251-3

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436

Kumagai, H., and Kouchi, H. (2003). Gene silencing by expression of hairpin RNA in *Lotus japonicus* roots and root nodules. *Mol. Plant Microbe Interact.* 16, 663–668. doi: 10.1094/MPMI.2003.16.8.663

Lamesch, P., Berardini, T. Z., Li, D. H., Swarbreck, D., Wilks, C., Sasidharan, R., et al. (2012). The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* 40, D1202–D1210. doi: 10.1093/nar/gkr1090

Lancelle, S. A., and Torrey, J. G. (1984). Early development of *Rhizobium*-induced root nodules of *Parasponia rigida*. I. Infection and early nodule initiation. *Protoplasma* 123, 26–37. doi: 10.1007/BF01283179

Lazo, G. R., Stein, P. A., and Ludwig, R. A. (1991). A DNA transformation-competent *Arabidopsis* genomic library in *Agrobacterium*. *Biotechnology* 9, 963–967. doi: 10.1038/nbt1091-963

Ledford, H. (2016). A better way to hack plant DNA. *Nature* 539, 16–17.

Li, H.-L., Wang, W., Mortimer, P. E., Li, R.-Q., Li, D.-Z., Hyde, K. D., et al. (2015). Large-scale phylogenetic analyses reveal multiple gains of actinorhizal nitrogen-fixing symbioses in angiosperms associated with climate change. *Sci. Rep.* 5:14023. doi: 10.1038/srep14023

Limpens, E., Ramos, J., Franken, C., Raz, V., Compaan, B., Franssen, H., et al. (2004). RNA interference in *Agrobacterium rhizogenes*-transformed roots of *Arabidopsis* and *Medicago truncatula*. *J. Exp. Bot.* 55, 983–992. doi: 10.1093/jxb/erh122

Liu, W., Kohlen, W., Lillo, A., Op den Camp, R., Ivanov, S., Hartog, M., et al. (2011). Strigolactone biosynthesis in *Medicago truncatula* and rice requires the symbiotic GRAS-type transcription factors NSP1 and NSP2. *Plant Cell* 23, 3853–3865. doi: 10.1105/tpc.111.089771

Lohar, D., Stiller, J., Kam, J., Stacey, G., and Gresshoff, P. M. (2009). Ethylene insensitivity conferred by a mutated *Arabidopsis* ethylene receptor gene alters nodulation in transgenic *Lotus japonicus*. *Ann. Bot.* 104, 277–285. doi: 10.1093/aob/mcp132

Mahonen, A. P., Bishopp, A., Higuchi, M., Nieminen, K. M., Kinoshita, K., Tormakangas, K., et al. (2006a). Cytokinin signaling and its inhibitor AHP6 regulate cell fate during vascular development. *Science* 311, 94–98.

Mahonen, A. P., Higuchi, M., Tormakangas, K., Miyawaki, K., Pischke, M. S., Sussman, M. R., et al. (2006b). Cytokinins regulate a bidirectional phosphorelay network in *Arabidopsis*. *Curr. Biol.* 16, 1116–1122.

Marvel, D. J., Torrey, J. G., and Ausubel, F. M. (1987). *Rhizobium* symbiotic genes required for nodulation of legume and nonlegume hosts. *Proc. Natl. Acad. Sci. U.S.A.* 84, 1319–1323. doi: 10.1073/pnas.84.5.1319

Matsumoto-Kitano, M., Kusumoto, T., Tarkowski, P., Kinoshita-Tsujimura, K., Václavíková, K., Miyawaki, K., et al. (2008). Cytokinins are central regulators of cambial activity. *Proc. Natl. Acad. Sci. U.S.A.* 105, 20027–20031. doi: 10.1073/pnas.0805619105

Miyata, K., Kawaguchi, M., and Nakagawa, T. (2013). Two distinct *EIN2* genes cooperatively regulate ethylene signaling in *Lotus japonicus*. *Plant Cell Physiol.* 54, 1469–1477. doi: 10.1093/pcp/pct095

Murray, J. D., Karas, B. J., Sato, S., Tabata, S., Amyot, L., and Szczyglowski, K. (2007). A cytokinin perception mutant colonized by *Rhizobium* in the absence of nodule organogenesis. *Science* 315, 101–104. doi: 10.1126/science.113 2514

Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J. D. G., and Kamoun, S. (2013). Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* 31, 691–693. doi: 10.1038/nbt.2655

Nieminen, K., Immanen, J., Laxell, M., Kauppinen, L., Tarkowski, P., Dolezal, K., et al. (2008). Cytokinin signaling regulates cambial development in poplar. *Proc. Natl. Acad. Sci. U.S.A.* 105, 20032–20037. doi: 10.1073/pnas.080561 7106

Nilsson, O., and Olsson, O. (1997). Getting to the root: the role of the *Agrobacterium rhizogenes rol* genes in the formation of hairy roots. *Physiol. Plant.* 100, 463–473. doi: 10.1111/j.1399-3054.1997.tb03050.x

Nukui, N., Ezura, H., and Minamisawa, K. (2004). Transgenic *Lotus japonicus* with an ethylene receptor gene *Cm-ERS1/H70A* enhances formation of infection threads and nodule primordia. *Plant Cell Physiol.* 45, 427–435. doi: 10.1093/pcp/pch046

Oldroyd, G. E. D. (2013). Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nat. Rev. Microbiol.* 11, 252–263. doi: 10.1038/nrmicro2990

Op, den Camp, R. H. M., Polone, E., Fedorova, E., Roelofsen, W., Squartini, A., et al. (2012). Nonlegume *Parasponia andersonii* deploys a broad *Rhizobium* host range strategy resulting in largely variable symbiotic effectiveness. *Mol. Plant Microbe Interact.* 25, 954–963. doi: 10.1094/MPMI-11-11-0304

Op den Camp, R., Streng, A., De Mita, S., Cao, Q., Polone, E., Liu, W., et al. (2011). LysM-type mycorrhizal receptor recruited for rhizobium symbiosis in nonlegume *Parasponia*. *Science* 331, 909–912. doi: 10.1126/science. 1198181

Penmetsa, R. V., and Cook, D. R. (1997). A legume ethylene-insensitive mutant hyperinfected by its rhizobial symbiont. *Science* 275, 527–530. doi: 10.1126/science.275.5299.527

Penmetsa, V. R., Uribe, P., Anderson, J., Lichtenzveig, J., Gish, J.-C., Nam, Y. W., et al. (2008). The *Medicago truncatula* ortholog of Arabidopsis EIN2, *sickle*, is a negative regulator of symbiotic and pathogenic microbial associations. *Plant J.* 55, 580–595. doi: 10.1111/j.1365-313X.2008.03531.x

Plet, J., Wasson, A., Ariel, F., Le Signor, C., Baker, D., Mathesius, U., et al. (2011). MtCRE1-dependent cytokinin signaling integrates bacterial and plant cues to coordinate symbiotic nodule organogenesis in *Medicago truncatula*. *Plant J.* 65, 622–633. doi: 10.1111/j.1365-313X.2010.04447.x

Price, M. N., Dehal, P. S., and Arkin, A. P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* 26, 1641–1650. doi: 10.1093/molbev/msp077

Rudich, J., Kedar, N., and Halevy, A. H. (1972). Ethylene evolution from cucumber plants as related to sex expression. *Plant Physiol.* 49, 998–999. doi: 10.1104/pp. 49.6.998

Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J. X., Mitros, T., Nelson, W., et al. (2010). Genome sequence of the palaeopolyploid soybean. *Nature* 463, 178–183. doi: 10.1038/nature08670

Shtark, O. Y., Sulima, A. S., Zhernakov, A. I., Kliukova, M. S., Fedorina, J. V., Pinaev, A. G., et al. (2016). *Arbuscular mycorrhiza* development in pea (*Pisum sativum* L.) mutants impaired in five early nodulation genes including putative orthologs of NSP1 and NSP2. *Symbiosis* 68, 129–144. doi: 10.1007/s13199-016-0382-2

Smit, P., Raedts, J., Portyanko, V., Debellé, F., Gough, C., Bisseling, T., et al. (2005). NSP1 of the GRAS protein family is essential for rhizobial Nod factor-induced transcription. *Science* 308, 1789–1791. doi: 10.1126/science.1111025

Soepadmo, E. (1974). Ulmaceae. *Flora Malesiana* 8, 31–76.

Tang, H. B., Krishnakumar, V., Bidwell, S., Rosen, B., Chan, A. N., Zhou, S. G., et al. (2014). An improved genome release (version Mt4.0) for the model legume *Medicago truncatula*. *BMC Genomics* 15:312. doi: 10.1186/1471-2164-15-312

Tanurdzic, M., and Banks, J. A. (2004). Sex-determining mechanisms in land plants. *Plant Cell* 16, S61–S71. doi: 10.1105/tpc.016667

Thatcher, L. F., Carrie, C., Andersson, C. R., Sivasithamparam, K., Whelan, J., and Singh, K. B. (2007). Differential gene expression and subcellular targeting of Arabidopsis glutathione S-transferase F8 is achieved through alternative transcription start sites. *J. Biol. Chem.* 282, 28915–28928. doi: 10.1074/jbc. M702207200

Timmers, A. C. J., Auriac, M.-C., and Truchet, G. (1999). Refined analysis of early symbiotic steps of the *Rhizobium-Medicago* interaction in relationship with microtubular cytoskeleton rearrangements. *Development* 126, 3617–3628.

Tirichine, L., Sandal, N., Madsen, L. H., Radutoiu, S., Albrektsen, A. S., Sato, S., et al. (2007). A gain-of-function mutation in a cytokinin receptor triggers spontaneous root nodule organogenesis. *Science* 315, 104–107. doi: 10.1126/science.1132397

Trinick, M. J. (1973). Symbiosis between *Rhizobium* and the non-legume. *Trema aspera*. *Nature* 244, 459–460. doi: 10.1038/244459a0

Trinick, M. J. (1980). Growth of *Parasponia* in agar tube culture and symbiotic effectiveness of isolates from *Parasponia* spp. *New Phytol.* 85, 37–45. doi: 10.1111/j.1469-8137.1980.tb04446.x

Trinick, M. J. (1981). "The effective rhizobium symbiosis with the non-legume *Parasponia andersonii*," in *Current Perspectives in Nitrogen Fixation*, eds A. H. Gibson and W. E. Newton (Canberra: Australian Academy of Sciences), 480.

Trinick, M. J., and Hadobas, P. A. (1989). "Biology of the *Parasponia-Bradyrhizobium* symbiosis," in *Nitrogen Fixation with Non-Legumes*, eds J. K. Ladha and K. A. Malik (Netherlands: Springer), 25–33. doi: 10.1007/978-94-009-0889-5_3

Tuskan, G. A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., et al. (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313, 1596–1604. doi: 10.1126/science.1128691

van Velzen, R., Holmer, R., Bu, F., Rutten, L., van Zeijl, A., Liu, W., et al. (2017). Parallel loss of symbiosis genes in relatives of nitrogen-fixing non-legume *Parasponia*. *bioRxiv* 169706. doi: 10.1101/169706

van Zeijl, A. (2017). *Dissecting Hormonal Pathways in Nitrogen-Fixing Rhizobium Symbioses*. Ph.D. thesis, Wageningen University: Wageningen, doi: 10.18174/419674

van Zeijl, A., Liu, W., Xiao, T. T., Kohlen, W., Yang, W.-C., Bisseling, T., et al. (2015a). The strigolactone biosynthesis gene *DWARF27* is co-opted in rhizobium symbiosis. *BMC Plant Biol.* 15:260. doi: 10.1186/s12870-015-0651-x

van Zeijl, A., Op den Camp, R. H. M., Deinum, E. E., Charnikhova, T., Franssen, H., Op den Camp, H. J. M., et al. (2015b). Rhizobium lipo-chitooligosaccharide signaling triggers accumulation of cytokinins in *Medicago truncatula* roots. *Mol. Plant* 8, 1213–1226. doi: 10.1016/j.molp.2015.03.010

Vernie, T., Kim, J., Frances, L., Ding, Y., Sun, J., Guan, D., et al. (2015). The NIN transcription factor coordinates diverse nodulation programs in different tissues of the *Medicago truncatula* root. *Plant Cell* 27, 3410–3424. doi: 10.1105/tpc.15.00461

Wang, H., Moore, M. J., Soltis, P. S., Bell, C. D., Brockington, S. F., Alexandre, R., et al. (2009). Rosid radiation and the rapid rise of angiosperm-dominated forests. *Proc. Natl. Acad. Sci. U.S.A.* 106, 3853–3858. doi: 10.1073/pnas.0813376106

Webster, G., Poulton, P. R., Cocking, E. C., and Davey, M. R. (1995). The nodulation of micro-propagated plants of *Parasponia andersonii* by tropical legume rhizobia. *J. Exp. Bot.* 46, 1131–1137. doi: 10.1093/jxb/46.9.1131

Xiao, T. T., Schilderink, S., Moling, S., Deinum, E. E., Kondorosi, E., Franssen, H., et al. (2014). Fate map of *Medicago truncatula* root nodules. *Development* 141, 3517–3528. doi: 10.1242/dev.110775

Yin, T. J., and Quinn, J. A. (1995). Tests of a mechanistic model of one hormone regulating both sexes in *Cucumis sativus* (Cucurbitaceae). *Am. J. Bot.* 82, 1537–1546. doi: 10.1002/j.1537-2197.1995.tb13856.x

Young, N. D., Debelle, F., Oldroyd, G. E., Geurts, R., Cannon, S. B., Udvardi, M. K., et al. (2011). The *Medicago genome* provides insight into the evolution of rhizobial symbioses. *Nature* 480, 520–524. doi: 10.1038/nature10625

Zhang, Y. X., van Dijk, A. D. J., Scaffidi, A., Flematti, G. R., Hofmann, M., Charnikhova, T., et al. (2014). Rice cytochrome P450 MAX1 homologs catalyze distinct steps in strigolactone biosynthesis. *Nat. Chem. Biol.* 10, 1028–1033. doi: 10.1038/nchembio.1660

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Genetic and Molecular Mechanisms Underlying Symbiotic Specificity in Legume-Rhizobium Interactions

Qi Wang†, Jinge Liu† and Hongyan Zhu*

Department of Plant and Soil Sciences, University of Kentucky, Lexington, KY, United States

Legumes are able to form a symbiotic relationship with nitrogen-fixing soil bacteria called rhizobia. The result of this symbiosis is to form nodules on the plant root, within which the bacteria can convert atmospheric nitrogen into ammonia that can be used by the plant. Establishment of a successful symbiosis requires the two symbiotic partners to be compatible with each other throughout the process of symbiotic development. However, incompatibility frequently occurs, such that a bacterial strain is unable to nodulate a particular host plant or forms nodules that are incapable of fixing nitrogen. Genetic and molecular mechanisms that regulate symbiotic specificity are diverse, involving a wide range of host and bacterial genes/signals with various modes of action. In this review, we will provide an update on our current knowledge of how the recognition specificity has evolved in the context of symbiosis signaling and plant immunity.

Keywords: legume, nodulation, nitrogen fixation, rhizobial symbiosis, host specificity

INTRODUCTION

The legume-rhizobial symbiosis starts with a signal exchange between the host plant and its microsymbiont (Oldroyd, 2013). Recognition of compatible bacteria by the host induces cortical cell divisions to form root nodule primordia, and simultaneously initiates an infection process to deliver the bacteria into the nodule cells. Infection of most legumes involves the development of plant-made infection threads that initiate in the root hair. The infection threads harboring dividing bacteria grow through the epidermal cell layer into the nodule cells, where the bacteria are released and internalized in an endocytosis-like process. In nodule cells, individual bacteria are enclosed by a membrane of plant origin, forming an organelle-like structure called the symbiosome, within which the bacteria further differentiate into nitrogen-fixing bacteroids (Jones et al., 2007; Oldroyd et al., 2011).

Symbiotic nodule development involves synchronous differentiation of both nodule and bacterial cells. Legume nodules can be grouped into two major types: indeterminate (e.g., pea, clovers, and *Medicago*) and determinate (e.g., soybeans, common bean, and *Lotus*) (Nap and Bisseling, 1990; Hirsch, 1992). Indeterminate nodules originate from cell divisions in the inner cortex and possess a persistent apical meristem. Consequently, indeterminate nodules are cylindrical in shape, with a developmental gradient from the apex to the base of the nodule, which can be divided into different nodule zones (Nap and Bisseling, 1990). In contrast, determinate nodules result from cell divisions in the middle/outer cortex of the root, lack a persistent meristem, and are spherical in shape. Cell divisions of a determinate nodule cease at early developmental

stages and the mature nodule develops through cell enlargement; as such, the infected cells develop more or less synchronously to the nitrogen-fixing stage. In both nodule types, the symbiotic nodule cells undergo genome endoreduplication, leading to polyploidization and cell enlargement. Parallel to the nodule cell development is the differentiation of the nitrogen-fixing bacteroids. Depending on the host, but independent of the nodule type, such bacterial differentiation can be terminal or reversible. Terminal differentiation is featured by genome endoreduplication, cell elongation, increased membrane permeability, and loss of reproductive ability, while in reversible differentiation the bacteroids retain cell size and DNA content similar to free-living bacteria (Kereszt et al., 2011; Oldroyd et al., 2011; Haag et al., 2013). Compared to free-living bacteria, the bacteroids display dramatic changes in transcriptome, cell surface structure and metabolic activities so that they become better adapted to the intracellular environment and dedicated to nitrogen fixation (Mergaert et al., 2006; Prell and Poole, 2006; Haag et al., 2013).

Both legumes and rhizobial bacteria are phylogenetically diverse. No single rhizobial strains can form symbiosis with all legumes, and vice versa. Specificity occurs at both species and genotypic levels (Broughton et al., 2000; Perret et al., 2000; Wang et al., 2012). This can take place at early stages of the interaction so that the same bacterial strains can infect and nodulate one host plant but not another (Yang et al., 2010; Wang et al., 2012; Tang et al., 2016; Fan et al., 2017). Incompatibility also frequently happens at later stages of nodule development such that nitrogen-fixing efficiency differs significantly between different plant-bacteria combinations (Wang et al., 2012, 2017, 2018; Yang et al., 2017). Symbiotic specificity results from the changing of signals from both host and bacterial sides; as such, various recognition mechanisms have evolved during the process of co-adaptation. Knowledge of the genetic and molecular basis of symbiotic specificity is important for developing tools for genetic manipulation of the host or bacteria in order to enhance nitrogen fixation efficiency. In this review, we will discuss our current understanding of the evolution of specificity in the root nodule symbiosis.

SPECIFICITY MEDIATED BY FLAVONOIDS AND THE FLAVONOID-NodD RECOGNITION

Under nitrogen-limiting conditions, legume roots secrete a cocktail of flavonoid compounds into the rhizosphere, and they serve to activate the expression of a group of bacterial nodulation (*nod*) genes, leading to the synthesis of the Nod factor, a lipochitooligosaccharidic signal that is essential for initiating symbiotic development in most legumes (Oldroyd et al., 2011). Induction of *nod* gene expression is mediated by the flavonoid-activated NodD proteins, which are LysR-type transcription regulators (Long, 1996). NodDs activate *nod* gene expression through binding to the conserved DNA motifs (*nod* boxes) upstream of the *nod* operons (Rostas et al., 1986; Fisher et al., 1988).

NodD proteins from different rhizobia are adapted to recognizing different flavonoids secreted by different legumes, and this recognition specificity defines an early checkpoint of the symbiosis (Peck et al., 2006). Despite the absence of direct evidence for physical interaction between the two molecules, flavonoids have been shown to be able to stimulate the binding of NodD to *nod* gene promoters in *Sinorhizobium meliloti* (Peck et al., 2006). It is well documented that inter-strain exchange of *nodD* genes can alter the response of the recipient strain to a different set of flavonoid inducers and hence the host range (Horváth et al., 1987; Perret et al., 2000). For example, the transfer of *nodD1* from the broad host range symbiont *Rhizobium* sp. NGR234 to the restricted host range strain *Rhizobium leguminosarum* biovar *trifolii* ANU843 enabled the recipient strain to nodulate the non-legume *Parasponia*, because the wide-host-range NodD1 protein is capable of recognizing a broader spectrum of flavonoid inducers (Bender et al., 1988).

The evidence for the importance of flavonoids in determining host range primarily comes from bacterial genetics, and the plant genes involved are less studied. Since legume roots secrete a complex mixture of flavonoid compounds, it is difficult to pinpoint which flavonoids play a more critical role, and when and where they are produced. Recent studies in soybeans and *Medicago truncatula* have highlighted key flavonoids required for rhizobial infection (reviewed in Liu and Murray, 2016). These so called "infection flavonoids" are strong inducers of *nod* genes, secreted by roots, highly accumulated at the infection sites, and show increased biosynthesis in response to infection by compatible rhizobia. Although luteolin was the first flavonoid identified that can induce *nod* gene expression across a wide range of rhizobial strains, it is not legume-specific, mainly produced in germinating seeds, and has not been detected in root exudates or nodules. In contrast, methoxychalcone has been shown to be one of the strong host infection signals from *Medicago* and closely related legumes that form indeterminate nodules, while genistein and daidzein are crucial signals from soybeans that form determinate nodules. Part of the flavonoid compounds may also function as phytoalexins, acting to reinforce symbiosis specificity (Liu and Murray, 2016). For example, *Bradyrhizobium japonicum* and *Mesorhizobium loti*, but not the *Medicago* symbiont *S. meliloti*, are susceptible to the isoflavonoid medicarpin produced by *Medicago* spp. (Pankhurst and Biggs, 1980; Breakspear et al., 2014), and the soybean symbionts *B. japonicum* and *S. fredii* are resistant to glyceollin when exposed to genistein and daidzein (Parniske et al., 1991).

SPECIFICITY MEDIATED BY Nod-FACTOR PERCEPTION

Nod factors produced by rhizobia are an essential signaling component for symbiosis development in most legumes. Nod factors are lipochito-oligosaccharides, consisting of four or five 1,4-linked *N*-acetyl-glucosamine residues that carry a fatty acyl chain of varying length attached to the C-2 position of the non-reducing end and various species-specific chemical decorations at both the reducing and non-reducing ends (Dénarié et al., 1996).

The common *nodABC* genes contribute to the synthesis of the chitin backbone, while other strain-specific *nod* genes act to modify the backbone by changing the size and saturation of the acyl chain, or adding to the terminal sugar units with acetyl, methyl, carbamoyl, sulfuryl or glycosyl groups. Structural variations in Nod factors are a key determinant of host range, because these Nod factors have to be recognized by the host in order to initiate infection and nodulation (Perret et al., 2000; D'Haeze and Holsters, 2002).

Nod factors are perceived by Nod-factor receptors (e.g., NFR1 and NFR5 in *Lotus japonicus*), which are LysM-domain-containing receptor kinases (Limpens et al., 2003; Madsen et al., 2003; Radutoiu et al., 2003). Direct binding of Nod factors to the extracellular LysM domains of the receptor complex leads to activation of the downstream nodulation signaling pathways (Broghammer et al., 2012). Specificity in Nod-factor binding is widely thought to be critical for recognition between the prospective symbiotic partners. This hypothesis has been strongly supported by genetic evidence even though such binding specificity has not been demonstrated. The best examples are from the pea-*R. leguminosarum* symbiosis where bacterial *nod* gene mutants that lead to changed Nod factor composition or structure exhibited genotype-specific nodulation (Firmin et al., 1993; Bloemberg et al., 1995). This alteration of host range corresponds to allelic variations at the *Sym2/Sym37/PsK1* locus, an orthologous region of *NFR1* that contains a cluster of LysM receptor kinases (Zhukov et al., 2008; Li et al., 2011). In this case, allelic variation coupled with gene duplication and diversification contribute to alterations in symbiotic compatibility.

Nod factor recognition presumably plays a more critical role in determining host range at species level, which has been best illustrated on the bacterial side. However, natural polymorphisms in Nod-factor receptors that are responsible for nodulation specificity between different legumes have not been well studied at the genetic level, simply because the plants cannot be interbred. Nevertheless, transferring *NFR1* and *NFR5* of *L. japonicus* into *M. truncatula* enables nodulation of the transformants by the *L. japonicus* symbiont *Mesorhizobium loti* (Radutoiu et al., 2007).

SPECIFICITY MEDIATED BY PERCEPTION OF RHIZOBIAL EXOPOLYSACCHARIDES

In addition to Nod factors, rhizobial surface polysaccharides such as exopolysaccharides (EPS), lipopolysaccharides (LPS), and capsular polysaccharides (KPS) are also thought to be important for establishing symbiotic relationships (Fraysse et al., 2003; Becker et al., 2005; Jones et al., 2007; Gibson et al., 2008). These surface components are proposed to be able to suppress plant defense, but their active roles in promoting bacterial infection and nodulation remain elusive and are dependent on the specific interactions studied.

Exopolysaccharides have been shown to be required for rhizobial infection in multiple symbiotic interactions. This has been best illustrated in the *Sinorhizobium-Medicago* symbiosis,

in which succinoglycan, a major EPS produced by *S. meliloti*, is required for the initiation and elongation of infection threads, and increased succinoglycan production enhances nodulation capacity (Leigh et al., 1985; Reinhold et al., 1994; Cheng and Walker, 1998; Jones, 2012). However, the symbiotic role of EPS is very complicated in the *Mesorhizobium-Lotus* interaction (Kelly et al., 2013). For instance, a subset of EPS mutants of *M. loti* R7A displayed severe nodulation deficiencies on *L. japonicus* and *L. corniculatus*, whereas other mutants formed effective nodules (Kelly et al., 2013). In particular, R7A mutants deficient in production of an acidic octasaccharide EPS were able to normally nodulate *L. japonicus*, while *exoU* mutants producing a truncated pentasaccharide EPS failed to invade the host. It was proposed that full-length EPS serves as a signal to compatible hosts to modulate plant defense responses and allow bacterial infection, and R7A mutants that make no EPS could avoid or suppress the plant surveillance system and therefore retain the ability to form nodules. In contrast, strains that produce modified or truncated EPS trigger plant defense responses resulting in block of infection (Kelly et al., 2013).

EPS production is common in rhizobial bacteria, and the composition of EPS produced by different species varies widely (Skorupska et al., 2006). Several studies have suggested the involvement of the EPS structures in determining infective specificity (Hotter and Scott, 1991; Kannenberg et al., 1992; Parniske et al., 1994; Kelly et al., 2013). Recently, an EPS receptor (EPR3) has been identified in *L. japonicus*, which is a cell surface-localized protein containing three extracellular LysM domains and an intracellular kinase domain (Kawaharada et al., 2015). EPR3 binds rhizobial EPS in a structurally specific manner. Interestingly, *Epr3* gene expression is contingent on Nod-factor signaling, suggesting that the bacterial entry to the host is controlled by two successive steps of receptor-mediated recognition of Nod factor and EPS signals (Kawaharada et al., 2015, 2017). The receptor-ligand interaction supports the notion that EPS recognition plays a role in regulation of symbiosis specificity. However, natural variation in host-range specificity that results from specific recognition between host receptors and strain-specific EPS has not been demonstrated in any legume-rhizobial interactions. It is noteworthy that acidic EPS of bacterial pathogens also promote infection to cause plant disease (Newman et al., 1994; Yu et al., 1999; Aslam et al., 2008; Beattie, 2011). Thus, rhizobial EPS might also be recognized by host immune receptors to induce defense responses that negatively regulate symbiosis development.

SPECIFICITY MEDIATED BY HOST INNATE IMMUNITY

Symbiotic and pathogenic bacteria often produce similar signaling molecules to facilitate their invasion of the host (Deakin and Broughton, 2009). These molecules include conserved microbe-associated molecular patterns (MAMPs) and secreted effectors (D'Haeze and Holsters, 2004; Fauvart and Michiels, 2008; Deakin and Broughton, 2009; Soto et al., 2009; Downie,

2010; Wang et al., 2012; Okazaki et al., 2013). The host has evolved recognition mechanisms to distinguish between, and respond differently to, pathogens and symbionts (Bozsoki et al., 2017; Zipfel and Oldroyd, 2017). However, this discrimination is not always successful; as a result, recognition specificity frequently occurs in both pathogenic and symbiotic interactions. In the legume-rhizobial interaction, effector- or MAMP-triggered plant immunity mediated by host receptors also plays an important role in regulating host range of rhizobia (Yang et al., 2010; Wang et al., 2012; Faruque et al., 2015; Kawaharada et al., 2015; Tang et al., 2016).

Several dominant genes have been cloned in soybeans (e.g., *Rj2*, *Rfg1*, and *Rj4*) that restrict nodulation by specific rhizobial strains (Yang et al., 2010; Tang et al., 2016; Fan et al., 2017). In these cases, restriction of nodulation is controlled in a similar manner as 'gene-for-gene' resistance against plant pathogens. *Rj2* and *Rfg1* are allelic genes that encode a typical TIR-NBS-LRR resistance protein conferring resistance to multiple *B. japonicum* and *Sinorhizobium fredii* strains (Yang et al., 2010; Fan et al., 2017). *Rj4* encodes a thaumatin-like defense-related protein that restricts nodulation by specific strains of *B. elkanii* (Tang et al., 2016). The function of these genes is dependent on the bacterial type III secretion system and its secreted effectors (Krishnan et al., 2003; Okazaki et al., 2009; Yang et al., 2010; Tsukui et al., 2013; Tsurumaru et al., 2015; Tang et al., 2016; Yasuda et al., 2016). These studies indicate an important role of effector-triggered immunity in the regulation of nodulation specificity in soybeans. As discussed earlier, rhizobial Nod factors and surface polysaccharides could play a role in suppression of defense responses (Shaw and Long, 2003; D'Haeze and Holsters, 2004; Tellström et al., 2007; Jones et al., 2008; Liang et al., 2013; Cao et al., 2017), but these signaling events apparently are not strong enough to evade effector-trigged immunity in incompatible interactions.

Many rhizobial bacteria use the type III secretion system to deliver effectors into host cells to promote infection, and in certain situations, the delivered effector(s) are required for Nod-factor independent nodulation as demonstrated in the soybean-*B. elkanii* symbiosis (Deakin and Broughton, 2009; Okazaki et al., 2013, 2016). On the other hand, however, recognition of the effectors by host resistance genes triggers immune responses to restrict rhizobial infection. The nodulation resistance genes occur frequently in natural populations, raising a question why host evolve and maintain such seemingly unfavorable alleles. This could happen because of balancing selection, as the same alleles may also contribute to disease resistance against pathogens, considering that some rhizobial effectors are homologous to those secreted by bacterial pathogens (Dai et al., 2008; Kambara et al., 2009). Alternatively, legume may take advantage of *R* genes to exclude nodulation with less efficient nitrogen-fixing strains and selectively interact with strains with high nitrogen fixation efficiency, which is the case of the soybean *Rj4* allele.

A single dominant locus, called *NS1*, was also identified in *M. truncatula* that restricts nodulation by *S. meliloti* strain Rm41 (Liu et al., 2014). Unlike *R* gene-controlled host specificity in

soybeans, which depends on bacterial type III secretion system, Rm41 strain lacks genes encoding such a system. It will be interesting to know what host gene(s) control this specificity and what bacterial signals are involved.

SPECIFICITY IN NITROGEN FIXATION

Symbiotic specificity is not confined to the early recognition stages; incompatible host-strain combinations can lead to formation of nodules that are defective in nitrogen fixation (Fix-). For example, a screen of a core collection of *Medicago* accessions using multiple *S. meliloti* strains showed that ∼40% of the plant-strain combinations produced nodules but failed to fix nitrogen (Liu et al., 2014). The Fix- phenotype was not due to a lack of infection but caused by bacteroid degradation after differentiation (Yang et al., 2017; Wang et al., 2017, 2018).

Host genetic control of nitrogen fixation specificity is very complicated in the *Medicago-Sinorhizobium* symbiosis, involving multiple linked loci with complex epistatic and allelic interactions. By using the residual heterozygous lines identified from a recombination inbred line population, Zhu and colleagues were able to clone two of the underlying genes, namely *NFS1* and *NFS2*, that regulate strain-specific nitrogen fixation concerning the *S. meliloti* strains Rm41 and A145 (Wang et al., 2017, 2018; Yang et al., 2017). *NFS1* and *NFS2* both encode nodule-specific cysteine-rich (NCR) peptides (Mergaert et al., 2003). The NFS1 and NFS2 peptides function to provoke bacterial cell death and early nodule senescence in an allele-specific and rhizobial strain-specific manner, and their function is dependent on host genetic background. NCRs were previously shown to be positive regulators of symbiotic development, essential for terminal bacterial differentiation and for maintenance of bacterial survival in the nodule cells (Van de Velde et al., 2010; Wang et al., 2010; Horváth et al., 2015; Kim et al., 2015). The discovery of NFS1 and NFS2 revealed a negative role that NCRs play in regulation of symbiotic persistence, and showed that NCRs are host determinants of symbiotic specificity in *M. truncatula* and possibly also in closely related legumes that subject their symbiotic bacteria to terminal differentiation.

The genomes of *M. truncatula* and closely related galegoid legumes contain a large number of NCR-encoding genes that are expressed exclusively in the infected nodule cells (Montiel et al., 2017). These *NCR* genes, similar to bacterial type III effectors or MAMPs, can play both positive and negative roles in symbiotic development and both roles are associated with the antimicrobial property of the peptides. On one hand, the host uses this antimicrobial strategy for promoting terminal bacteroid differentiation to enhance nitrogen fixation efficiency (Oono and Denison, 2010; Oono et al., 2010; Van de Velde et al., 2010; Wang et al., 2010). On the other hand, some rhizobial strains cannot survive the antibacterial activity of certain peptide isoforms. The vulnerability of particular bacterial strains in response to a peptide is contingent on the genetic constitution of the bacteria as well as the genetic background of the host. It was proposed that this host-strain adaptation drives the coevolution
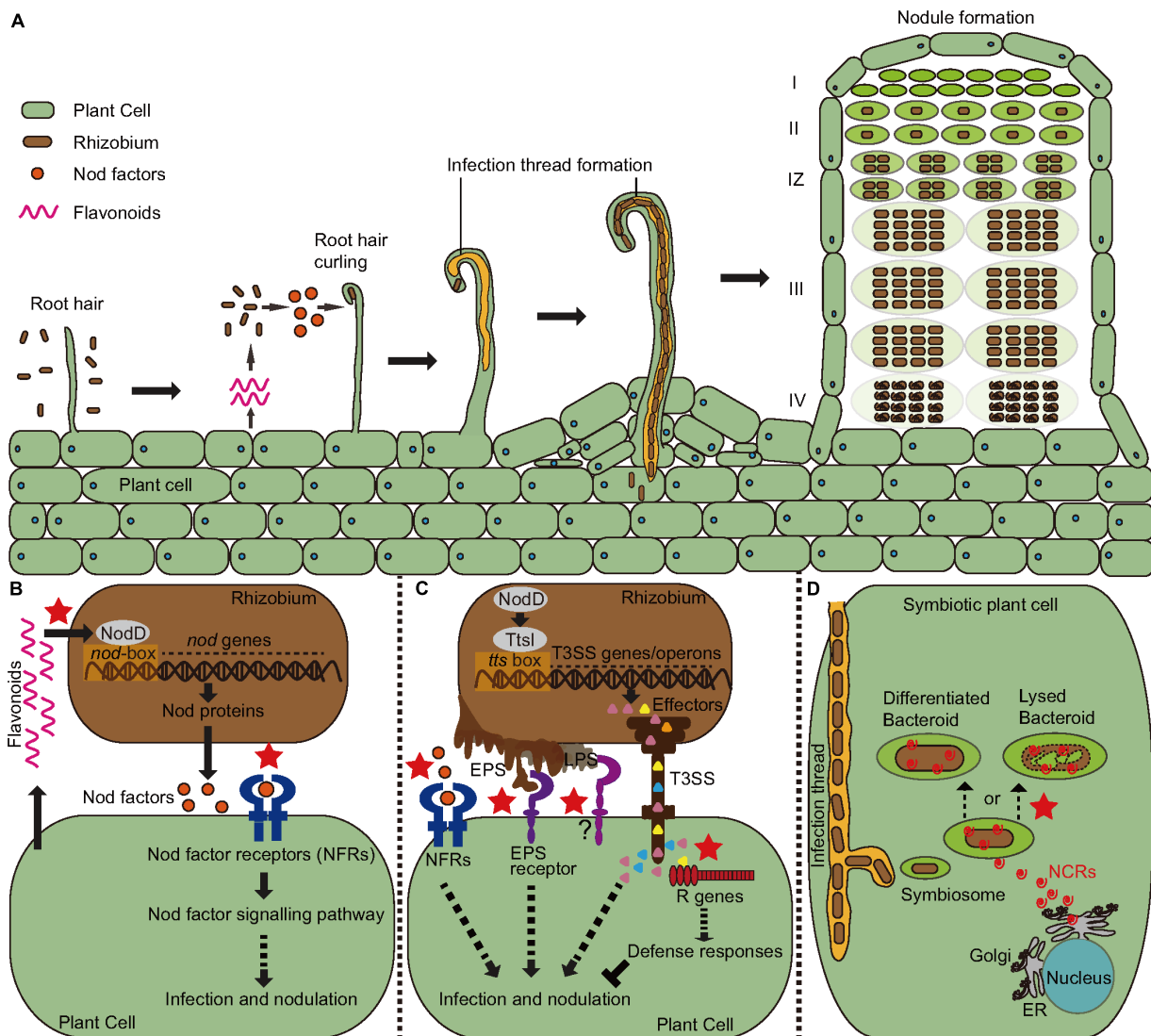
FIGURE 1 | Symbiosis signaling and plant immunity involved in recognition specificity in the legume-rhizobial interactions (indicated by the red stars).
(A) The process of infection and nodule development. A mature indeterminate nodule contains a meristem zone (I), an infection zone (II), an interzone (IZ), a nitrogen fixing zone (III), and a senescent zone (IV). **(B)** The host secretes flavonoids to induce the expression of bacterial nodulation (*nod*) gene through the activation of NodD proteins. The enzymes encoded by the *nod* genes lead to the synthesis of Nod factors (NF) that are recognized by host Nod factor receptors (NFRs). Recognition specificity occurs both between Flavonoids and NodDs and between NF and NFRs. **(C)** In addition to NF signaling, bacteria also produce extracellular polysaccharides (EPS) and type III effectors to facilitate their infection in compatible interactions, but these molecules may also induce immune responses causing resistance to infection in incompatible interactions. **(D)** Certain legumes such as *Medicago* encode antimicrobial nodule-specific cysteine-rich (NCR) peptides to drive their bacterial partners to terminal differentiation that is required for nitrogen fixation. However, some rhizobial strains cannot survive the antibacterial activity of certain peptide isoforms, leading to formation of nodules defective in nitrogen fixation.

of both symbiotic partners, leading to the rapid amplification and diversification of the *NCR* genes in galegoid legumes (Wang et al., 2017; Yang et al., 2017).

Host-range specificity in the ability to fix nitrogen has also been documented in legumes (e.g., soybeans) where the bacteria undergo reversible differentiation. In soybeans, this type of incompatibility was associated with the induction of phytoalexin accumulation and hypersensitive reaction in the nodule cells (Parniske et al., 1990). No *NCR* genes exist in the soybean genome, implying the involvement of novel genetic mechanisms

that control this specificity. Work is in progress in our lab to identify the host genes that are involved.

CONCLUSION AND FUTURE PERSPECTIVES

Specificity in the legume-rhizobial symbiosis results from a suite of signal exchanges between the two symbiotic partners (summarized in **Figure 1**). Recent studies have just begun to

reveal the underlying molecular mechanisms that regulate this specificity, and there are many challenging questions waiting to be answered. Effector-triggered immunity has been shown to be an important factor in determining host range of rhizobia in soybeans but the cognate effectors have not been clearly defined. In addition, what are the genes that control nodulation specificity in the *Medicago-Sinorhizobium* interaction where the bacterial partner lacks the type III secretion system? Cloning and characterization of the *NS1* locus in *M. truncatula* (Liu et al., 2014) will provide novel insights into this question. We now know that NCR peptides regulate nitrogen fixation specificity in *Medicago* and possibly in other closely related legumes, but we lack mechanistic understanding of how these peptides work. Do the pro- and anti-symbiotic peptides interact with the same bacterial targets? How do the amino-acid substitutions affect the peptide structure and function? How is nitrogen fixation specificity regulated in the NCR-lacking legumes such as soybeans where bacteria undergo reversible differentiation? These are just a handful of outstanding questions that need to be addressed. Answering these questions will certainly enrich our knowledge of how specificity is controlled and allow us to use such knowledge to develop tools for genetic improvement of symbiotic nitrogen fixation in legumes.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

REFERENCES

Aslam, S. N., Newman, M. A., Erbs, G., Morrissey, K. L., Chinchilla, D., Boller, T., et al. (2008). Bacterial polysaccharides suppress induced innate immunity by calcium chelation. *Curr. Biol.* 18, 1078–1083. doi: 10.1016/j.cub.2008.06.061

Beattie, G. A. (2011). Water relations in the interaction of foliar bacterial pathogens with plants. *Annu. Rev. Phytopathol.* 49, 533–555. doi: 10.1146/annurev-phyto-073009-114436

Becker, A., Fraysse, N., and Sharypova, L. (2005). Recent advances in studies on structure and symbiosis-related function of rhizobial K-antigens and lipopolysaccharides. *Mol. Plant Microbe Interact.* 18, 899–905. doi: 10.1094/MPMI-18-0899

Bender, G. L., Nayudu, M., Le Strange, K. K., and Rolfe, B. G. (1988). The nodD1 gene from Rhizobium strain NGR234 is a key determinant in the extension of host range to the nonlegume *Parasponia*. *Mol. Plant Microbe Interact.* 1, 259–266. doi: 10.1094/MPMI-1-259

Bloemberg, G. V., Kamst, E., Harteveld, M., Drift, K. M., Haverkamp, J., Thomas-Oates, J. E., et al. (1995). A central domain of *Rhizobium* NodE protein mediates host specificity by determining the hydrophobicity of fatty acyl moieties of nodulation factors. *Mol. Microbiol.* 16, 1123–1136. doi: 10.1111/j.1365-2958.1995.tb02337.x

Bozsoki, Z., Cheng, J., Feng, F., Gysel, K., Vinther, M., Andersen, K. R., et al. (2017). Receptor-mediated chitin perception in legume roots is functionally separable from Nod factor perception. *Proc. Natl. Acad. Sci. U.S.A.* 114, 8118–8127. doi: 10.1073/pnas.1706795114

Breakspear, A., Liu, C., Roy, S., Stacey, N., Rogers, C., Trick, M., 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. doi: 10.1105/tpc.114.133496

Broghammer, A., Krusell, L., Blaise, M., Sauer, J., Sullivan, J. T., Maolanon, N., et al. (2012). Legume receptors perceive the rhizobial lipochitin oligosaccharide signal molecules by direct binding. *Proc. Natl. Acad. Sci. U.S.A.* 109, 13859–13864. doi: 10.1073/pnas.1205171109

Broughton, W. J., Jabbouri, S., and Perret, X. (2000). Keys to symbiotic harmony. *J. Bacteriol.* 182, 5641–5652. doi: 10.1128/JB.182.20.5641-5652.2000

Cao, Y., Halane, M. K., Gassmann, W., and Stacey, G. (2017). The role of plant innate immunity in the legume-rhizobium symbiosis. *Annu. Rev. Plant Biol.* 68, 535–561. doi: 10.1146/annurev-arplant-042916-041030

Cheng, H. P., and Walker, G. C. (1998). Succinoglycan is required for initiation and elongation of infection threads during nodulation of alfalfa by *Rhizobium meliloti*. *J. Bacteriol.* 180, 5183–5191.

Dai, W. J., Zeng, Y., Xie, Z. P., and Staehelin, C. (2008). Symbiosis-promoting and deleterious effects of NopT, a novel type 3 effector of *Rhizobium* sp. strain NGR234. *J. Bacteriol.* 190, 5101–5110. doi: 10.1128/JB.00306-08

Deakin, W. J., and Broughton, W. J. (2009). Symbiotic use of pathogenic strategies: rhizobial protein secretion systems. *Nat. Rev. Microbiol.* 7, 312–321. doi: 10.1038/nrmicro2091

Dénarié, J., Debellé, F., and Promé, J. C. (1996). Rhizobium lipo-chitooligosaccharide nodulation factors: signaling molecules mediating recognition and morphogenesis. *Annu. Rev. Biochem.* 65, 503–535. doi: 10.1146/annurev.bi.65.070196.002443

D'Haeze, W., and Holsters, M. (2002). Nod factor structures, responses, and perception during initiation of nodule development. *Glycobiology* 12, 79R–105R. doi: 10.1093/glycob/12.6.79R

D'Haeze, W., and Holsters, M. (2004). Surface polysaccharides enable bacteria to evade plant immunity. *Trends Microbiol.* 12, 555–561. doi: 10.1016/j.tim.2004.10.009

Downie, J. A. (2010). The roles of extracellular proteins, polysaccharides and signals in the interactions of rhizobia with legume roots. *FEMS Microbiol. Rev.* 34, 150–170. doi: 10.1111/j.1574-6976.2009.00205.x

Fan, Y., Liu, J., Lyu, S., Wang, Q., Yang, S., and Zhu, H. (2017). The soybean *Rfg1* gene restricts nodulation by *Sinorhizobium fredii* USDA193. *Front. Plant Sci.* 8:1548. doi: 10.3389/fpls.2017.01548

Faruque, O. M., Miwa, H., Yasuda, M., Fujii, Y., Kaneko, T., Sato, S., et al. (2015). Identification of *Bradyrhizobium elkanii* genes involved in incompatibility with soybean plants carrying the *Rj4* allele. *Appl. Environ. Microbiol.* 81, 6710–6717. doi: 10.1128/AEM.01942-15

Fauvart, M., and Michiels, J. (2008). Rhizobial secreted proteins as determinants of host specificity in the rhizobium–legume symbiosis. *FEMS Microbiol. Lett.* 28, 1–9. doi: 10.1111/j.1574-6968.2008.01254.x

Firmin, J. L., Wilson, K. E., Carlson, R. W., Davies, A. E., and Downie, J. A. (1993). Resistance to nodulation of cv. Afghanistan peas is overcome by *nodX*, which mediates an O-acetylation of the *Rhizobium leguminosarum* lipo-oligosaccharide nodulation factor. *Mol. Microbiol.* 10, 351–360. doi: 10.1111/j.1365-2958.1993.tb01961.x

Fisher, R. F., Egelhoff, T. T., Mulligan, J. T., and Long, S. R. (1988). Specific binding of proteins from *Rhizobium meliloti* cell-free extracts containing NodD to DNA sequences upstream of inducible nodulation genes. *Genes Dev.* 2, 282–293. doi: 10.1101/gad.2.3.282

Fraysse, N., Couderc, F., and Poinsot, V. (2003). Surface polysaccharide involvement in establishing the rhizobium–legume symbiosis. *FEBS J.* 270, 1365–1380. doi: 10.1046/j.1432-1033.2003.03492.x

Gibson, K. E., Kobayashi, H., and Walker, G. C. (2008). Molecular determinants of a symbiotic chronic infection. *Annu. Rev. Genet.* 42, 413–441. doi: 10.1146/annurev.genet.42.110807.091427

Haag, A. F., Arnold, M. F., Myka, K. K., Kerscher, B., Dall'Angelo, S., Zanda, M., et al. (2013). Molecular insights into bacteroid development during *Rhizobium*–legume symbiosis. *FEMS Microbial. Rev.* 37, 364–383. doi: 10.1111/1574-6976.12003

Hirsch, A. M. (1992). Developmental biology of legume nodulation. *New Phytol.* 122, 211–237. doi: 10.1111/j.1469-8137.1992.tb04227.x

Horváth, B., Bachem, C. W., Schell, J., and Kondorosi, A. (1987). Host-specific regulation of nodulation genes in *Rhizobium* is mediated by a plant-signal, interacting with the *nodD* gene product. *EMBO J.* 6, 841–848. doi: 10.1002/j.1460-2075.1987.tb04829.x

Horváth, B., Domonkos, Á., Kereszt, A., Szűcs, A., Ábrahám, E., Ayaydin, F., et al. (2015). Loss of the nodule-specific cysteine rich peptide, NCR169, abolishes symbiotic nitrogen fixation in the *Medicago truncatula* dnf7 mutant. *Proc. Natl. Acad. Sci. U.S.A* 112, 15232–15237. doi: 10.1073/pnas.1500777112

Hotter, G. S., and Scott, D. B. (1991). Exopolysaccharide mutants of Rhizobium loti are fully effective on a determinate nodulating host but are ineffective on an indeterminate nodulating host. *J. Bacteriol.* 173, 851–859. doi: 10.1128/jb.173.2.851-859.1991

Jones, K. M. (2012). Increased production of the exopolysaccharide succinoglycan enhances *Sinorhizobium meliloti* 1021 symbiosis with the host plant *Medicago truncatula*. *J. Bacteriol.* 194, 4322–4331. doi: 10.1128/JB.00751-12

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. *Nat. Rev. Microbiol.* 5, 619–633. doi: 10.1038/nrmicro1705

Jones, K. M., Sharopova, N., Lohar, D. P., Zhang, J. Q., VandenBosch, K. A., and Walker, G. C. (2008). Differential response of the plant *Medicago truncatula* to its symbiont *Sinorhizobium meliloti* or an exopolysaccharide-deficient mutant. *Proc. Natl. Acad. Sci. U.S.A.* 105, 704–709. doi: 10.1073/pnas.0709338105

Kambara, K., Ardissone, S., Kobayashi, H., Saad, M. M., Schumpp, O., Broughton, W. J., et al. (2009). Rhizobia utilize pathogen-like effector proteins during symbiosis. *Mol. Microbiol.* 71, 92–106. doi: 10.1111/j.1365-2958.2008.06507.x

Kannenberg, E. L., Rathbun, E. A., and Brewin, N. J. (1992). Molecular dissection of structure and function in the lipopolysaccharide of *Rhizobium leguminosarum* strain 3841 using monoclonal antibodies and genetic analysis. *Mol. Microbiol.* 6, 2477–2487. doi: 10.1111/j.1365-2958.1992.tb01424.x

Kawaharada, Y., Kelly, S., Nielsen, M. W., Hjuler, C. T., Gysel, K., Muszynski, A., et al. (2015). Receptor-mediated exopolysaccharide perception controls bacterial infection. *Nature* 523, 308–312. doi: 10.1038/nature14611

Kawaharada, Y., Nielsen, M. W., Kelly, S., James, E. K., Andersen, K. R., Rasmussen, S. R., et al. (2017). Differential regulation of the Epr3 receptor coordinates membrane-restricted rhizobial colonization of root nodule primordia. *Nat. Commun.* 8:14534. doi: 10.1038/ncomms14534

Kelly, S. J., Muszyński, A., Kawaharada, Y., Hubber, A. M., Sullivan, J. T., Sandal, N., et al. (2013). Conditional requirement for exopolysaccharide in the *Mesorhizobium–Lotus* symbiosis. *Mol. Plant Microbe Interact.* 26, 319–329. doi: 10.1094/MPMI-09-12-0227-R

Kereszt, A., Mergaert, P., and Kondorosi, E. (2011). Bacteroid development in legume nodules: evolution of mutual benefit or of sacrificial victims? *Mol. Plant Microbe Interact.* 24, 1300–1309. doi: 10.1094/MPMI-06-11-0152

Kim, M., Chen, Y., Xi, J., Waters, C., Chen, R., and Wang, D. (2015). An antimicrobial peptide essential for bacterial survival in the nitrogen-fixing symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* 112, 15238–15243. doi: 10.1073/pnas.1500123112

Krishnan, H. B., Lorio, J., Kim, W. S., Jiang, G., Kim, K. Y., DeBoer, M., et al. (2003). Extracellular proteins involved in soybean cultivar-specific nodulation are associated with pilus-like surface appendages and exported by a type III protein secretion system in *Sinorhizobium fredii* USDA257. *Mol. Plant Microbe Interact.* 16, 617–625. doi: 10.1094/MPMI.2003.16.7.617

Leigh, J. A., Signer, E. R., and Walker, G. C. (1985). Exopolysaccharide-deficient mutants of *Rhizobium meliloti* that form ineffective nodules. *Proc. Natl. Acad. Sci. U.S.A.* 82, 6231–6235. doi: 10.1073/pnas.82.18.6231

Li, R., Knox, M. R., Edwards, A., Hogg, B., Ellis, T. N., Wei, G., et al. (2011). Natural variation in host-specific nodulation of pea is associated with a haplotype of the SYM37 LysM-type receptor-like kinase. *Mol. Plant-Microbe Interact.* 24, 1396–1403. doi: 10.1094/MPMI-01-11-0004

Liang, Y., Cao, Y., Tanaka, K., Thibivilliers, S., Wan, J., Choi, J., et al. (2013). Nonlegumes respond to rhizobial Nod factors by suppressing the innate immune response. *Science* 341, 1384–1387. doi: 10.1126/science.1242736

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. doi: 10.1126/science.1090074

Liu, C. W., and Murray, J. D. (2016). The role of flavonoids in nodulation host-range specificity: an update. *Plants* 5:E33. doi: 10.3390/plants5030033

Liu, J., Yang, S., Zheng, Q., and Zhu, H. (2014). Identification of a dominant gene in *Medicago truncatula* that restricts nodulation by *Sinorhizobium meliloti* strain Rm41. *BMC Plant Biol.* 14:167. doi: 10.1186/1471-2229-14-167

Long, S. R. (1996). Rhizobium symbiosis: nod factors in perspective. *Plant Cell* 8, 1885–1898. doi: 10.1105/tpc.8.10.1885

Madsen, E. B., Madsen, L. H., Radutoiu, S., Olbryt, M., Rakwalska, M., Szczyglowski, K., et al. (2003). A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals. *Nature* 425, 637–640. doi: 10.1038/nature02045

Mergaert, P., Nikovics, K., Kelemen, Z., Maunoury, N., Vaubert, D., Kondorosi, A., et al. (2003). A novel family in *Medicago truncatula* consisting of more than 300 nodule-specific genes coding for small, secreted polypeptides with conserved cysteine motifs. *Plant Physiol.* 132, 161–173. doi: 10.1104/pp.102.018192

Mergaert, P., Uchiumi, T., Alunni, B., Evanno, G., Cheron, A., Catrice, O., et al. (2006). Eukaryotic control on bacterial cell cycle and differentiation in the *Rhizobium*–legume symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* 103, 5230–5235. doi: 10.1073/pnas.0600912103

Montiel, J., Downie, J. A., Farkas, A., Bihari, P., Herczeg, R., Bálint, B., et al. (2017). Morphotype of bacteroids in different legumes correlates with the number and type of symbiotic NCR peptides. *Proc. Natl. Acad. Sci. U.S.A.* 114, 5041–5046. doi: 10.1073/pnas.1704217114

Nap, J. P., and Bisseling, T. (1990). Developmental biology of a plant-prokaryote symbiosis: the legume root nodule. *Science* 250, 948–954. doi: 10.1126/science.250.4983.948

Newman, M. A., Conrads-Strauch, J., Scofield, G., Daniels, M. J., and Dow, J. M. (1994). Defense-related gene induction in *Brassica campestris* in response to defined mutants of *Xanthomonas campestris* with altered pathogenicity. *Mol. Plant Microbe Interact.* 7, 553–563. doi: 10.1094/MPMI-7-0553

Okazaki, S., Kaneko, T., Sato, S., and Saeki, K. (2013). Hijacking of leguminous nodulation signaling by the rhizobial type III secretion system. *Proc. Natl. Acad. Sci. U.S.A.* 110, 17131–17136. doi: 10.1073/pnas.1302360110

Okazaki, S., Tittabutr, P., Teulet, A., Thouin, J., Fardoux, J., Chaintreuil, C., et al. (2016). Rhizobium–legume symbiosis in the absence of Nod factors: two possible scenarios with or without the T3SS. *ISME J.* 10, 64–74. doi: 10.1038/ismej.2015.103

Okazaki, S., Zehner, S., Hempel, J., Lang, K., and Göttfert, M. (2009). Genetic organization and functional analysis of the type III secretion system of *Bradyrhizobium elkanii*. *FEMS Microbiol. Lett.* 295, 88–95. doi: 10.1111/j.1574-6968.2009.01593.x

Oldroyd, G. E. (2013). Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nat. Rev. Microbiol.* 11, 252–264. doi: 10.1038/nrmicro2990

Oldroyd, G. E., Murray, J. D., Poole, P. S., and Downie, J. A. (2011). The rules of engagement in the legume-rhizobial symbiosis. *Annu. Rev. Genet.* 45, 119–144. doi: 10.1146/annurev-genet-110410-132549

Oono, R., and Denison, R. F. (2010). Comparing symbiotic efficiency between swollen versus nonswollen rhizobial bacteroids. *Plant Physiol.* 154, 1541–1548. doi: 10.1104/pp.110.163436

Oono, R., Schmitt, I., Sprent, J. I., and Denison, R. F. (2010). Multiple evolutionary origins of legume traits leading to extreme rhizobial differentiation. *New Phytol.* 187, 508–520. doi: 10.1111/j.1469-8137.2010.03261.x

Pankhurst, C. E., and Biggs, D. R. (1980). Sensitivity of *Rhizobium* to selected isoflavonoids. *Can. J. Microbiol* 26, 542–545. doi: 10.1139/m80-092

Parniske, M., Ahlborn, B., and Werner, D. (1991). Isoflavonoid-inducible resistance to the phytoalexin glyceollin in soybean rhizobia. *J. Bacteriol.* 173, 3432–3439. doi: 10.1128/jb.173.11.3432-3439.1991

Parniske, M., Schmidt, P., Kosch, K., and Müller, P. (1994). Plant defense responses of host plants with determinate nodules induced by EPS-defective exoB mutants of *Bradyrhizobium japonicum*. *Mol. Plant Microbe Interact.* 7, 631–638. doi: 10.1094/MPMI-7-0631

Parniske, M., Zimmermann, C., Cregan, P. B., and Werner, D. (1990). Hypersensitive reaction of nodule cells in the *Glycine* sp./*Bradyrhizobium japonicum*-symbiosis occurs at the genotype-specific level. *Plant Biol.* 103, 143–148. doi: 10.1111/j.1438-8677.1990.tb00140.x

Peck, M. C., Fisher, R. F., and Long, S. R. (2006). Diverse flavonoids stimulate NodD1 binding to nod gene promoters in *Sinorhizobium meliloti. J. Bacteriol.* 188, 5417–5427. doi: 10.1128/JB.00376-06

Perret, X., Staehelin, C., and Broughton, W. J. (2000). Molecular basis of symbiotic promiscuity. *Microbiol. Mol. Biol. Rev.* 64, 180–201. doi: 10.1128/MMBR.64.1.180-201.2000

Prell, J., and Poole, P. (2006). Metabolic changes of rhizobia in legume nodules. *Trends Microbiol.* 14, 161–168. doi: 10.1016/j.tim.2006.02.005

Radutoiu, S., Madsen, L. H., Madsen, E. B., Felle, H. H., Umehara, Y., Grønlund, M., et al. (2003). Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *Nature* 425, 585–592. doi: 10.1038/nature02039

Radutoiu, S., Madsen, L. H., Madsen, E. B., Jurkiewicz, A., Fukai, E., Quistgaard, E. M., et al. (2007). LysM domains mediate lipochitin–oligosaccharide recognition and Nfr genes extend the symbiotic host range. *EMBO J.* 26, 3923–3935. doi: 10.1038/sj.emboj.7601826

Reinhold, B. B., Chan, S. Y., Reuber, T. L., Marra, A., Walker, G. C., and Reinhold, V. N. (1994). Detailed structural characterization of succinoglycan, the major exopolysaccharide of *Rhizobium meliloti* Rm1021. *J. Bacteriol.* 176, 1997–2002. doi: 10.1128/jb.176.7.1997-2002.1994

Rostas, K., Kondorosi, E., Horvath, B., Simoncsits, A., and Kondorosi, A. (1986). Conservation of extended promoter regions of nodulation genes in *Rhizobium. Proc. Natl. Acad. Sci. U.S.A.* 83, 1757–1761. doi: 10.1073/pnas.83.6.1757

Shaw, S. L., and Long, S. R. (2003). Nod factor inhibition of reactive oxygen efflux in a host legume. *Plant Physiol.* 132, 2196–2204. doi: 10.1104/pp.103.021113

Skorupska, A., Janczarek, M., Marczak, M., Mazur, A., and Król, J. (2006). Rhizobial exopolysaccharides: genetic control and symbiotic functions. *Microb. Cell Fact.* 5:7. doi: 10.1186/1475-2859-5-7

Soto, M. J., Domínguez-Ferreras, A., Pérez-Mendoza, D., Sanjuán, J., and Olivares, J. (2009). Mutualism versus pathogenesis: the give-and-take in plant–bacteria interactions. *Cell. Microbiol.* 11, 381–388. doi: 10.1111/j.1462-5822.2009.01282.x

Tang, F., Yang, S., Liu, J., and Zhu, H. (2016). Rj4, a gene controlling nodulation specificity in soybeans, encodes a thaumatin-like protein, but not the one previously reported. *Plant Physiol.* 170, 26–32. doi: 10.1104/pp.15.01661

Tellström, V., Usadel, B., Thimm, O., Stitt, M., Küster, H., and Niehaus, K. (2007). The lipopolysaccharide of *Sinorhizobium meliloti* suppresses defense-associated gene expression in cell cultures of the host plant *Medicago truncatula. Plant Physiol.* 143, 825–837. doi: 10.1104/pp.106.090985

Tsukui, T., Eda, S., Kaneko, T., Sato, S., Okazaki, S., Kakizaki-Chiba, K., et al. (2013). The type III secretion system of *Bradyrhizobium japonicum* USDA122 mediates symbiotic incompatibility with Rj2 soybean plants. *Appl. Environ. Microbiol.* 79, 1048–1051. doi: 10.1128/AEM.03297-12

Tsurumaru, H., Hashimoto, S., Okizaki, K., Kanesaki, Y., Yoshikawa, H., and Yamakawa, T. (2015). A putative type III secretion system effector encoded by the MA20_12780 gene in *Bradyrhizobium japonicum* Is-34 causes

incompatibility with Rj4 genotype soybeans. *Appl. Environ. Microbiol.* 81, 5812–5819. doi: 10.1128/AEM.00823-15

Van de Velde, W., Zehirov, G., Szatmari, A., Debreczeny, M., Ishihara, H., Kevei, Z., et al. (2010). Plant peptides govern terminal differentiation of bacteria in symbiosis. *Science* 327, 1122–1126. doi: 10.1126/science.1184057

Wang, D., Griffitts, J., Starker, C., Fedorova, E., Limpens, E., Ivanov, S., et al. (2010). A nodule specific protein secretory pathway required for nitrogen-fixing symbiosis. *Science* 327, 1126–1129. doi: 10.1126/science.1184096

Wang, D., Yang, S., Tang, F., and Zhu, H. (2012). Symbiosis specificity in the legume–rhizobial mutualism. *Cell Microbiol.* 14, 334–342. doi: 10.1111/j.1462-5822.2011.01736

Wang, Q., Liu, J., Li, H., Yang, S., Körmöczi, P., Kereszt, A., et al. (2018). Nodule-specific cysteine-rich peptides negatively regulate nitrogen-fixing symbiosis in a strain-specific manner in *Medicago truncatula. Mol. Plant Microbe Interact.* 31, 240–248. doi: 10.1094/MPMI-08-17-0207-R

Wang, Q., Yang, S., Liu, J., Terecskei, K., Ábrahám, E., Gombár, A., et al. (2017). Host-secreted antimicrobial peptide enforces symbiotic selectivity in *Medicago truncatula. Proc. Natl. Acad. Sci. U.S.A.* 114, 6854–6859. doi: 10.1073/pnas.1700715114

Yang, S., Tang, F., Gao, M., Krishnan, H. B., and Zhu, H. (2010). R gene-controlled host specificity in the legume–rhizobia symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18735–18740. doi: 10.1073/pnas.1011957107

Yang, S., Wang, Q., Fedorova, E., Liu, J., Qin, Q., Zheng, Q., et al. (2017). Microsymbiont discrimination mediated by a host-secreted peptide in *Medicago truncatula. Proc. Natl. Acad. Sci. U.S.A.* 114, 6848–6853. doi: 10.1073/pnas.1700460114

Yasuda, M., Miwa, H., Masuda, S., Takebayashi, Y., Sakakibara, H., and Okazaki, S. (2016). Effector-triggered immunity determines host genotype-specific incompatibility in legume–rhizobium symbiosis. *Plant Cell Physiol.* 57, 1791–1800. doi: 10.1093/pcp/pcw104

Yu, J., Peñaloza-Vázquez, A., Chakrabarty, A. M., and Bender, C. L. (1999). Involvement of the exopolysaccharide alginate in the virulence and epiphytic fitness of *Pseudomonas syringae* pv. syringae. *Mol. Microbiol* 33, 712–720. doi: 10.1046/j.1365-2958.1999.01516.x

Zhukov, V., Radutoiu, S., Madsen, L. H., Rychagova, T., Ovchinnikova, E., Borisov, A., et al. (2008). The pea Sym37 receptor kinase gene controls infection-thread initiation and nodule development. *Mol. Plant Microbe Interact.* 21, 1600–1608. doi: 10.1094/MPMI-21-12-1600

Zipfel, C., and Oldroyd, G. E. (2017). Plant signalling in symbiosis and immunity. *Nature* 543, 328–336. doi: 10.1038/nature22009

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The Hydrophobin-Like OmSSP1 May Be an Effector in the Ericoid Mycorrhizal Symbiosis

Salvatore Casarrubia[1], Stefania Daghino[1], Annegret Kohler[2], Emmanuelle Morin[2], Hassine-Radhouane Khouja[1], Yohann Daguerre[2], Claire Veneault-Fourrey[2,3], Francis M. Martin[2], Silvia Perotto[1] and Elena Martino[1,2]*

[1] Department of Life Sciences and Systems Biology, University of Turin, Turin, Italy, [2] INRA (Institut National de la Recherche Agronomique), UMR 1136 INRA-Université de Lorraine Interactions Arbres/Microorganismes, Laboratoire d'Excellence ARBRE, Centre INRA-Lorraine, Champenoux, France, [3] Université de Lorraine, UMR 1136 INRA-Université de Lorraine Interactions Arbres/Microorganismes, Laboratoire d'Excellence ARBRE, Faculté des Sciences et Technologies, Vandoeuvre les Nancy, France

Mutualistic and pathogenic plant-colonizing fungi use effector molecules to manipulate the host cell metabolism to allow plant tissue invasion. Some small secreted proteins (SSPs) have been identified as fungal effectors in both ectomycorrhizal and arbuscular mycorrhizal fungi, but it is currently unknown whether SSPs also play a role as effectors in other mycorrhizal associations. Ericoid mycorrhiza is a specific endomycorrhizal type that involves symbiotic fungi mostly belonging to the Leotiomycetes (Ascomycetes) and plants in the family Ericaceae. Genomic and RNASeq data from the ericoid mycorrhizal fungus *Oidiodendron maius* led to the identification of several symbiosis-upregulated genes encoding putative SSPs. OmSSP1, the most highly symbiosis up-regulated SSP, was found to share some features with fungal hydrophobins, even though it lacks the Pfam hydrophobin domain. Sequence alignment with other hydrophobins and hydrophobin-like fungal proteins placed OmSSP1 within Class I hydrophobins. However, the predicted features of OmSSP1 may suggest a distinct type of hydrophobin-like proteins. The presence of a predicted signal peptide and a yeast-based signal sequence trap assay demonstrate that OmSSP1 is secreted. OmSSP1 null-mutants showed a reduced capacity to form ericoid mycorrhiza with *Vaccinium myrtillus* roots, suggesting a role as effectors in the ericoid mycorrhizal interaction.

Keywords: ericoid mycorrhiza, *Oidiodendron maius*, small secreted proteins, hydrophobins, homologous recombination

INTRODUCTION

Fungi secrete a wide range of enzymatic and non-enzymatic proteins that function in the break-down of complex organic molecules but also in the interaction with microbial competitors or with animal and plant partners (Stergiopoulos and de Wit, 2009; Tian et al., 2009; Scherlach et al., 2013; Talbot et al., 2013). Fungi can establish different types of interactions with plants, ranging from mutualistic to antagonistic. Whatever their lifestyle, plant-colonizing fungi are recognized by the plant immune system through invariant molecular patterns known as microbe- or pathogen-associated molecular patterns (MAMPS or PAMPs) (Jones and Dangl, 2006). To

successfully colonize plant tissues, fungi must prevent the PAMP-triggered immunity (PTI) reaction (Lo Presti et al., 2015). For this purpose, fungi secrete effector molecules that may play different functions depending on the fungal lifestyles. For example, they can be toxic compounds that kill the host plant (in necrotrophs), or secreted proteins that shield the fungus and suppress the host immune response, or proteins that manipulate the host cell metabolism to allow plant tissue invasion and nutrient uptake (de Jonge et al., 2011; Giraldo et al., 2013; Selin et al., 2016). Many small secreted proteins (SSPs) have been reported to function as effectors (Lo Presti et al., 2015).

Effectors were initially considered as virulence factors secreted exclusively by pathogens (van Esse et al., 2008; Stergiopoulos and de Wit, 2009). However, it has become apparent that effectors can manipulate the plant immune system also in mutualistic associations (Kim et al., 2016). Mutualistic fungi establish intimate contacts with plants by forming specialized fungal structures involved in nutrient exchange with the host (Martin et al., 2017). Effector-like SSPs have been functionally characterized as effector-like molecules both in arbuscular (AM) and ectomycorrhizal (ECM) fungi, as well as in some endophytic fungi (Plett and Martin, 2015). For example, the ECM fungus *Laccaria bicolor* requires MiSSP7 (Mycorrhiza-induced Small Secreted Protein 7) to establish symbiosis. MiSSP7 suppresses the plant defense reactions by interacting with the jasmonate co-receptor JAZ6 (Plett et al., 2011, 2014). Similarly, the AM fungus *Rhizophagus irregularis* secretes SP7, an effector protein that counteracts the plant immune program by interacting with the pathogenesis-related transcription factor ERF19, leading to increased mycorrhization (Kloppholz et al., 2011). Tsuzuki et al. (2016) also showed that host-induced gene silencing of SlS1, a putative secreted *R. irregularis* SSP expressed in symbiosis, resulted in suppression of colonization and formation of stunted arbuscules. In a similar manner, the candidate effector protein PIIN_08944, secreted by the fungal endophyte *Piriformospora indica* during colonization of both *Arabidopsis* and barley plants, was demonstrated to play a crucial role in reducing the expression of PTI genes and of the salicylic acid defense pathway (Akum et al., 2015).

Protein effectors are normally secreted following the endoplasmic reticulum-Golgi apparatus pathway, and bioinformatic identification of effector candidates can thus be based on the presence of the N-terminal signal peptide (Lo Presti et al., 2015), even though alternative secretion pathways have been reported for *Magnaporthe oryzae* and *Phytopthora infestans* (Giraldo et al., 2013; Wang et al., 2017). General SSPs features are: the presence of a signal peptide and the absence of transmembrane domains or GPI-anchor sites; a small size, with a mature length smaller than 300 amino acids; a richness in cysteine residues and, sometimes, the presence of conserved motifs (Martin et al., 2008; Stergiopoulos and de Wit, 2009; Hacquard et al., 2012; Zuccaro et al., 2014; Lo Presti et al., 2015).

Bioinformatic analyses of about 50 fungal genomes has highlighted that, when compared with saprotrophic and pathogenic fungi, the ECM fungal secretome is enriched in SSPs and contains species-specific SSPs likely dedicated to the molecular cross-talk between fungal and plant partners (Pellegrin et al., 2015). In line with this finding, a comparative *in silico*

analysis of the AM fungi *Rhizophagus clarus, R. irregularis* and *Gigaspora rosea* highlighted the presence of shared SSPs (Sedzielewska Toro and Brachmann, 2016; Kamel et al., 2017), supporting a general conserved role of SSPs in AM. These data suggest that effector SSPs may represent an important fungal "toolkit" that enables the establishment/maintenance of host plant colonization in mycorrhiza (Plett and Martin, 2015; Martin et al., 2016). However, there is increasing awareness that SSPs in mycorrhizal fungi are likely involved in additional functions unrelated to symbiosis. For example, several SSPs are secreted by the ECM fungi *L. bicolor* and *Hebeloma cylindrosporum* during the free-living phase (Vincent et al., 2012; Doré et al., 2015). Moreover, large-scale transcriptomic and genomic analyses including fungi with different lifestyles revealed a wide array of SSPs in most saprotrophic fungi (Pellegrin et al., 2015; Valette et al., 2017), suggesting a possible role for SSPs in competition and rhizospheric communication (Rovenich et al., 2014).

Whereas effector SSPs have been characterized in AM and ECM fungi, there is currently no information on the occurrence of SSPs in ericoid mycorrhizal (ERM) fungi and on their potential role in symbiosis. ERM fungi are soil-borne fungi mostly belonging to Leotiomycetes (Ascomycetes). They form a peculiar endomycorrhizal type by colonizing the root epidermal cells of plants within the family Ericaceae and promote growth of their host plant in stressful habitats (Perotto et al., 2012). A common ERM fungal species is *Oidiodendron maius* (Dalpé, 1986) and *O. maius* strain Zn, an isolate from a metal polluted soil whose genome and transcriptome have been recently sequenced (Kohler et al., 2015), has become a model system to investigate metal stress tolerance in these fungi (Perotto et al., 2012; Daghino et al., 2016; Ruytinx et al., 2016). Genomic data have revealed a large set of carbohydrate-active enzymes (CAZymes) in *O. maius*, with many plant cell wall degrading enzymes being expressed during symbiosis (Kohler et al., 2015). As these enzymes could potentially elicit defense reactions through oligosaccharide release, symbiosis development likely requires a tight control of the plant defense reactions and, based on our current knowledge of arbuscular and ectomycorrhizal interactions, effectors to control plant immunity. Aim of this work was to identify, through the analysis of *O. maius* genomic and transcriptomic data, fungal SSPs potentially involved in the molecular dialog governing the ERM symbiosis.

MATERIALS AND METHODS

Fungal Strains and Growth Conditions

Oidiodendron maius strain Zn (hereafter *O. maius*) was isolated from the roots of *V. myrtillus* growing in the Niepolomice Forest (Poland), and first described by Martino et al. (2000). This *O. maius* strain is deposited at the Mycotheca Universitatis Taurinensis collection (MUT1381; University of Turin, Italy) and at the American Type Culture Collection (ATCC MYA-4765; Manassas, VA, US), and was maintained on Czapek-Dox solid medium (NaNO$_3$ 2 g L^{-1}, KCl 0.5 g L^{-1}, glycerol phosphate*H$_2$O 0.5 g L^{-1}, K$_2$HPO$_4$ 0.35 g L^{-1}, FeSO$_4$ 0.01 g L^{-1}, sucrose 30 g L^{-1}, agar 10 gL^{-1}, adjusted to pH 6). *O. maius* and *Om*Δ*SSP1*-null mutants were also grown in the presence of different stressor compounds. Czapek-Dox medium was

supplemented with 0.3 mM Cd (as $3CdSO_4^*8H_2O$), 15 mM Zn (as $ZnSO_4^*7H_2O$), 117.6 mM H_2O_2, 0.75 mM menadione, 0.1% (w:v) caffeic acid, 0.5% (w:v) tannic acid, 0.5% (w:v) gallic acid and 0.5% (w:v) quercetin. Prior to fungal inoculation, sterile cellophane membranes were placed on the agar surface to provide a convenient means of removing the mycelium from the plate. The membranes were first boiled for 15 min in 10 mM EDTA (disodium salt, dihydrate, SIGMA), rinsed and then autoclaved in ddH_2O. Fungal colonies were removed after 30 days, dried over-night and weighted.

In Vitro Mycorrhizal Synthesis

Axenic *V. myrtillus* seedlings were obtained from seeds (Les Semences du Puy, Le Puy-En-Velay, France) surface sterilized in 70% ethanol (v:v) 0.2% Tween 20 for 3 min, rinsed with sterile water, submerged in 0.25% sodium hypochlorite for 15 min and rinsed again with sterile water. Seeds were germinated on 1% water agar for 2 weeks in darkness before transfer to a growth chamber for 1 month.

Mycorrhiza was synthesized in petri plates containing Modified Melin-Norkrans (MMN) medium with the following composition: KH_2PO_4 0.5 gL^{-1}, Bovine Serum Albumin (BSA) 0.1 gL^{-1}, $CaCl_2^*2H_2O$ 0.066 gL^{-1}, NaCl 0.025 gL^{-1}, $MgSO_4^*7H_2O$ 0.15 gL^{-1}, thiamine-HCl 0.1 gL^{-1}, $FeCl_3^*6H_2O$ 0.001 gL^{-1}, agar 10 gL^{-1}, final pH 4.7. Sterile cellophane membranes, prepared as described before, were placed on the agar surface before fungal inoculation. A suspension of *O. maius* conidia in sterile deionised water was distributed on the cellophane membranes in the bottom half of the MMN petri plates. Ten germinated *V. myrtillus* seedlings were then transferred just above the conidia suspension. Plates were sealed and placed in a growth chamber (16-h photoperiod, light at 170 μmol m^{-2} s^{-1}, temperatures at 23°C day and 21°C night). Roots were collected and the percentage of mycorrhization evaluated after 45 day.

As a control for the asymbiotic condition, *O. maius* was grown on the same medium used for mycorrhizal synthesis. Plates covered by cellophane membranes were inoculated with 5 mm fungal plugs and fungal colonies were removed after 45 days. Three biological replicates were prepared for each sample for the RNA-Seq experiment.

RNA Extraction and RNA-Seq Data Analyses

Total RNA was extracted and quantified from 100 mg aliquots of *O. maius* mycelium and *O. maius*-inoculated *V. myrtillus* collected 45 days after inoculation, frozen in liquid nitrogen and mechanically ground. Total RNA was extracted from *O. maius* mycelium using a Tris-HCl extraction buffer and from *V. myrtillus* mycorrhizal roots using the CTAB method, as described by Kohler et al. (2015).

Preparation of libraries from total RNA and 2 × 100 bp Illumina HiSeq sequencing (RNA-Seq) was performed by IGA Technology Services (Udine, Italy). Raw reads were trimmed and aligned to the respective reference transcripts available at the JGI MycoCosm database (http://genome.jgi-psf.org/programs/fungi/index.jsf) using CLC Genomics Workbench v6. For mapping, the minimum length fraction was 0.9, the minimum similarity

fraction 0.8 and the maximum number of hits for a read was set to 10. The unique and total mapped reads number for each transcript were determined, and then normalized to RPKM (Reads Per Kilobase of exon model per Million mapped reads). A summary of the aligned reads is given in Table S1. The complete data set has been deposited in NCBI Gene Expression Omnibus and is accessible through GEO Series accession numbers GSE63947. To identify differentially regulated transcripts in mycorrhizal tissues compared to free-living mycelium, the Baggerly's Test (Baggerly et al., 2003) implemented in CLC Genomic workbench was used. This test compares the proportions of counts in a group of samples against those of another group of samples. The samples are given different weights depending on their sizes (total counts). The weights are obtained by assuming a Beta distribution on the proportions in a group, and estimating these, along with the proportion of a binomial distribution, by the method of moments. The result is a weighted t-type test statistic. In addition, Benjamini & Hochberg multiple-hypothesis testing corrections with False Discovery Rate (FDR) were used. Transcripts with a more then 5-fold change and a FDR corrected $p < 0.05$ were kept for further analysis.

O. maius Small Secreted Proteins (SSPs) were identified using a custom pipeline including SignalP v4 (1), WolfPSort (2), TMHMM, TargetP (3), and PS-Scan algorithms (4) as reported in Pellegrin et al. (2015). To assess whether symbiosis-regulated transcripts were conserved or lineage-specific (i.e., orphan genes with no similarity to known sequences in DNA databases), their protein sequences were queried against the protein repertoires of 59 fungal genomes using BLASTP with e-value 1e-5. Proteins were considered as orthologs of symbiosis-regulated transcripts pending they showed 70% coverage over the regulated sequence and at least 30% amino acid identity.

cDNA Synthesis and Quantitative RT-PCR (RT-qPCR)

The expression of seven selected SSPs was evaluated by RT-qPCR. cDNA was obtained from about 1,000 ng of total RNA with a reaction mix containing 10 μM random primers, 0.5 mM dNTPs, 4 μl 5× buffer, 2 μl 0.1 M DTT, and 1 μl Superscript II Reverse Transcriptase (Invitrogen) in a final volume of 20 μl. Temperature regime was: 65°C for 5 min, 25°C for 10 min, 42°C for 50 min, and 70°C for 15 min. Possible DNA contamination was tested with an additional PCR reaction using a specific primer pair that amplifies an intron containing region of the *O. maius Elongation Factor1α* (*OmEF1α*) (Table S2). RT-qPCR was performed with the Rotor-Gene Q (Qiagen) apparatus. The reactions were carried out in a final volume of 15 μl with 7.5 μl of iQ SYBR Green Supermix (Bio-rad), 5.5 μl of forward and reverse primers (10 μM stock concentration; Table S2) and 2 μl of cDNA (diluted 1:10). qPCR cycling program consisted of a 10 min/95°C holding step followed by 40 cycles of two steps (15 s/95°C and 1 min/60°C). The relative expression of the target transcript was measured using the $2^{-\Delta Ct}$ method (Livak and Schmittgen, 2001). The *Omβ-Tubulin* (*OmβTub*) (Table S2) was used as reference house-keeping gene. Three to five biological replicates and two technical replicates were analyzed for each condition tested. qPCR primers were

designed with Primer3Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) and checked for specificity and secondary structure formation with PrimerBlast (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) and OligoAnalyzer (eu.idtdna.com/calc/analyzer). Primers were synthesized by Eurogentec (Belgium).

Construction of the *OmSSP1*-Disruption Vector and *Agrobacterium*-Mediated Transformation

OmSSP1-null mutants were obtained through *Agrobacterium tumefaciens*-mediated (ATM) homologous recombination. PCR reactions were used to produce the 5′ upstream flanking region (1502 bp) and the 3′ downstream flanking region (1,533 bp) of the *OmSSP1* gene. PCR reactions were carried out in a final volume of 50 μl containing: 50 ng of genomic DNA of *O. maius Zn,* 1 μl dNTPs 10 mM, 2.5 μl of each primer (10 μM stock concentration; Table S2), 10 μl of 5× Phusion HF Buffer and 0.5 units of Phusion Hot Start II High-Fidelity (Thermo Scientific). The PCR program was as follows: 30 s at 98°C for 1 cycle, 10 s at 98°C, 30 s at 60°C, 45 s at 72°C for 30 cycles, 10 min at 72°C for 1 cycle. Amplicons were then purified with Wizard® SV gel and PCR clean-up system (PROMEGA) following the manufacturer's instructions. PCRs amplicons were cut with XmaI-HindIII (for the 5′) and BglII-HpaI (for the 3′) and cloned into the pCAMBIA0380_HYG vector (Fiorilli et al., 2016) in order to obtain the pCAMBIA0380_HYG-ΔOmSSP1 vector (Figure S1). The restriction reactions were performed in 30 μl final volume containing 0.5 μg of DNA (1 μg for the plasmid), 0.5 μl of each enzyme (from PROMEGA), 0.3 μl of BSA 100X and 3 μl of buffer 10X, overnight at 37°C. The ligase reaction was carried out in 20 μl final volume containing 50 ng of vector, 18 ng of the amplicon, 2 μl of buffer 10X and 1 μl of T4 enzyme (PROMEGA), overnight at 4°C. The vector sequence was checked by PCR and DNA sequencing.

The vector was cloned into *Agrobacterium tumefaciens* LBA1100, that was used to transform ungerminated *O. maius* conidia according to the protocol described in Abbà et al. (2009).

Identification of OmSSP1-Null Mutants by PCR and Southern Blot

Fungal transformants were screened by PCR. A small portion of each fungal colony was collected and boiled for 15 min in 20 μl of 10 mM Tris HCl pH 8.2, vortexed for 1 min and centrifuged 15 min at room temperature. Then, 2 μl of the supernatant were used directly for PCR amplification without any further purification, using two sets of primers (Table S2). The first primer set was designed to amplify the *OmSSP1* gene (OmSSPb1r and OmSSPb1f) whereas the second set (Hyg4f e Hyg2r) was designed to amplify the portion of the *hph* gene corresponding to the Hyg-probe (Figure S1). A OmSSP1-null mutant would yield an amplified product only with the second primer set. The putative OmSSP1-null mutants identified were validated by PCR using primers (Table S2) designed on the genome at the 5′ and 3′ of the homologous recombination site (respectively preOmSSPb1f3/Hyg6r and

TABLE 1 | Expression level, biochemical and bioinformatic features of seven selected *O. maius* mycorrhiza-induced SSPs.

Name		Expression				Biochemical features					No. of orthologs (see also Table S4c)	
Selected SSPs	Protein ID	FLM means	MYC means	Fold change	*p*-value	Predicted domains, repeats PFAM motifs	C residues (%)	GRAVY index	Protein length	Molecular weight (kDa)	*O. maius*	Total (60 genomes)
OmSSP1	182936	0.685	13734	20056	5.07 E-20	1 LCR, NO PFAM	8 (8.60)	+ 0.53	93	9.2	1	4
OmSSP2	160794	0.103	409	3961	8.84 E-44	1 LCR, PFAM: DUF4360	4 (1.58)	− 0.29	253	27.6	3	82
OmSSP3	56093	0.044	28	628	2.96 E-10	NO PFAM	1 (0.54)	+ 0.40	188	19.7	1	1
OmSSP4	20160	0	105	MS	7.56 E-10	2 internal repeats, NO PFAM	0 (0.00)	− 0.48	102	11.3	1	1
OmSSP5	23797	0	48	MS	1.55 E-06	2 internal repeats, NO PFAM	0 (0.00)	− 0.24	108	11.8	1	1
OmSSP6	125454	0	34	MS	1.50 E-22	NO PFAM	2 (0.91)	− 0.06	220	23.0	2	14
OmSSP7	25811	0	16	MS	0.003	1 LCR, NO PFAM	6 (2.37)	− 0.19	253	26.5	3	172

LCR, Low complexity region; MS, Transcripts could not be identified in the free living mycelium (FLM) and were considered as mycorrhiza specific.

postOmSSPb1r3/Hyg3f). The positive OmSSP1-null mutants were further analyzed through Southern blot hybridization to verify single-copy integration of the disruption cassette in the genome. Fifteen micrograms of genomic DNA from the deletion mutants and from the wild-type were digested with BamHI (PROMEGA) and size-fractionated on 1% (w:v) agarose TAE 1X gel. The separated restriction fragments were blotted onto a nylon membrane following standard procedures (see Abbà et al., 2009). Hybridization with a probe designed on the Hygromycin-resistance cassette (Hyg-probe, Figure S1) was performed with a chemiluminescent detection system (ECL direct DNA labeling and detection system; GE Healthcare, U.K.) according to the manufacturer's recommendations.

Quantification of the Degree of Mycorrhization

To determine differences in root colonization between _O. maius_ wild-type and OmSSP1-null mutants, the percentage of mycorrhization was recorded after 1.5 months. The roots of 3–6 seedlings colonized by each mutant strain were collected and the whole root system was stained overnight in a solution of lactic acid:glycerol:H_2O (14:1:1) containing acid fuchsin 0.01% (w:v), destained twice with 80% lactic acid and observed using a Nikon Eclipse E400 optical microscope. The magnified intersections method (Villarreal-Ruiz et al., 2004) was adapted to quantify the percentage of fungal colonization of _V. myrtillus_ hair roots, under the microscope, using the rectangle around the cross-hair as intersection area at 40× magnification. A total of 60 intersections per seedling root system were scored. Counts were recorded as percentage of root cells colonized (RC) by the fungus using the formula: RC% = 100 × Σ of coils counted for all the intersections/Σ of epidermal cells counted for all the intersections.

Phylogenetic and Bioinformatic Analyses

The aminoacid sequences comprised between cysteine 1 and cysteine 8 of the fungal proteins listed in Table S3 were aligned using MUSCLE (MUltiple Sequence Comparison by Log-Expectation; Gap Open penalty−2, Edgar, 2004) tool implemented in MEGA7 (Tamura et al., 2007). Maximum likelihood analysis (Guindon and Gascuel, 2003) was conducted using www.phylogeny.fr in advanced mode (Dereeper et al., 2008). The phylogenetic tree was reconstructed using the maximum likelihood method implemented in the PhyML program (v3.1/3.0 aLRT).

Bioinformatic analyses of protein primary sequences were performed using online tools. Blastp searches on the Uniprot database (The UniProt Consortium, 2017) identified the closest protein matches. Hydropathy profiles were generated with ProtScale tools (http://web.expasy.org/protscale/), while the grand average of hydropathicity (GRAVY), the predicted amino acid number and molecular weight were calculated using the ProtParam tool (http://web.expasy.org/protparam/). The intrinsic solubility profiles were obtained with the camsolintrinsic calculator (http://www-mvsoftware.ch.cam.ac.uk/index.php/camsolintrinsic). The representation of the residues hydrophobicity of the aligned sequences of each clade was obtained using http://www.ibi.vu.nl (Simossis and Heringa, 2005).

The Yeast Signal Sequence Trap Assay

Functional validation of the predicted signal peptide of OmSSP1 was conducted with a yeast signal sequence trap assay (Plett et al., 2011). The pSUC2-GW gateway vector carries a truncated invertase (SUC2) lacking both its initiation methionine and signal peptide. cDNA encoding the predicted OmSSP1 signal peptide was cloned into pSUC2-GW plasmids using BP/LR Gateway technologies (Invitrogen). Then, yeast cells
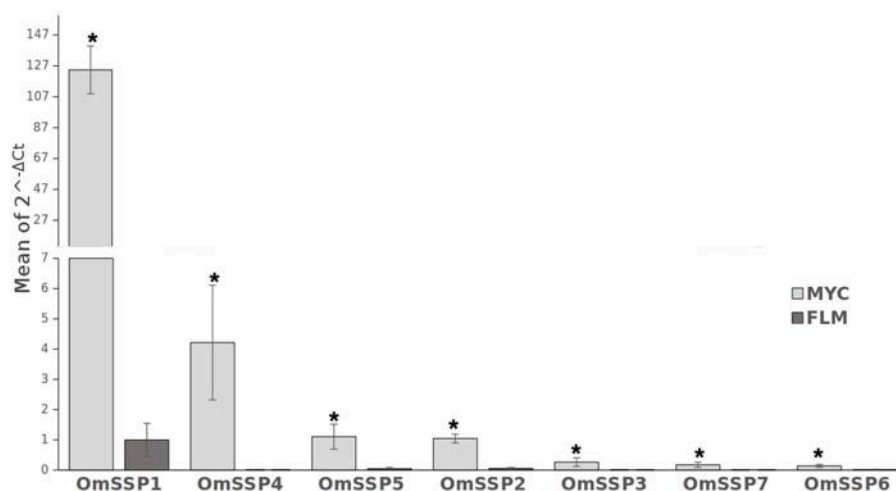
FIGURE 1 | RT-qPCR validation of selected _O. maius_ SSPs expression level in mycorrhizal tissues of _V. myrtillus_ (MYC) as compared to the free living mycelium (FLM). Relative expression (mean of ΔCt values) of 7 _O. maius_ SSPs was normalized to _OmTub_ transcripts. Bars represent the mean ±SD. Asterisks (*) indicates a statistically significant difference ($p < 0.05$) (ANOVA, Tukey's _post hoc_ test).
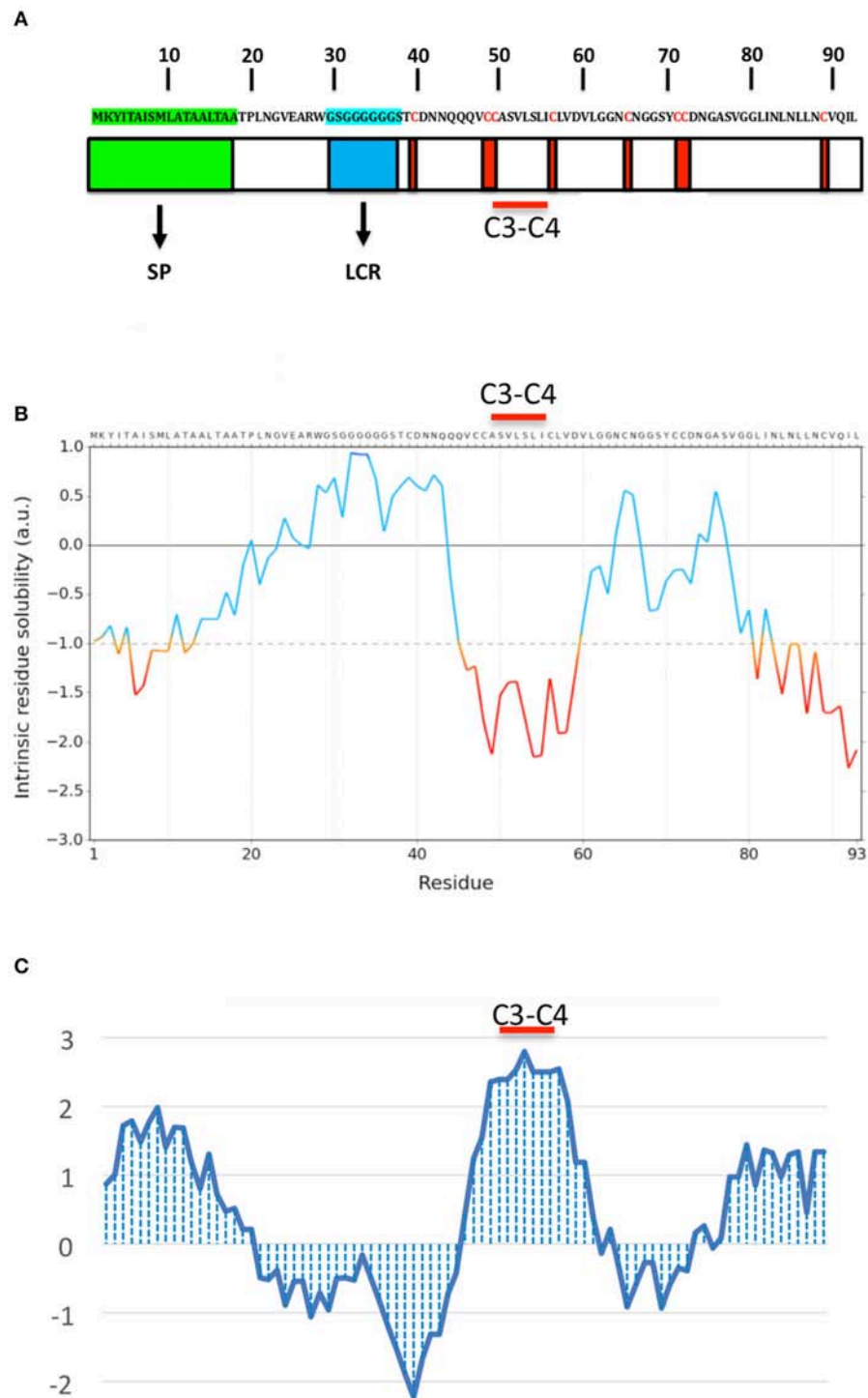
FIGURE 2 | OmSSP1 structural and biochemical features predicted through bioinformatics tools. (A) Schematic representation of OmSSP1 primary amino acid sequence. The green bar represents the signal peptide (SP), the blue bar the low complexity region (LCR) and the red bars the cysteine (C) residues. (B) Calculation of the intrinsic solubility profile (ISP) and (C) hydrophaty profile (HP). For the ISP, scores larger than 1 indicate highly soluble regions, while scores smaller than −1 indicate poorly soluble regions. For the HP, hydrophobic regions show positive peaks with values above 0 whereas hydrophilic regions show negative peaks. The position of the C3–C4 loop, characterized by a conserved hydrophobic core, is indicated. The sequences of A–C have been graphically aligned in order to show the correspondence between the elements of the primary sequence (A), their solubility (B) and hydrophobicity profile (C).

(YTK12 strain) were transformed with 200 ng of the pSUC2-GW/OmSSP1 plasmid using the lithium acetate method (Gietz and Schiestl, 2007). Transformants were grown on SD-W yeast minimal medium (6.7 gL^{-1} Yeast Nitrogen Base without amino acids, 0.7 gL^{-1} tryptophan dropout supplement, 20 gL^{-1} glucose, 20 gL^{-1} agar, pH 5.6) and on YPGA medium (10 gL^{-1} yeast extract, 20 gL^{-1} peptone, 20 gL^{-1} agar amended with 20 gL^{-1} glucose and 60 µg mL^{-1} antimycin A after autoclaving, pH 6.5). To assay for invertase secretion, colonies were grown overnight at 30°C with shaking (200 rpm) and diluted to an OD$_{600}$ = 1, then 5 µl of serial dilution of the yeast culture were plated onto YPSA medium containing sucrose (10 gL^{-1} yeast extract, 20 gL^{-1} peptone, 20 gL^{-1} agar amended with 2 gL^{-1} sucrose and 60 µg mL^{-1} antimycin A after autoclaving, pH 6.5).

Statistical Analyses

The significance of differences among the different treatments was statistically evaluated by ANOVA with Tukey's pairwise comparison as post-hoc test for multiple comparisons for normally distributed data. Statistical elaborations of growth and biomass data were performed using PAST statistical package, version 2.17 (Hammer et al., 2001). The differences were considered significant at a probability level of $p < 0.05$.

RESULTS

The *O. maius* Genome Contains Several SSPs Up-Regulated During Mycorrhizal Symbiosis With *V. myrtillus*

Among the 16,703 genes found in the *O. maius* genome (Kohler et al., 2015), 445 genes (∼38% of the total *O. maius* predicted secretome) code for putatively secreted proteins smaller than 300 amino acids (Table S4a). The transcriptomic analysis of *O. maius* under free-living conditions (FLM) and in symbiosis with *V. myrtillus* (MYC) indicated that about 24% (278/1163) of the genes coding for putatively secreted proteins were induced in symbiosis (Fold Change > 5, *p < 0.05*), 90 of them corresponding to SSPs. Many of these mycorrhiza-induced SSPs (MiSSP) were strongly up-regulated in symbiosis (13 with FC ≥ 400) or mycorrhiza-specific (Table S4b). Only 32 were cysteine enriched (C > 3%), a feature normally attributed to SSPs (Kim et al., 2016). About half (49/90) contained PFAM motifs specific of CAZymes (especially glycoside hydrolases, GHs), lipases, hydrophobins and peptidases, whereas the remaining 41 contained no known PFAM

domain. None of these MiSSPs contained a nuclear localization signal motif but 6 of them featured a KR rich sequence, i.e., a motif characterized by basic aminoacids (like lysine and arginine) supporting entry in the plant nucleus (Table S4b).

Genes orthologous to the *O. maius* MiSSPs were identified by genomic comparative analyses with 59 taxonomically and ecologically distinct fungi (Table S4d), including three other ERM fungi in the Leotiomycetes (*Meliniomyces bicolor, M. variabilis* and *Rhizoscyphus ericae*) belonging to the "*R. ericae*" aggregate (Vrålstad et al., 2002). Ten of the 90 *O. maius* symbiosis-induced SSPs were specific for *O. maius*, whereas 2 to 834 orthologous genes were found for the other 80 SSPs (Table S4b). Many *O. maius* SSPs orthologs were found in the other three ERM fungal species, although no ERM specific SSPs could be identified. The highest number of *O. maius* SSPs orthologs was found in the genomes of pathogenic and saprotrophic fungi (64 and 73 respectively), whereas only 38 orthologous genes were found in the genomes of 12 ECM fungi (Table S4b).

We selected seven *O. maius* SSPs for further analyses, with a preference for those uniquely or highly expressed in symbiosis that did not contain PFAM motifs with known functions (**Table 1**). Three different software (PrediSi, SIgnalP 4.1, Phobius) confirmed the presence of a signal peptide (**Table 1**), but alignment of the primary sequences indicated very low similarities. Cysteine enrichment > 3% was only observed for OmSSP1 (8.6%), the most highly symbiosis-induced SSP (**Table 1**). The expression of these OmSSPs was investigated in free-living and in mycorrhizal *O. maius* by RT-qPCR, which confirmed a significant up-regulation of all selected SSP genes in symbiosis, and in particular a very strong induction of OmSSP1 (**Figure 1**).

OmSSP1 Shows Molecular Similarities With Fungal Proteins Annotated as Hydrophobins

OmSSP1 is a single copy gene located in the scaffold 14 of the *O. maius* genome. The coding region contains 354 nucleotides, with 2 exons and 1 intron. In addition to the signal peptide, OmSSP1 is rich in glycine (16.1%) and leucine (12.9%) and contains 8 cysteine residues (**Figure 2**). The intrinsic calculated solubility and solvent accessibility highlighted at least three poorly water soluble regions. The majority of hydrophobic residues in OmSSP1 are clustered between amino acid residues

TABLE 2 | Results of BlastP searches on the Uniprot database using the OmSSP1 sequence as query.

Accession	Organism	Submitted name	Identity (%)	Score	E-value
UNIPROT DATABASE					
A0A014P1Y3	*Metarhizium robertsii*	Hydrophobin-like protein	57	60.5	1e-09
G9N067	*Hypocrea virens (=Trichoderma virens)*	Hydrophobin	55	57.8	2e-08
A9NIV6	*Fusarium culmorum*	Hydrophobin 3	55	54.3	2e-07
A0A1J7IK23	*Coniochaeta ligniaria*	Hydrophobin 3	54	51.6	2e-06
A0A0M9EQ60	*Fusarium langsethiae*	Hydrophobin 3	54	51.6	2e-06
A0A179FBJ9	*Pochonia chlamydosporia*	Fungal hydrophobin domain-containing protein	55	50.1	9e-06

41–56 and 79–93 (**Figure 2**). OmSSP1 secondary structure predictions indicated a disordered folded state for 46% of the protein, partly due to the presence of a low complexity region (LCR) before the first cysteine (**Figure 2**), whereas 16 and 20% of the protein structure could form helix and beta-sheet structures, respectively (not shown).

Although no known domains could be found in the predicted OmSSP1 protein sequence by PFAM database searching, nor a

FIGURE 3 | Phylogenetic tree of OmSSP1 and *O. maius* hydrophobins with other annotated hydrophobins from Ascomycetes. The analysis included protein sequences annotated or described as Ascomycetes class I and class II hydrophobins (listed in Table S3). This sequence alignment considered the complete amino acid sequence comprised between C1 and C8. Muscle algorithm implemented in MEGA7 (Tamura et al., 2007) was used to generate the multiple protein sequence alignment. The phylogenetic tree was reconstructed on the Phylogeny.fr platform (Dereeper et al., 2008) using the maximum likelihood method (Guindon and Gascuel, 2003) implemented in the PhyML program (v3.1/3.0 aLRT). The WAG substitution model was selected assuming an estimated proportion of invariant sites (of 0.088) and 4 gamma-distributed rate categories to account for rate heterogeneity across sites. The gamma shape parameter was estimated directly from the data (gamma = 7.097). Reliability for internal branch was assessed using the aLRT test (SH-Like). Graphical representation and edition of the phylogenetic tree were performed with TreeDyn (v198.3; Chevenet et al., 2006) A–D, terminal clades..

functional classification in the InterPro database, BlastP searches both in the UniProt and in the RefSeq databases yielded, as closest protein matches, fungal hydrophobins or hydrophobin-like proteins (**Table 2**). Indeed, the two *OmSSP1* orthologous genes in *T. terrestris* (protein ID 2089872 and 209296) showed high homology with fungal hydrophobins. The hydropathy profile of OmSSP1 was analyzed and a positive GRAVY value was found (0.53), indicating overall hydrophobicity (**Table 1**). Therefore, we compared the protein sequence of OmSSP1 with those of the four *O. maius* annotated hydrophobins and of other fungal hydrophobins (Linder et al., 2005; Seidl-Seiboth et al., 2011; Grigoriev et al., 2014). Three out of the four *O. maius* hydrophobins showed a GRAVY score above 0.6, indicating an overall hydrophobicity higher than OmSSP1 (0.53), whereas the GRAVY score for *O. maius* hydrophobin 4 was only 0.32 (Table S5).

Two classes of fungal hydrophobins have been distinguished by Wessels (1994) on the basis of the hydropathy profile and of slightly different motifs between the eight positionally conserved cysteine residues: Class I ($C-X_{5-8}-CC-X_{17-39}-C-X_{8-23}-C-X_{5-6}-CC-X_{6-18}-C-X_{2-13}$), and Class II ($C-X_{9-10}-CC-X_{11}-C-X_{16}-C-X_{8-9}-CC-X_{10}-C-X_{6-7}$). OmSSP1 bears a signature motif similar to class I hydrophobins ($C-X_7-CC-X_7-C-X_8-C-X_5-CC-X_{16}-C-X_2$), with the exception of a shorter sequence (only 7 aminoacids) between cysteines 3 and 4, expected to host the most hydrophobic protein region. Although shorter, the C_3-C_4 loop of OmSSP1 showed a hydrophobic core composed by valine (V) and leucine (L), thus explaining the negative values in solubility profile and positive values in hydropathy profile (**Figure 2**). Hydrophobins are amphiphilic molecules (Whiteford and Spanu, 2002; Rineau et al., 2017), and OmSSP1 has a very hydrophilic stretch rich in G before the C_3-C_4 loop, corresponding to the LCR (**Figure 2**).

To better understand the phylogenetic relatedness of OmSSP1 with fungal hydrophobins, we aligned the OmSSP1 protein sequence with the four annotated *O. maius* hydrophobins and with Class I and Class II annotated hydrophobins from Ascomycetes (listed in Table S3). A phylogenetic tree built on the complete C_1-C_8 sequence alignment is shown in **Figure 3**. However, since the different proteins showed a highly variable sequence length of the C_3-C_4 loop, ranging from 4 to 39 amino acids, a more conserved phylogenetic tree was generated by Maximum Likelihood (ML) without this protein region, to avoid possible bias due to the different protein lengths (Figure S2). In both trees, class II hydrophobins and two of the four *O. maius* hydrophobins (Oidma2 and Oidma3) grouped in a single, well supported cluster. The two other *O. maius* hydrophobins grouped in a well-supported cluster together with characterized Class I hydrophobins (**Figure 3** and Figure S2). Although the position of some fungal proteins differed in the two ML trees, most terminal clades (A–D) were maintained and well supported. OmSSP1 clustered in Clade C (**Figure 3** and Figure S2) together with other proteins reported as hydrophobins (Table S3) and featuring a short C_3-C_4 loop (X_{4-9}). The complete C_1-C_8 sequence alignment of proteins in Clade C is shown in **Figure 4**. As in OmSSP1, most amino acids in the short C_3-C_4 loop of other proteins in this clade were hydrophobic.
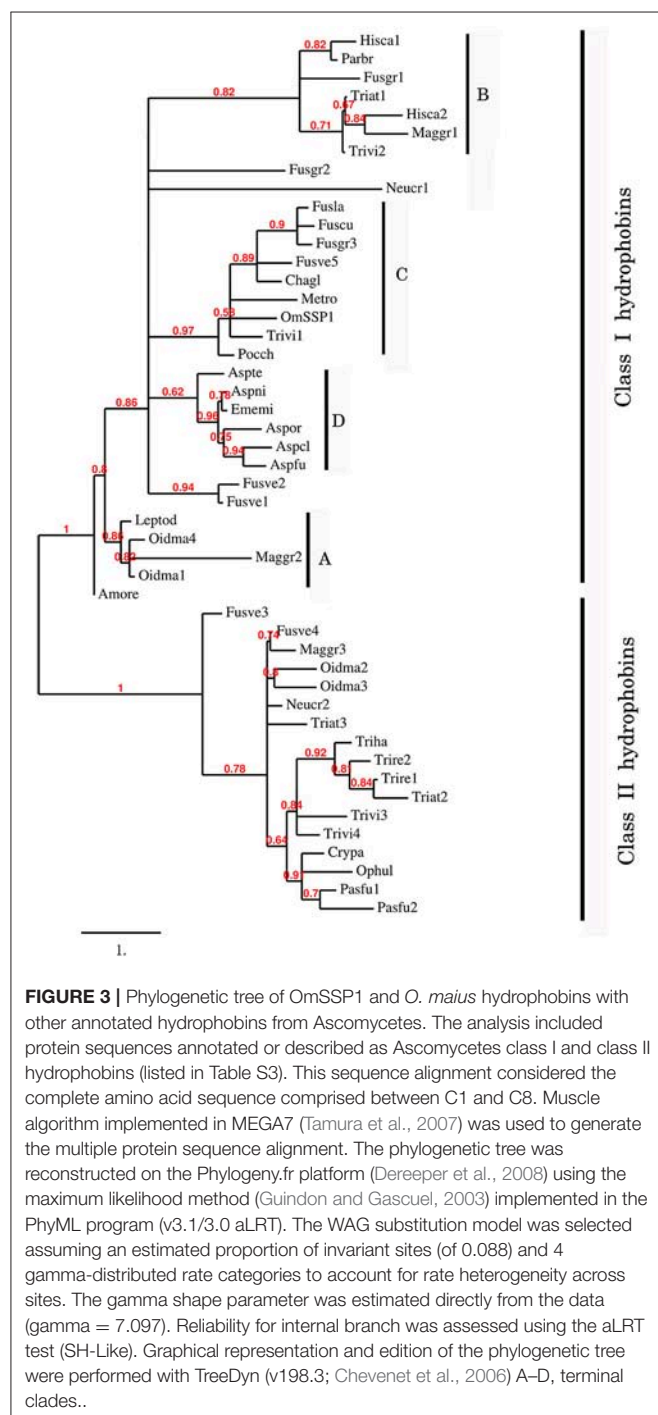
Clade B (**Figure 3** and Figure S2) was another well-supported clade containing proteins also featuring a very short (X_{8-9}) C_3-C_4 loop (**Figure 4**). Clade B included a *Trichoderma atroviride* hydrophobin (Triat1) described by Seidl-Seiboth et al. (2011) as a member of a novel subclass in Class I hydrophobins. However, unlike Clade C, proteins in Clade B featured mainly hydrophilic amino acids in the C_3-C_4 loop (**Figure 4**). Clade A included the two *O. maius* hydrophobins Oidma1 and Oidma4.

The Yeast Invertase Secretion Assay Indicates That OmSSP1 Is Secreted

The predicted OmSSP1 signal peptide was functionally validated in a yeast signal sequence trap assay. This test is based on the yeast requirement for a secreted invertase (SUC) to grow on sucrose amended media (Klein et al., 1996). The pSUC-GW (Jacobs et al., 1997) gateway vector carries a truncated invertase that lacks its signal peptide (SUC2). The cDNA coding for the putative OmSSP1 signal peptide (OmSSP1_SP) was fused in frame to the yeast SUC2 invertase, and the recombinant pSUC-GW vector was transformed into the invertase-deficient yeast strain YTK12. As positive control, the MiSSP7 (Plett et al., 2011) signal peptide (MiSSP7_SP) and the yeast wild-type signal peptide (SUC2_SP$^+$) were used, whereas the empty vector was used as negative (mock) control. All transformants grew on SD-W and YPGA control media containing glucose, whereas only OmSSP1_SP and the two positive controls rescued YTK12 yeast's growth on YPSA media containing sucrose, indicating that OmSSP1 contains a secretion signal that is functional in yeast (**Figure 5**).

Growth of *Om△SSP1* Mutants Is Not Impaired Under Stressful Conditions, but They Have a Reduced Capability to Colonize *V. myrtillus* Roots

O. maius can be genetically transformed and gene disruption can be obtained by homologous recombination (Martino et al., 2007; Abbà et al., 2009). To investigate the biological function of OmSSP1 in *O. maius*, knock-out mutants (*Om△SSP1*) were obtained through AMT transformation using a vector containing the *hpd*-cassette (Figure S1). Hygromycin-resistant colonies were screened by PCR, and eight putative homologous recombinants out of 742 screened fungal transformants could be identified. Southern blot hybridization with a probe to the Hygromycin cassette showed a single band of the expected size for four out of the eight candidate mutants (Figure S3). To further confirm the vector insertion site, PCR amplifications were performed with primers designed to amplify the genome regions flanking the inserted pCAMBIA0380△OmSSP1 disruption cassette (Figure S1, Table S2), followed by sequencing of the amplicons. Three *O. maius* transformants (*Om△SSP1^{150}*, *Om△SSP1^{377}*, *Om△SSP1^{412}*) were confirmed as *OmSSP1* deletion mutants. These three OmSSP1-null mutants were not affected in mycelium morphology or growth rate when inoculated on Czapek-Dox solid medium (not shown). *OmSSP1* deletion did not modify the wettability phenotype of the *O. maius* mycelium (Figure S4), but it should be noted that *OmSSP1* gene expression was very low in the FLM (Table S4b).

It has been recently suggested that, in FLM, SSPs may increase fungal tolerance to toxic compounds (such as aromatic
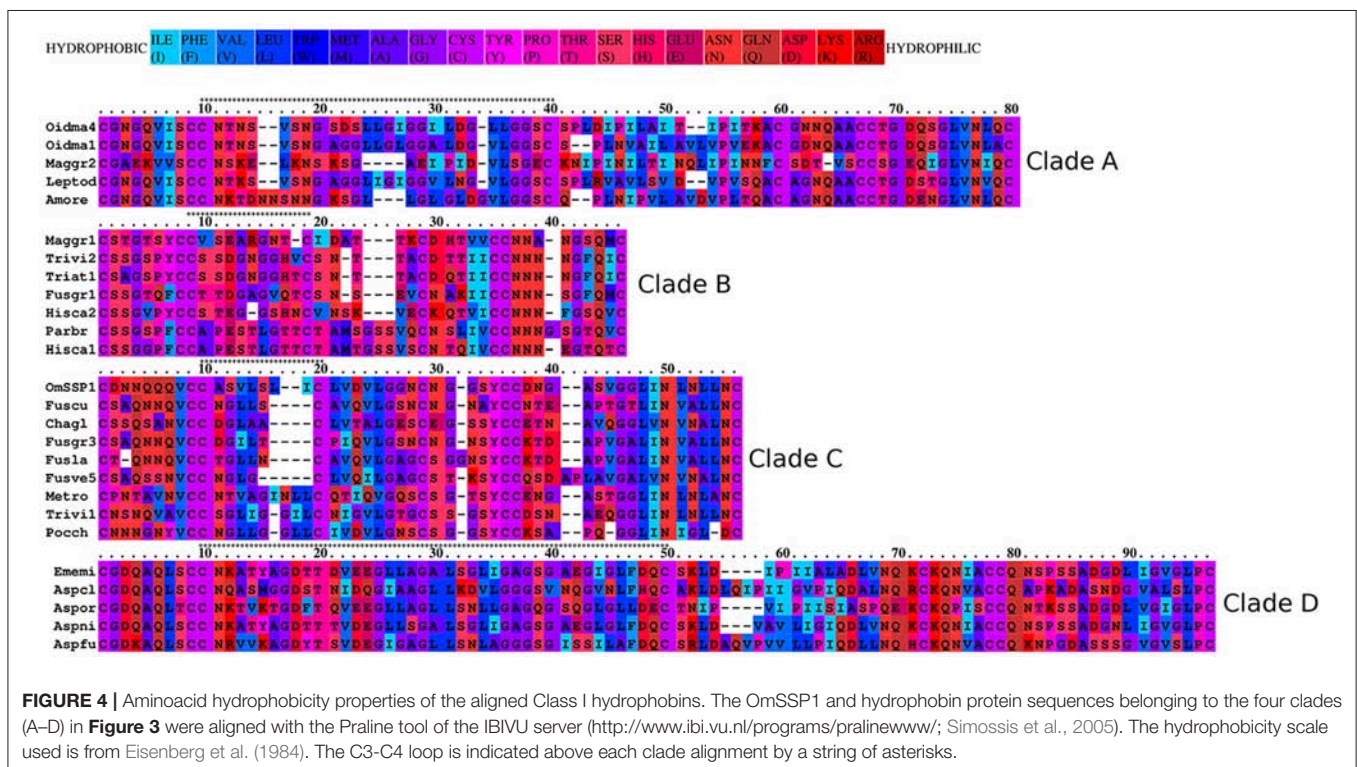
FIGURE 4 | Aminoacid hydrophobicity properties of the aligned Class I hydrophobins. The OmSSP1 and hydrophobin protein sequences belonging to the four clades (A–D) in **Figure 3** were aligned with the Praline tool of the IBIVU server (http://www.ibi.vu.nl/programs/pralinewww/; Simossis et al., 2005). The hydrophobicity scale used is from Eisenberg et al. (1984). The C3-C4 loop is indicated above each clade alignment by a string of asterisks.

compounds or reactive oxygen species) released during substrate degradation (Valette et al., 2017). We therefore investigated whether deletion of the *OmSSP1* gene reduced *O. maius* fitness when exposed to different stress inducers. As shown in Figure S5, growth of the three *OmΔSSP1* mutants was not significantly affected by any of the stress conditions tested, namely two toxic heavy metals (Cd and Zn), molecules causing oxidative stress (H_2O_2 and menadione) and plant-derived organic compounds displaying toxic/antimicrobical effects (caffeine, tannic acid, gallic acid, quercetin and caffeic acid).

We then investigated the symbiotic ability of the three *OmΔSSP1* mutants on seedlings of the host plant *V. myrtillus* as compared with the wild-type *O. maius* strain (**Figure 6**). Although there were no statistically significant differences in plant biomass after 45 days of co-culture, a statistically significant ($p < 0.05$) reduction in the percentage of root colonization was measured for all *OmΔSSP1* mutants, when compared with the wild-type *O. maius* strain (**Figure 6**).

DISCUSSION

Similar to Other Mycorrhizal Fungi, the ERM Fungus *O. maius* Has a Wide Array of SSPs

Effector-like SSPs have been found to be secreted by ECM and AM fungi and to be instrumental for plant colonization (Plett and Martin, 2015; Martin et al., 2016). SSPs are also encoded in the genome of the model ERM fungus *O. maius*. Overall, the *O. maius* genome contains 445 SSPs, corresponding to 2.6% of the total number of *O. maius* genes. Similar percentages were reported for the ECM fungus *L. bicolor* (Pellegrin et al., 2015) and for saprotrophic fungi (Valette et al., 2017). The 90 mycorrhiza-induced SSPs (MiSSPs) correspond to about 20% of the total *O. maius* SSPs, a percentage similar to the ECM fungus *L. bicolor* (Kohler et al., 2015) and the AM fungus *R. irregularis* (Tisserant et al., 2013).

Bioinformatic analysis of *O. maius* MiSSPs revealed that 45.5% (41/90) correspond to orphans genes with no known PFAM domains. Many fungal effector SSPs are targeted to the host plant nucleus (Lo Presti et al., 2015). Although none of the *O. maius* MiSSPs contained a nuclear localization signal motif, six *O. maius* MiSSPs featured a KR reach sequence supporting a localization in the plant nucleus.

Comparative genomics revealed that 27% of the total *O. maius* SSPs are species-specific (Table S4a). Species-specific SSPs (SSSPs), defined as SSPs with no homology in other species, have been found in both AM (Salvioli et al., 2016; Sedzielewska Toro and Brachmann, 2016; Tang et al., 2016) and ECM fungi (Pellegrin et al., 2015) and they are considered to be likely involved in the promotion of host-specific interactions (Pellegrin et al., 2015). The number of SSSPs found in *O. maius* falls in the proportion predicted by Kim et al. (2016) for symbiotic organisms (25-50%), although only 10 out of the 90 *O. maius* SSPs up-regulated in symbiosis were species-specific (Table S4b).

Whereas lifestyle-specific SSPs have been found in ECM fungi (Pellegrin et al., 2015), no SSPs shared by all four ERM fungi (with no orthologs in the other fungi used for comparative analysis) could be found. ERM and ECM fungi also differed because comparative analyses showed that ECM fungi share most of their SSPs with saprotrophs such as brown rot, white rot, and litter decayers (Pellegrin et al., 2015), whereas the highest number of *O. maius* symbiosis-induced SSPs orthologous belong to pathogenic and saprotrophic fungi.

SSPs have been recently found in most fungal species regardless of their lifestyle, suggesting that they could be involved in a variety of processes, both common and lifestyle-specific (Pellegrin et al., 2015). Interestingly, three orthologous genes of *OmSSP1*, the most highly up-regulated (ca. 20,000-folds) *O. maius* MiSSP, were found not in other ERM fungi but in the two saprotrophic fungi *Neurospora crassa* and *Thelavia terrestris*. Switching among endophytic, pathogenic, and saprotrophic lifestyles, likely controlled by both environmental and host factors, were suggested for *N. crassa*

FIGURE 5 | OmSSP1 contains a signal sequence that is functional in yeast. **(A)** Upon sucrose selection (YPSA medium) the OmSSP1 signal peptide (OmSSP1_SP) rescues the functionality of the yeast invertase, like the other two positive controls (The MiSSP7 signal peptide - MiSSP7_SP - and the wild-type sequence of the yeast signal peptide - SUC2_SP $^+$). The empty vector (mock) was used as a negative control. Control media **(B)** SD-W and **(C)** YPGA, both containing glucose, restore the growth ability of all yeast transformants. Five microliter of undiluted (ud) or serially diluted (10^{-1} and 10^{-2}) yeast culture were plated.

(Kuo et al., 2014). *T. terrestris*, a thermophilic saprotrophic fungus phylogenetically related to *N. crassa*, shows important cellulases and hemicellulases activities (Berka et al., 2015). Thus, *N. crassa* and *T. terrestris* share important characteristics with *O. maius*, which besides being an endomycorrhizal fungus, is also commonly isolated from roots of other plants as well as from substrates rich in plant-derived organic matter (Rice and Currah, 2006). Moreover, the *O. maius* gene content for cell wall degrading enzymes, proteases and lipases places this fungus closer to saprotrophs and pathogens than to other mycorrhizal fungi (Martino et al., 2018), and may explain the fact that OmSSP1 orthologs are also found in SAP S/L/O (soil/litter/organic matter saprotrophs) fungi.

OmSSP1, the Most Highly Expressed *O. maius* MiSSP, May be a Distinctive Type of Hydrophobin

Hydrophobins are small secreted proteins less than 200 amino acids long, with a secretion signal and a pattern of eight cysteine residues recurring in the sequence (Whiteford and

FIGURE 6 | Om△SSP1 mutants have a reduced capability to colonize *V. myrtillus* roots. *V. myrtillus* roots were observed after 1.5 months of co-culture with *O. maius* WT and with three OmSSP1 null-mutant strains. **(A)** The percentage of the root colonization was significantly lower for the OmSSP1 null-mutants as compared to the *O. maius* WT. **(B)** Quantification of fresh plant biomass (roots - gray bars - and aboveground portions - white bars) of *V. myrtillus* plants grown alone, in the presence of the *O. maius* WT strain or of the OmSSP1 null-mutants. Bars represent the mean ± SD, $n = 5$. Each biological replicate represents the total biomass of eight *V. myrtillus* seedlings grown in an individual plate. Different letters indicate statistically significant difference ($p < 0.05$, ANOVA, Tukey's *post hoc* test).

Spanu, 2002). A unique three-dimensional folding comes from these features, keeping exposed the hydrophobic residues and rendering them amphiphilic (Rineau et al., 2017). Two classes of hydrophobins have been recognized (Wessels, 1994): Class I, with higher sequence variability and more stable superstructures, is found in both Asco- and Basidiomycetes, while Class II has only been found in Ascomycetes (Kershaw and Talbot, 1998). Hydrophobins are abundantly expressed during fungal development, pathogenesis and symbiosis (Wösten, 2001; Whiteford and Spanu, 2002). Being amphiphilic, they could behave as biosurfactants and facilitate fungal adhesion to organic matter and its decomposition (Rineau et al., 2017). Hydrophobins are also instrumental for fungal hyphae to form aerial structures and to adhere to each other and/or to hydrophobic surfaces, such as the plant leaf surface during pathogenesis. Symbiosis-upregulated hydrophobins have been found in the ECM fungi *Pisolithus tinctorius* (Tagu et al., 2001) and in *L. bicolor* (Martin et al., 2008; Plett et al., 2012), where they could play a role in establishing hyphal aggregation in the symbiotic interfaces (Raudaskoski and Kothe, 2015).

The *O. maius* genome features four annotated hydrophobins containing the PFAM and InterPro hydrophobin domains. Rineau et al. (2017) suggested that all *O. maius* hydrophobins belong to Class I, but our phylogenetic analysis (that also included Class II hydrophobins) showed that two proteins (Oidma1 and Oidma4) belong to Class I and two (Oidma2 and Oidma3) to Class II hydrophobins. OmSSP1 shares some features with Class I hydrophobins and clusters with annotated hydrophobins in this Class, but it was not identified as a hydrophobin because it lacks the corresponding PFAM and InterPro domains, possibly because of the shorter C3–C4 region.

Amino acid features, such as charge and hydrophobicity, can influence hydrophobin structure and function. Thus, the amino acidic composition of the C_3-C_4 loop as well as of the N-terminal region of hydrophobins may influence the wettability and the substrate-attachment preference of the protein (Linder et al., 2005; Kwan et al., 2006). In this respect, it is interesting to note that proteins in Clade B and Clade C, both showing a C_3-C_4 loop unusually short for Class I hydrophobins, feature amino acid sequences with very different hydrophobicity (**Figure 4**), suggesting they may represent structurally and functionally diverse subclasses of Class I hydrophobins. The low complexity region found in OmSSP1 is also unusual for hydrophobins and could be considered a recently evolved trait of this protein (Toll-Riera et al., 2012). Its presence suggests for OmSSP1 a low propension to aggregate and to form alpha-helices and beta-sheets, three properties often correlated with the ability of hydrophobins to pile up in needle-like (amyloid) structures (Rineau et al., 2017).

A high number of hydrophilic residues (asparagine especially) were found in the N-amino terminal region of OmSSP1. According to Linder et al. (2005), the amino terminal region of hydrophobins could have important roles in the specific function of individual proteins. For example, hydrophobins featuring high number of exposed hydrophilic residues at the N-terminal region were found to be overexpressed in mycorrhizal tissues (Whiteford and Spanu, 2002; Rineau et al., 2017).

OmSSP1 Null-Mutants Have a Reduced Ability to Colonize *V. myrtillus* Roots

There is increasing awareness that SSPs may play important roles during saprotrophic fungal growth, as they have been identified in saprotrophic fungi and they can be expressed by mycorrhizal fungi during asymbiotic growth (Vincent et al., 2012; Doré et al., 2015; Valette et al., 2017). However, although a limited range of growth conditions were tested, OmSSP1 did not appear to be necessary in the FLM, as the three OmSSP1-null mutants were not affected in mycelium morphology or growth rate, even when they were exposed to toxic and oxidative chemical compounds. By contrast, when they were tested for symbiotic capabilities on *V. myrtillus* plants, a significant reduction in the percentage of mycorrhization (from about 37% to about 23-24%) was measured as compared with the wild-type strain, thus suggesting a specific role of *OmSSP1* in the mycorrhization process. The OmSSP1 deletion did not fully prevent root mycorrhization. However, 20% of *O. maius* SSPs are induced in symbiosis, and although OmSSP1 was the most highly up-regulated, we cannot exclude a functional redundancy, as already reported for the effectors of pathogenic fungi (Selin et al., 2016). Thus, the absence of OmSSP1 could be partly compensated by other *O. maius* SSPs with similar function, thus lowering the impact of the *OmSSP1* deletion.

CONCLUSIONS

In conclusion, the genome of the ERM fungus *O. maius* contains several SSPs that are up-regulated in symbiosis. Decreased colonization of *V. myrtillus* roots by OmSSP1-null mutants indicates that this protein, the most highly induced in the ERM symbiosis, is a hydrophobin-like effector that participates in the molecular fungal-plant interaction occurring during mycorrhizal formation. Our data demonstrate for the first time the importance of MiSSPs in ERM, although several questions remain open on the cellular localisation of OmSSP1 and its role in symbiosis.

In ECM, hydrophobins likely play an important role in hyphal aggregation during the formation of the extraradical fungal mantle and the Hartig net (Tagu et al., 2001). However, a similar role is unlikely for OmSSP1 because ERM fungi do not form any extraradical hyphal aggregate and individual ERM fungal hyphae take contact with the hydrophilic surface of root epidermal cells (the only cell type colonized in ERM) prior to cell wall penetration (Perotto et al., 2012).

Other intriguing roles have been proposed for fungal hydrophobins in plant-microbe interactions. For example, Whiteford and Spanu (2002) suggested that, by forming a layer on the hyphal surface, hydrophobins could shield potential microbe-associated molecular patterns (MAMPs) recognized by the host and thus avoid plant defense reactions. In *Trichoderma*, a genus comprising several soil borne plant-growth promoting fungi, adherence to the root surface and root colonization have been suggested to be mediated by hydrophobins (Hermosa et al., 2012). In fact, deletion of the Class I hydrophobin TasHyd1 decreased root attachment and intercellular growth in *T. asperellum* (Viterbo and Chet, 2006). Similarly we could speculate that a role of OmSSP1 in ERM may be to enhance *O. maius* attachment to the root surface and to protect the fungal hypha from plant defense compounds.

AUTHOR CONTRIBUTIONS

SC, SD, ElM, and SP planned and designed the research; SC, SD, ElM, SP, AK, EmM, H-RK, YD, and CV-F performed experiments or sequencing, collected, analyzed or interpreted the data; ElM, SP, and FM contributed reagents, materials, analysis; SC, ElM, and SP wrote the manuscript; SD, CV-F, and FM revised it critically. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2018.00546/full#supplementary-material

REFERENCES

Abbà, S., Khouja, H. R., Martino, E., Archer, D. B., and Perotto, S. (2009). SOD1-targeted gene disruption in the ericoid mycorrhizal fungus *Oidiodendron maius* reduces conidiation and the capacity for mycorrhization. *Mol. Plant Microbe Int.* 22, 1412–1421. doi: 10.1094/MPMI-22-11-1412

Akum, F. N., Steinbrenner, J., Biedenkopf, D., Imani, J., and Kogel, K.-H. (2015). The *Piriformospora indica* effector PIIN_08944 promotes the mutualistic Sebacinalean symbiosis. *Front. Plant Sci.* 6:906. doi: 10.3389/fpls.2015.00906

Baggerly, K. A., Deng, L., Morris, J. S., and Aldaz, C. M. (2003). Differential expression in SAGE: accounting for normal between-library variation. *Bioinformatics* 19, 1477–1483. doi: 10.1093/bioinformatics/btg173

Berka, R. M., Grigoriev, I. V., Otillar, R., Salamov, A., Grimwood, J., Reid, I., et al. (2015). Comparative genomic analysis of the thermophilic biomass-degrading fungi *Myceliophthora thermophila* and *Thielavia terrestris*. *Nat. Biotechnol.* 29, 922–927. doi: 10.1038/nbt.1976

Chevenet, F., Brun, C., Banuls, A. L., Jacq, B., and Chisten, R. (2006). TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics* 7:439. doi: 10.1186/1471-2105-7-439

Daghino, S., Martino, E., and Perotto, S. (2016). Model systems to unravel the molecular mechanisms of heavy metal tolerance in the ericoid mycorrhizal symbiosis. *Mycorrhiza* 26, 263–274. doi: 10.1007/s00572-015-0675-y

Dalpé, Y. (1986). Axenic synthesis of ericoid mycorrhiza in *Vaccinium angustifolium* ait. by oidiodendron species. *New Phytol.* 103, 391–396. doi: 10.1111/j.1469-8137.1986.tb00624.x

de Jonge, R., Bolton, M. D., and Thomma, B. P. (2011). How filamentous pathogens co-opt plants: the ins and outs of fungal effectors. *Curr. Op. Plant Biol.* 14, 400–406. doi: 10.1016/j.pbi.2011.03.005

Dereeper, A., Guignon, V., Blanc, G., Audic S., Buffet, S., Chevenet, F., et al. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucl. Acids Res.* 1, W465–W469. doi: 10.1093/nar/gkn180

Doré, P., Perraud, M., Dieryckx, C., Kohler, A., Morin, E., Henrissat, B., et al. (2015). Comparative genomics, proteomics and transcriptomics give new insight into the exoproteome of the basidiomycete *Hebeloma cylindrosporum* and its involvement in ectomycorrhizal symbiosis. *New Phytol.* 208, 1169–1187. doi: 10.1111/nph.13546

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Eisenberg, D., Schwarz, E., Komaromy, M., and Wall, R. (1984). Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J. Mol. Biol.* 179, 125–142. doi: 10.1016/0022-2836(84)90309-7

Fiorilli, V., Belmondo, S., Khouja, H.-R., Abb,à, S., Faccio, A., Daghino, S., et al. (2016). RiPEIP1, a gene from the arbuscular mycorrhizal fungus *Rhizophagus irregularis*, is preferentially expressed in planta and may be involved in root colonization. *Mycorrhiza* 26, 609–621. doi: 10.1007/s00572-016-0697-0

Gietz, R. D., and Schiestl, R. H. (2007). High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* 2, 31–34. doi: 10.1038/nprot.2007.13

Giraldo, M. C., Dagdas, Y. F., Gupta, Y. K., Mentlak, T. A., Yi, M., Martinez-Rocha, A. L., et al. (2013). Two distinct secretion systems facilitate tissue invasion by the rice blast fungus *Magnaporthe oryzae*. *Nature Comm.* 4:1996. doi: 10.1038/ncomms2996

Grigoriev, I. V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., et al. (2014). MycoCosm portal: gearing up for 1000 fungal genomes. *Nucl. Acids Res.* 42, D699–D704. doi: 10.1093/nar/gkt1183

Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704. doi: 10.1080/10635150390235520

Hacquard, S., Joly, D. L., Lin, Y.-C., Tisserant, E., Feau, N., Delaruelle, C., et al. (2012). A comprehensive analysis of genes encoding small secreted proteins identifies candidate effectors in *Melampsora larici-populina* (poplar leaf rust). *Mol. Plant-Microbe Int.* 25, 279–293. doi: 10.1094/MPMI-09-11-0238

Hammer, Ø., Harper, D. A. T., and Ryan, P. D. (2001). PAST - Paleontological statistics software. package for education and data analysis. *Palaeont. Electr.* 4:9. Available online at: http://palaeo-electronica.org/2001_1/past/issue1_01.htm

Hermosa, R., Viterbo, A., Chet, I., and Monte, E. (2012). Plant-beneficial effects of *Trichoderma* and of its genes. *Microbiology* 158, 17–25. doi: 10.1099/mic.0.052274-0

Jacobs, K. A., Collins-Racie, L. A., Colbert, M., Duckett, M., Golden-Fleet, M., Kelleher, K., et al. (1997). A genetic selection for isolating cDNAs encoding secreted proteins. *Gene* 198, 289–296. doi: 10.1016/S0378-1119(97)00330-2

Jones, J. D., and Dangl, J. L. (2006). The plant immune system. *Nature* 444, 323–329. doi: 10.1038/nature05286

Kamel, L., Tang, N., Malbreil, M., San Clemente, H., Le Marquer, M., Roux, C., et al. (2017). The comparison of expressed candidate secreted proteins from two arbuscular mycorrhizal fungi unravels common and specific molecular tools to invade different host plants. *Front. Plant Sci.* 8:124. doi: 10.3389/fpls.2017.00124

Kershaw, M. J., and Talbot, N. J. (1998). Hydrophobins and repellents: proteins with fundamental roles in fungal morphogenesis. *Fungal Gen. Biol.* 23, 18–33. doi: 10.1006/fgbi.1997.1022

Kim, K.-T., Jeon, J., Choi, J., Cheong, K., Song, H., Choi, G., et al. (2016). Kingdom-wide analysis of fungal small secreted proteins (SSPs) reveals their potential role in host association. *Front. Plant Sci.* 7:186. doi: 10.3389/fpls.2016.00186

Klein, R. D., Gu, Q., Goddard, A., and Rosenthal, A. (1996). Selection for genes encoding secreted proteins and receptors. *Proc. Natl. Acad. Sci. U.S.A.* 93, 7108–7113. doi: 10.1073/pnas.93.14.7108

Kloppholz, S., Kuhn, H., and Requena, N. (2011). A secreted fungal effector of *Glomus intraradices* promotes symbiotic biotrophy. *Curr. Biol.* 21, 1204–1209. doi: 10.1016/j.cub.2011.06.044

Kohler, A., Kuo, A., Nagy, L. G., Morin, E., Barry, K. W., Buscot, F., et al. (2015). Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nat. Gen.* 47, 410–415. doi: 10.1038/ng.3223

Kuo, H.-C., Hui, S., Choi, J., Asiegbu, F. O., Valkonen, J. P. T., and Lee, Y.-H. (2014). Secret lifestyles of *Neurospora crassa*. *Sci. Rep.* 4:5135. doi: 10.1038/srep05135

Kwan, A. H., Winefield, R. D., Sunde, M., Matthews, J. M., Haverkamp, R. G., Templeton, M. D., et al. (2006). Structural basis for rodlet assembly in fungal hydrophobins. *Proc. Nat. Acad. Sci. U.S.A.* 103, 3621–3626. doi: 10.1073/pnas.0505704103

Linder, M. B., Szilvay, G. R., Nakari-Setälä, T., and Penttilä, M. E. (2005). Hydrophobins: the protein-amphiphiles of filamentous fungi. *FEMS Microbiol. Rev.* 29, 877–896. doi: 10.1016/j.femsre.2005.01.004

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., et al. (2015). Fungal effectors and plant susceptibility. *Ann. Rev. Plant Biol.* 66, 513–545. doi: 10.1146/annurev-arplant-043014-114623

Martin, F. M., Uroz, S., and Barker, D. G. (2017). Ancestral alliances: plant mutualistic symbioses with fungi and bacteria. *Science* 356:eaad4501. doi: 10.1126/science.aad4501

Martin, F., Aerts, A., Ahrén, D., Brun, A., Danchin, E. G., Duchaussoy, F., et al. (2008). The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature* 452, 88–92. doi: 10.1038/nature06556

Martin, F., Kohler, A., Murat, C., Veneault-Fourrey, C., and Hibbett, D. S. (2016). Unearthing the roots of ectomycorrhizal symbioses. *Nat. Rev. Microbiol.* 14, 760–773. doi: 10.1038/nrmicro.2016.149

Martino, E., Morin, E., Grelet, G. A., Kuo, A., Kohler, A., Daghino, S., et al. (2018). Comparative genomics and transcriptomics depict ericoid mycorrhizal fungi as versatile saprotrophs and plant mutualists. *New Phytol.* 217, 1213–1229. doi: 10.1111/nph.14974

Martino, E., Murat, C., Vallino, M., Bena, A., Perotto, S., and Spanu, P. (2007). Imaging mycorrhizal fungal transformants that express EGFP during ericoid endosymbiosis. *Curr. Genet.* 52, 65–75. doi: 10.1007/s00294-007-0139-9

Martino, E., Turnau, K., Girlanda, M., Bonfante, P., and Perotto, S. (2000). Ericoid mycorrhizal fungi from heavy metal polluted soils: their identification and growth in the presence of zinc ions. *Mycol. Res.* 104, 338–344. doi: 10.1017/S0953756299001252

Pellegrin, C., Morin, E., Martin, F. M., and Veneault-Fourrey, C. (2015). Comparative analysis of secretomes from ectomycorrhizal fungi with an emphasis on small-secreted proteins. *Front. Microbiol.* 6:1278. doi: 10.3389/fmicb.2015.01278

Perotto, S., Martino, E., Abb,à, S., and Vallino, M. (2012). "14 Genetic diversity and functional aspects of ericoid mycorrhizal fungi," in *Fungal Associations*, ed B. Hock (Berlin, Heidelberg: Springer), 255–285.

Plett, J. M., and Martin, F. (2015). Reconsidering mutualistic plant–fungal interactions through the lens of effector biology. *Current Op. Plant Biol.* 26, 45–50. doi: 10.1016/j.pbi.2015.06.001

Plett, J. M., Daguerre, Y., Wittulsky, S., Vayssières, A., Deveau, A., Melton, S. J., et al. (2014). Effector MiSSP7 of the mutualistic fungus *Laccaria bicolor* stabilizes the *Populus* JAZ6 protein and represses jasmonic acid (JA) responsive genes. *Proc. Nat. Acad. Sci. U.S.A.* 111, 8299–8304. doi: 10.1073/pnas.1322671111

Plett, J. M., Gibon, J., Kohler, A., Duffy, K., Hoegger, P. J., Velagapudi, R., et al. (2012). Phylogenetic, genomic organization and expression analysis of hydrophobin genes in the ectomycorrhizal basidiomycete *Laccaria bicolor*. *Fungal Gen. Biol.* 49, 199–209. doi: 10.1016/j.fgb.2012.01.002

Plett, J. M., Kemppainen, M., Kale, S. D., Kohler, A., Legué, V., Brun, A., et al. (2011). A secreted effector protein of *Laccaria bicolor* is required for symbiosis development. *Curr. Biol.* 21, 1197–1203. doi: 10.1016/j.cub.2011. 05.033

Punt, P. J., Oliver, R. P., Dingemanse, M. A., Pouwels, P. H., and van den Hondel, C. A. (1987). Transformation of *Aspergillus* based on the hygromycin B resistance marker from *Escherichia coli*. *Gene* 56,117–124. doi: 10.1016/0378-1119(87)90164-8

Raudaskoski, M., and Kothe, E. (2015). Novel findings on the role of signal exchange in arbuscular and ectomycorrhizal symbioses. *Mycorrhiza* 25, 243–252. doi: 10.1007/s00572-014-0607-2

Rice, A. V., and Currah, R. S. (2006). "*Oidiodendron maius*: saprobe in sphagnum peat, mutualist in ericaceous roots?" in *Microbial Roots Endophytes*, eds B. J. E. Shulz, C. J. C. Boyle, and T. N. Sieber (Berlin: Springer-Verlag), 227–246.

Rineau, F., Lmalem, H., Ahren, D., Shah, F., Johansson, T., Coninx, L., et al. (2017). Comparative genomics and expression levels of

hydrophobins from eight mycorrhizal genomes. *Mycorrhiza* 27, 383–396. doi: 10.1007/s00572-016-0758-4

Rovenich, H., Boshoven, J. C., and Thomma, B. P. (2014). Filamentous pathogen effector functions: of pathogens, hosts and microbiomes. *Curr. Opin. Plant Biol.* 20, 96–103. doi: 10.1016/j.pbi.2014.05.001

Ruytinx, J., Martino, E., Rozpadek, P., Daghino, S., Turnau, K., Colpaert, J., et al. (2016). "Homeostasis of trace elements in mycorrhizal fungi," in *Molecular Mycorrhizal Symbiosis*, ed F. Martin (Hoboken, NJ: John Wiley & Sons, Inc.), 277–298.

Salvioli, A., Ghignone, S., Novero, M., Navazio, L., Bagnaresi, P., and Bonfante, P. (2016). Symbiosis with an endobacterium increases the fitness of a mycorrhizal fungus, raising its bioenergetic potential. *ISME J.* 10, 130–144. doi: 10.1038/ismej.2015.91

Scherlach, K., Graupner, K., and Hertweck, C. (2013). Molecular bacteria-fungi interactions: effects on environment, food, and medicine. *Ann. Rev. Microbiol.* 67, 375–397. doi: 10.1146/annurev-micro-092412-155702

Sedzielewska Toro, K., and Brachmann, A. (2016). The effector candidate repertoire of the arbuscular mycorrhizal fungus *Rhizophagus clarus*. *BMC Genomics* 17:101. doi: 10.1186/s12864-016-2422-y

Seidl-Seiboth, V., Gruber, S., Sezerman, U., Schwecke, T., Albayrak, A., Neuhof, T., et al. (2011). Novel hydrophobins from *Trichoderma* define a new hydrophobin subclass: protein properties, evolution, regulation and processing. *J. Mol. Evol.* 72, 339–351. doi: 10.1007/s00239-011-9438-3

Selin, C., de Kievit, T. R., Belmonte, M. F., and Fernando, W. G. (2016). Elucidating the role of effectors in plant-fungal interactions: progress and challenges. *Front. Microbiol.* 7:600. doi: 10.3389/fmicb.2016.00600

Simossis, V. A., and Heringa, J. (2005). PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. *Nucl. Acids Res.* 33, W289–W294. doi: 10.1093/nar/gki390

Simossis, V. A., Kleinjung, J., and Heringa, J. (2005). Homology-extended sequence alignment. *Nucl. Acids Res.* 33, 816–824. doi: 10.1093/nar/gki233

Stergiopoulos, I., and de Wit, P. J. (2009). Fungal effector proteins. *Ann. Rev. Phytopathol.* 47, 233–263. doi: 10.1146/annurev.phyto.112408.132637

Tagu, D., De Bellis, R., Balestrini, R., De Vries, O. M. H., Piccoli, G., Stocchi, V., et al. (2001). Immunolocalization of hydrophobin HYDPt-1 from the ectomycorrhizal basidiomycete *Pisolithus tinctorius* during colonization of *Eucalyptus globulus* roots. *New Phytol.* 149, 127–135. doi: 10.1046/j.1469-8137.2001.00009.x

Talbot, J. M., Bruns, T. D., Smith, D. P., Branco, S., Glassman, S. I., Erlandson, S., et al. (2013). Independent roles of ectomycorrhizal and saprotrophic communities in soil organic matter decomposition. *Soil Biol. Biochem.* 57, 282–291. doi: 10.1016/j.soilbio.2012.10.004

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24, 1596–1599. doi: 10.1093/molbev/msm092

Tang, N., San Clemente, H., Roy, S., Bécard, G., Zhao, B., and Roux, C. (2016). A survey of the gene repertoire of *Gigaspora rosea* unravels conserved features among Glomeromycota for obligate biotrophy. *Front. Microbiol.* 7:233. doi: 10.3389/fmicb.2016.00233

The UniProt Consortium (2017). UniProt: the universal protein knowledgebase. *Nucl. Acids Res.* 45, D158–D169. doi: 10.1093/nar/gkw1099

Tian, C., Beeson, W. T., Iavarone, A. T., Sun, J., Marletta, M. A., Cate, J. H., et al. (2009). Systems analysis of plant cell wall degradation by the model filamentous fungus *Neurospora crassa*. *Proc. Nat. Acad. Sci. U.S.A.* 106, 22157–22162. doi: 10.1073/pnas.0906810106

Tisserant, E., Malbreil, M., Kuo, A., Kohler, A., Symeonidi, A., Balestrini, R., et al. (2013). Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. *Proc. Nat. Acad. Sci. U.S.A.* 110, 20117–20122. doi: 10.1073/pnas.1313452110

Toll-Riera, M., Radó-Trilla, N., Martys, F., and Albà, M. M. (2012). Role of low-complexity sequences in the formation of novel protein coding sequences. *Mol. Biol. Evol.* 29, 883–886. doi: 10.1093/molbev/msr263

Tsuzuki, S., Handa, Y., Takeda, N., and Kawaguchi, M. (2016). Strigolactone-induced putative secreted protein 1 is required for the establishment of symbiosis by the arbuscular mycorrhizal fungus *Rhizophagus irregularis*. *Mol. Plant Microbe Interact.* 29, 277–286. doi: 10.1094/MPMI-10-15-0234-R

Valette, N., Benoit-Gelber, I., Falco, M. D., Wiebenga, A., de Vries, R. P., Gelhaye, E., et al. (2017). Secretion of small proteins is species-specific within *Aspergillus* sp. *Microb. Biotech.* 10, 323–329. doi: 10.1111/1751-7915.12361

van Esse, H. P., van't Klooster, J. W., Bolton, M. D., Yadeta, K. A., van Baarlen, P., Boeren, S., et al. (2008). The *Cladosporium fulvum* virulence protein Avr2 inhibits host proteases required for basal defense. *Plant Cell Online* 20, 1948–1963. doi: 10.1105/tpc.108.059394

Villarreal-Ruiz, L., Anderson, I. C., and Alexander, I. J. (2004). Interaction between an isolate from the *Hymenoscyphus ericae* aggregate and roots of *Pinus* and *Vaccinium*. *New Phytol.* 164, 183–192. doi: 10.1111/j.1469-8137.2004.01167.x

Vincent, D., Kohler, A., Claverol, S., Solier, E., Joets, J., Gibon, J., et al. (2012). Secretome of the free-living mycelium from the ectomycorrhizal Basidiomycete *Laccaria bicolor*. *J. Proteome Res.* 11, 157–171. doi: 10.1021/pr200895f

Viterbo, A., and Chet, I. (2006). TasHyd1, a new hydrophobin gene from the biocontrol agent *Trichoderma asperellum*, is involved in plant root colonization. *Mol. Plant Pathol.* 7, 249–258. doi: 10.1111/j.1364-3703.2006.00335.x

Vrålstad, T., Myhre, E., and Schumacher, T. (2002). Molecular diversity and phylogenetic affinities of symbiotic root-associated ascomycetes of the Helotiales in burnt and metal polluted habitats. *New Phytol.* 155, 131–148. doi: 10.1046/j.1469-8137.2002.00444.x

Wang, S., Boevink, P. C., Welsh, L., Zhang, R., Whisson, S. C., and Birch, P. R. J. (2017). Delivery of cytoplasmic and apoplastic effectors from *Phytophthora infestans* haustoria by distinct secretion pathways. *New Phytol.* 216, 205–215. doi: 10.1111/nph.14696

Wessels, J. G. H. (1994). Developmental regulation of fungal cell wall formation. *Ann. Rev. Phytopathol.* 32, 413–437. doi: 10.1146/annurev.py.32.090194.002213

Whiteford, J. R., and Spanu, P. D. (2002). Hydrophobins and the interactions between fungi and plants. *Mol. Plant Pathol.* 3, 391–400. doi: 10.1046/j.1364-3703.2002.00129.x

Wösten, H. A. B. (2001). Hydrophobins: multipurpose proteins. *Ann. Rev. Microbiol.* 55, 625–646. doi: 10.1146/annurev.micro.55.1.625

Zuccaro, A., Lahrmann, U., and Langen, G. (2014). Broad compatibility in fungal root symbioses. *Curr. Op. Plant Biol.* 20, 135–145. doi: 10.1016/j.pbi.2014.05.013

Domain Swap Approach Reveals the Critical Roles of Different Domains of SYMRK in Root Nodule Symbiosis in *Lotus japonicus*

Hao Li, Mengxiao Chen, Liujian Duan, Tingting Zhang, Yangrong Cao and Zhongming Zhang*

State Key Laboratory of Agricultural Microbiology, College of Life Sciences and Technology, Huazhong Agricultural University, Wuhan, China

Symbiosis receptor kinase (SYMRK) is a cell membrane-localized protein kinase containing extracellular malectin-like domain (MLD) and leucine-rich repeat (LRR) domains, which is critically required for both root nodule symbiosis (RNS) and arbuscular mycorrhizal symbiosis (AMS). SYMRK is widely distributed in the genomes of different plant species; however, the contribution of different domains of SYMRK and its homologs from other plant species to RNS is largely unclear. In this study, SYMRK and its homologs from three typical plant species including *Medicago truncatula* (for both RNS and AMS), *Oryza sativa* (for AMS but not RNS), and *Arabidopsis thaliana* (for neither RNS or AMS) were investigated using domain swap approach in response to rhizobia in *Lotus japonicus*. Full-length SYMRK from rice and *Medicago* but not from *Arabidopsis* could complement *Lotus symrk-409* mutant plants to contribute RNS. The chimeric protein with the extracellular domain (ED) of LjSYMRK and cytoplasmic domains (CD) of SYMRK from both *Medicago* and rice but not *Arabidopsis* could contribute to RNS in *Lotus*, suggesting that the CD of SYMRK is required for symbiotic signaling. The chimeric receptors containing the CD of LjSYMRK (SYMRK^{CD}) and the EDs of MtDMI2 (MtDMI2^{ED}), OsSYMRK (OsSYMRK^{ED}), AtSYMRK (AtSYMRK^{ED}), NFR1 (NFR1^{ED}), and NFR5 (NFR5^{ED}) could complement *Lotus symrk-409* mutant plants to develop nodules. However, MtDMI2 could partially complement *Lotus symrk-409* mutants to form both effective nodules and ineffective bumps, which is similar to the complementation results from MtDMI2^{ED}-LjSYMRK^{CD} and LjSYMRK^{GDLC} in *Lotus symrk-409* mutants, suggesting that ED of SYMRK has a very fine-tune regulation for RNS in *Lotus*. The deletion of either MLD or LRR on SYMRK^{GDLC} (a mutant version of SYMRK with GDPC motif replaced by GDLC) could contribute to RNS when overexpressed in *Lotus symrk-409* mutants, suggesting that MLD and LRR domains might work together to be involved in symbiotic signaling and the LRR domain might play a negative role in LjSYMRK^{GDLC}-mediated RNS. By mutagenizing the conserved amino acids on LRR domain, five serine residues were found to be required for the function of LjSYMRK^{GDLC} in RNS. These finding precisely refine the molecular mechanisms of SYMRK function in symbiotic signaling in *L. japonicus*.

Keywords: *Lotus japonicus*, root nodule symbiosis, symbiosis receptor kinase, symbiotic signaling, domain swap, nitrogenase activity

INTRODUCTION

In barren soils, leguminous plants form symbiosis with rhizobia leading to the development of a new organ called a nodule, where rhizobia reside and subsequently reduce nitrogen into ammonium for the plant host in exchange for nutrients. In comparison to the root nodule symbiosis (RNS), which is mostly restricted in legumes, the arbuscular mycorrhizal symbiosis (AMS) is widespread in most landed plants, with a few exceptions (for example cruciferous plants). It is estimated that AMS originated about 400 million years ago, while RNS originated about 60 million years ago (Remy et al., 1994; Young et al., 2011), consistent with the generally accepted theory that RNS evolved from AMS and both of them evolved from plant–pathogens interaction (Parniske, 2008; Markmann and Parniske, 2009).

For almost all the symbiosis between legumes and rhizobia, the initiation of compatible interaction begins with a molecular dialog between two partners (Oldroyd, 2013). The flavonoids secreted by plant hosts induce the biosynthesis and secretion of lipo-chitooligosaccharide called Nod factor (NF) by rhizobia (Fisher and Long, 1992). In the root hairs, two LysM receptor-like kinases, for example, NFR1/NFR5 in *Lotus*, or LYK3/NFP in *Medicago*, could recognize the NF to initiate the symbiotic signaling leading to the formation of root nodules for rhizobial colonization (Limpens et al., 2003; Madsen et al., 2003; Radutoiu et al., 2003; Arrighi et al., 2006; Broghammer et al., 2012; Moling et al., 2014). Recently, another LysM-RLK (exopolysaccharide receptor 3, EPR3) was shown to recognize rhizobia EPS, indicating that plants could recognize at least two symbiotic signals to mediate symbiosis (Kawaharada et al., 2015, 2017). A leucine-rich repeat (LRR) receptor-like kinase, Lotus SYMRK or Medicago Does Not make Infections (DMI2), plays a key role in symbiotic signaling acting as a downstream component of Nod Factor Receptors (Ané et al., 2002; Stracke et al., 2002; Miwa et al., 2006). All these receptor-like kinases are required for calcium spiking and expression of symbiosis-relate genes *via* a common symbiosis pathway (CSP) (Oldroyd and Downie, 2008; Kouchi et al., 2010). Similar to RNS, LjNFR1/MtLYK3 also participates in AMS with a possible function to recognize Myc Factor produced by arbuscular mycorrhizal fungi (Zhang et al., 2015). The Myc and Nod factors initiate a common signaling pathway (CSP) with overlapped signaling components identified (Oldroyd, 2013; Genre and Russo, 2016). Symbiosis receptor kinase (SYMRK) is widespread in most plant species such as leguminous and cereal plants. Furthermore, in the *Arabidopsis* who could not have the capability of either RNS or AMS has two SYMRK homologs, i.e., AT1G67720.1 and AT2G37050.1 (Shiu and Bleecker, 2001).

In *Lotus*, knock-out of SYMRK completely abolished the ability to form ITs or produce nodules with rhizobia. However, in the *symrk-14* mutant plants that harbor a point mutation at GDPC motif linking malectin and LRR domains, epidermal responses including infection thread formation were significantly reduced, while the formation of nodule primordia and cortical infection were slightly changed (Kosuta et al., 2011), suggesting that epidermal response and cortical program might be differentially regulated during symbiotic interaction with rhizobia. Since GDPC motif was shown to be required

for the cleavage of malectin-like domain (MLD) from SYMRK protein (Antolín-Llovera et al., 2014), malectin and LRR might have a fine-tune regulation mechanism in mediating symbiotic response. The truncated version SYMRK$^{\Delta MLD}$ (lacking the MLD) was shown to be able to interact with NFR5 and transfer the symbiotic signal to downstream targets (Antolín-Llovera et al., 2014). In *Lotus*, overexpression of the full-length SYMRK could produce spontaneous nodules in the absence of rhizobia (Ried et al., 2014). Similarly, overexpression of the kinase domain of *Arachis hypogaea* SYMRK (AhSYMRK^{KD}) could also induce spontaneous nodule formation in the absence of rhizobia (Saha et al., 2014), suggesting that the kinase domain of SYMRK might have a dominant positive role in nodule organogenesis. However, the autophosphorylation of SYMRK seems not to be essential to mediate symbiotic signaling, since mutation at an autophosphorylation site (Tyr-670) in AhSYMRK could partially complement *symrk* mutants by producing infection threads and uninfected nodule primordia (Samaddar et al., 2013; Paul et al., 2014; Saha et al., 2016). In addition, expression of *OsSYMRK* under the control of the *SYMRK* promoter could only complement *symrk* mutants to produce AMS but not RNS in *Lotus* (Markmann et al., 2008). All these results not only indicated the critical role of SYMRK protein in RNS, but also suggested that SYMRK uncouples the rhizobia infection and nodule formation. However, how the extracellular domains (EDs) of SYMRK is involved in RNS is largely unknown.

In this study, we demonstrate that MtDMI2 (SYMRK homolog protein in *Medicago*) and OsSYMRK but not AtSYMRK could complement *Lotus SYMRK$^{-/-}$* plants. In addition, multiple chimeric SYMRK proteins were developed to analyze the function of different domains of SYMRK in RNS. The data refine the molecular mechanisms of each domain of SYMRK in RNS in *Lotus*.

RESULTS

Homology Analysis of SYMRK in Legumes and Non-legumes

SYMRK is critically required for both root nodule symbiosis (RNS) and AMS. To study its function involved in RNS, SYMRK homologs were identified based on sequence similarity with *Lotus* SYMRK using BLASTp as a search engine at the National Center for Biotechnology Information database (NCBI), and 18 SYMRK homologs were identified. A phylogenetic tree based on various SYMRK proteins was made using the MEGA 5.1 software (**Figure 1**). In the phylogenetic tree, SYMRK proteins and homologs fall into three different clades. Clade I contains SYMRK proteins from plants that could have both RNS and AMS; clade II contains SYMKR from plants having AMS but not RNS, while proteins from Cruciferae plants that could not possess either AMS or RNS fallen into clade III.

Both OsSYMRK and MtDMI2 but not AtSYMRK Could Complement *symrk-409* Mutant in *Lotus*

To study the function of SYMRK involved in RNS in *Lotus*, an *LORE1*-insertion *symrk* mutant were identified from the

FIGURE 1 | Phylogenetic tree of SYMRK proteins in different species. The Phylogenetic tree was constructed using MEGA 5.1 software with 1,000 bootstrap replications. Abbreviation of each species names were used to represent the SYMRK protein cloned from this species. Ma, *Melilotus albus*; Mt, *M. truncatula*; Ps, *Pisum sativum*; Ca, *Cicer arietinum*; Gm, *Glycine max*; Sr, *Sesbania rostrata*; Lj, *L. japonicus*; Sl, *Solanum lycopersicum*; Dc, *Dendrobium catenatum*; Zm, *Zea mays*; Sb, *Sorghum bicolor*; Os, *Oryza sativa*; Bd, *Brachypodium distachyon*; Br, *Brassica rapa*; Rs, *Raphanus sativus*; Cs, *Camelina sativa*; At, *Arabidopsis thaliana*; Al, *Arabidopsis lyrata*.

Centre for Carbohydrate Recognition and Signaling (CARB, http://users-mb.au.dk/pmgrp/) (Fukai et al., 2012; Urbanski et al., 2012; Małolepszy et al., 2016). The line 30010361 (*symrk-409*) contains an *LORE1*-insertion between exon 4 and intron 4, leading to the early termination of translation of SYMRK (Figures S1A–C). In response to rhizobia treatment, infection threads and nodules were not detected in the *symrk-409* mutant plants when compared with *Lotus* B-129 Gifu 7- and 21-day post-inoculated (DPI) with rhizobia (Figures S1D–G), indicating that the *symrk-409* is another knock-out mutant that could be used for future study.

To analyze the symbiotic function of various SYMRK proteins, one SYMRK protein from each clade-based phylogenetic tree, respectively, was chosen for study, i.e., MtDMI2 from *Medicago* that could have both RNS and AMS, OsSYMRK from rice that could only have AMS, and AtSYMRK from *Arabidopsis* that could have neither RNS or AMS.

To test the function of each SYMRK as indicated above, SYMRK proteins including LjSYMRK, MtDMI2, OsSYMRK, and AtSYMRK and control vector were transgenically expressed in the *symrk-409* mutant plants under the control of the *Ljubiquitin* promoter using hairy root transformation. The *Ljubiquitin* promoter could induce a strong expression of multiple genes in previous publications (Maekawa et al., 2009; Yoro et al., 2014; Wang et al., 2016). Compared with no nodules observed in the control transformation (**Figures 2A1–5**), overexpression of LjSYMRK or OsSYMRK or MtDMI2 but not AtSYMRK-induced nodule formation 21 DPI with rhizobia in *symrk-409* mutant plants (**Figure 2A**). There was no significant difference in the number of nodules per plant root transformed with

either LjSYMRK or MtDMI2 or OsSYMRK 21 DPI with rhizobia (**Figure 2B**). In comparison to wild type and LjSYMRK-transgenic *Lotus* nodules, decreased nitrogenase activities from transgenic nodules with expression of either MtDMI2 or OsSYMRK were measured. While no nitrogenase activity was detected in roots transformed with either vector control or AtSYMRK (**Figure 2C**). In consistence with the low levels of nitrogenase activity from MtDMI2-transgenic nodules, about 30% ineffective bumps with no rhizobia infection were observed (**Figure 2D**). However, no ineffective bumps were observed in OsSYMRK-transgenic Lotus *symrk-409* mutant roots. These data indicate that both MtDMI2 and OsSYMRK but not AtSYMRK could complement *symrk-409* mutant to form nodules.

Since both *Lotus* and *Medicago* belong to leguminous plants, we examined whether LjSYMRK could complement *MtDMI2*-knock-out mutant plants *dmi2-1* (or named TR25). Both LjSYMRK and MtDMI2 and control vector were introduced into *dmi2-1* using hairy root transformation. In comparison with no infection threads and nodules observed on the *dmi2-1* mutant transformed with empty vector (Figures S2A1–5,B), a significant number of nodules and infected nodule primordium was detected on roots transformed with either MtDMI2 or LjSYMRK (Figures S2A6–15,2B), which is consistent with nitrogenase activity measured in these two transgenic nodules (Figure S2C). However, about 20% nodule cases formed on *dmi2-1* mutant plants expressing *LjSYMRK* are ineffective since no rhizobia were detected using lacZ staining (Figure S2D), suggesting that MtDMI2 and LjSYMRK have overlapping function to complement each other but with exceptional roles in mediating symbiosis in their own hosts.

The CD of SYMRK Is Essential but the ED Is Important in RNS in *Lotus*

To study the function of each domain of SYMRK from different plant species, a variety of constructs was made using domain swaps and functionally tested in *symrk-409* mutant plants using hairy root transformation (Figure S3). To study the function of the cytoplasmic domain (CD) of SYMRK, the CD of LjSYMRK replaced with CDs of MtDMI2, OsSYMRK, and AtSYMRK (Figure S3). In *symrk-409*, roots transformed with LjSYMRK^{ED}-MtDMI2^{CD} and LjSYMRK^{ED}-OsSYMRK^{CD} formed normal infected nodule primordia and effective nodules (**Figure 3A1–10,B–D**) with similar nitrogenase activities measured (**Figure 3C**). However, LjSYMRK^{ED}-AtSYMRK^{CD} could not complement *symrk-409* to form infection threads and nodules (**Figures 3A11–15**). These data suggest that the CD of SYMRK plays an essential role in RNS.

To analyze the symbiotic function of the ED of SYMRK, the ED of LjSYMRK was replaced with the ED of MtDMI2, OsSYMRK, and AtSYMRK to obtain the chimeric protein MtDMI2^{ED}-LjSYMRK^{CD}, OsSYMRK^{ED}-LjSYMRK^{CD}, and AtSYMRK^{ED}-LjSYMRK^{CD}, respectively (Figure S3). After transgenically expressed in *symrk-409* roots, the function of these three chimeric proteins in nodulation were analyzed. All
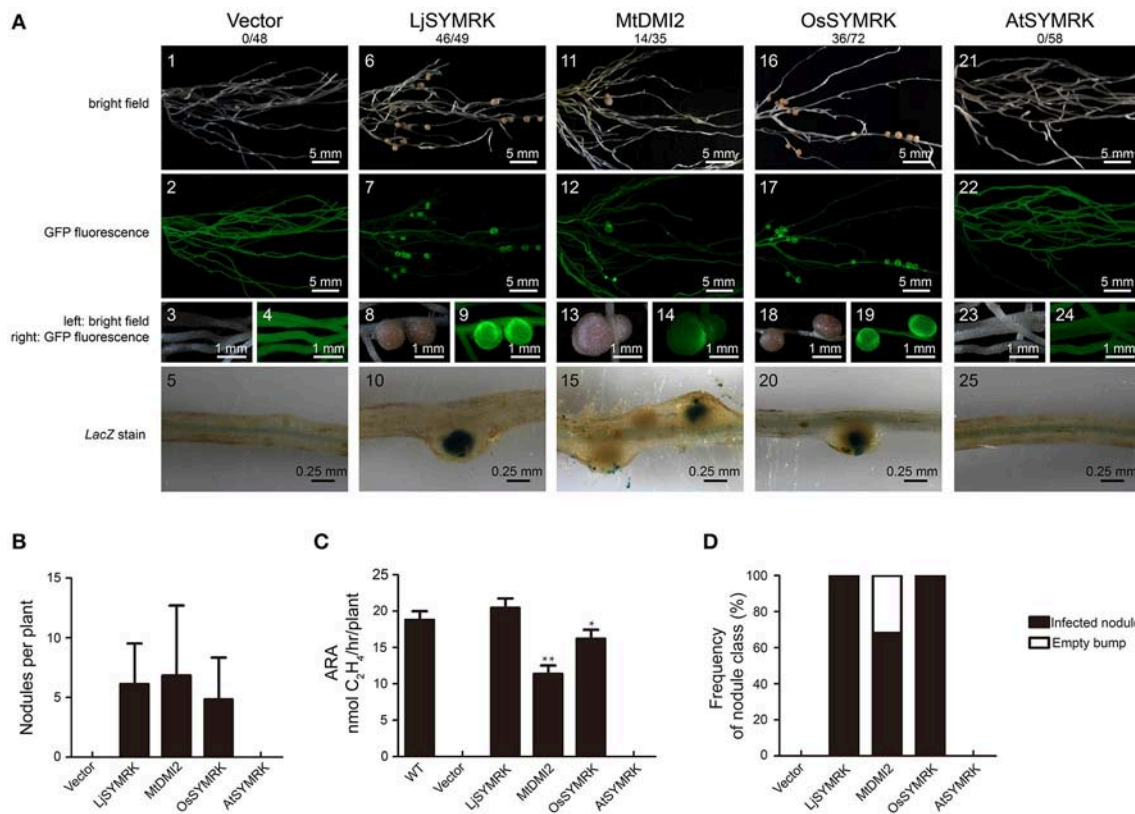
FIGURE 2 | Complementation of *Lotus symrk-409* mutants with LjSYMRK, MtDMI2, OsSYMRK, and AtSYMRK. **(A)** Representative micrographs of different transgenic nodules. Positive transgenic roots identified using GFP signals were inoculated with *M. loti* strain NZP2235 containing a *LacZ* report gene. Nodules and bumps were stained with X-gal. Images showed nodules produced on *symrk-409* roots transgenically expressed with empty control **(A1–5)**, *LjSYMRK* **(A6–10)**, *MtDMI2* **(A11–15)**, *OsSYMRK* **(A16–20)**, and *AtSYMRK* **(A21–25)**. Digital numbers on top of **(A)** indicate the number of nodulated plants out of the positive transgens. **(B)** Numbers of nodules and bumps generated on different transgenic roots 21 DPI with rhizobia. **(C)** Nitrogenase activity of root nodules 21 DPI with rhizobia determined using acetylene reduction method. *$P < 0.05$, **$P < 0.01$ indicate significant differences (*t*-test). **(D)** Frequency of nodules and bumps produced on the transgenic roots.

these three chimeric proteins could confer significant number of nodules formed on *symrk-409* roots 21 DPI with rhizobia (**Figures 4A,B**). However, MtDMI2^{ED}-LjSYMRK^{CD}-transgenic plants produced 30% bumps with no LacZ activity detected (**Figure 4D**) and lower nitrogenase activity than the other two transgenic nodules (**Figure 4C**). The ED of LjSYMRK was also replaced with ED of LjNFR1 or LjNFR5 to generate LjNFR1^{ED}-LjSYMRK^{CD} and LjNFR5^{ED}-LjSYMRK^{CD}, respectively, and transgenically expressed these chimeric proteins in *symrk-409* mutant plants. Surprisingly, normal nodules numbers and significantly high nitrogenase activities were measured in the *symrk-409* mutant transformed with either LjNFR1^{ED}-LjSYMRK^{CD} or LjNFR5^{ED}-LjSYMRK^{CD} (Figure S4). These data suggest that ED of SYMRK is important but might not be necessary while the CD of LjSYMRK is essential for RNS in *Lotus*.

MLD and LRR Domain Have a Fine-Tune Regulation on SYMRK-Mediated Symbiotic Response

Complementation with either MtDMI2 or MtDMI2^{ED}-LjSYMRK^{CD} in *symrk-409* plants could produce both effective

nodules and ineffective bumps. To test which domain is required for its function in negative regulate rhizobia infection, we made four truncated versions of LjSYMRK (Figure S3), i.e., LjSYMRK$^{\Delta ED}$ (LjSYMRK without ED), LjSYMRK$^{\Delta MLD}$ (LjSYMRK without MLD), LjSYMRK$^{\Delta LRR}$ (LjSYMRK without LRR domain and "GDPC" motif), and LjSYMRK^{GDLC} (Proline replaced with leucine). LjSYMRK deleted with MLD, or LRR, or ED could rescue the defects on nodulation phenotype in *symrk-409* mutant plants (**Figures 5A5–20**). The transgenic plants with overexpression of LjSYMRK$^{\Delta MLD}$, or LjSYMRK$^{\Delta LRR}$, or LjSYMRK$^{\Delta ED}$ formed a significant number of nodules with high nitrogenase activities measured (**Figures 5B–D**), suggesting that the EDs of SYMRK might not be necessary to be involved in symbiosis. However, LjSYMRK^{GDLC} (a mutation at the conserved residues "GDPC" motif) whose MLD was not detected to be cleaved during symbiosis could not completely complement *symrk-409* mutant plants with only ineffective bumps developed (**Figures 5A1–5,B–D**). These data suggest that MLD or LRR domain alone is dispensable but might work together to reduce rhizobial infection in mutant plants expressing LjSYMRK^{GDLC}.

FIGURE 3 | Complementation of *Lotus symrk-409* mutants with LjSYMRK^{ED}-MtDMI2^{CD}, LjSYMRK^{ED}-OsSYMRK^{CD}, and LjSYMRK^{ED}-AtSYMRK^{CD}. **(A)** Representative micrographs of different transgenic nodules. Positive transgenic roots identified using GFP signals were inoculated with *M. loti* strain NZP2235 containing a *LacZ* report gene. Nodules and bumps were stained with X-gal. Images showed nodules produced on *symrk-409* roots transgenically expressed LjSYMRK^{ED}-MtDMI2^{CD} **(A1–5)**, LjSYMRK^{ED}-OsSYMRK^{CD} **(A6–10)**, and LjSYMRK^{ED}-AtSYMRK^{CD} **(A11–15)**. Digital numbers on top of panel A indicate the number of nodulated plants out of total positive trasngens. **(B)** Numbers of nodules produced per transgenic roots 21 DPI with rhizobia. **(C)** Nitrogenase activities of root nodules 21 DPI with rhizobia determined using acetylene reduction method. **(D)** Frequency of nodules and bumps generated on the transgenic roots.

FIGURE 4 | Complementation of *Lotus symrk-409* mutants with MtDMI2^{ED}-LjSYMRK^{CD}, OsSYMRK^{ED}-LjSYMRK^{CD}, and AtSYMRK^{ED}-LjSYMRK^{CD}. **(A)** Representative micrographs of different transgenic nodules. Positive transgenic roots identified using GFP signals were inoculated with *M. loti* strain NZP2235 containing a *LacZ* report gene. Nodules and bumps were stained with X-gal. Images showed nodules produced on *symrk-409* roots transgenically expressed MtDMI2^{ED}-LjSYMRK^{CD} **(A1–5)**, OsSYMRK^{ED}-LjSYMRK^{CD} **(A6–10)**, and AtSYMR^{ED}-LjSYMRK^{CD} **(A11–15)**. Digital numbers on panel A indicate the number of nodulated plants per total number of GFP-positive plants. **(B)** Numbers of nodules produced per transgenic roots 21 DPI with rhizobia. **(C)** Nitrogenase activities of transgenic nodules 21 DPI with rhizobia determined using acetylene reduction method. **$P < 0.01$ (*t*-test). **(D)** Frequency of nodules and bumps generated on the transgenic roots.

The LRR Domain Plays a Crucial Role in LjSYMRK^{GDLC}-Mediated RNS

A previous conclusion indicated that mutation at the GDPC motif that connects MLD and LRR disrupts the cleavage of MLD from SYMRK leading to the block of RNS in *Lotus*. We sought to examine whether LRR domain is required for its function in symbiosis. Chimeric constructs containing LjSYMRK^{GDLC} with LRR domains of MtDMI2, OsSYMRK, and AtSYMRK were made to generate LjSYMRK^{GDLC}-MtLRR, LjSYMRK^{GDLC}-OsLRR, and LjSYMRK^{GDLC}-AtLRR, respectively (Figure S3), and functionally tested in *symrk-409* mutant plants using hairy root transformation approach. Transgenic roots overexpressing either LjSYMRK^{GDLC}-OsLRR or LjSYMRK^{GDLC}-AtLRR but not LjSYMRK^{GDLC}-MtLRR could generate effective nodules with high nitrogenase activity per root tested (**Figure 6**). These findings implicate that the LRR domain is critical among SYMRK homologous proteins in the RNS.

In order to verify the importance of LRR domains, the alignment of amino acid sequence of LRR domains of SYMRK from different plant species were made (Figure S5) and nine conserved residues were chosen for point mutation into alanine based on the construct containing LjSYMRK^{GDLC}. All these variants were transgenically overexpressed in *symrk-409* mutant plants and nodulation phenotype was assayed 21 DPI with rhizobia. Transgenic roots expressing LjSYMRK^{GDLC} with mutations at H437, P448, Y460, and E469, respectively, only developed ineffective nodules with no nitrogenase activity detected (**Figure 7**), which is similar to the phenotype on LjSYMRK^{GDLC}-expressing transgenic roots. However, mutations at five serine residues (S422, S450, S451, S455, and S470) in LjSYMRK^{GDLC} induced the formation of nodules in *symrk-409* mutant plants, some of which are rhizobia-infected nodules with measurable nitrogenase activity (**Figure 7**), indicating that these five serine residues are critical in SYMRK^{GDLC}.
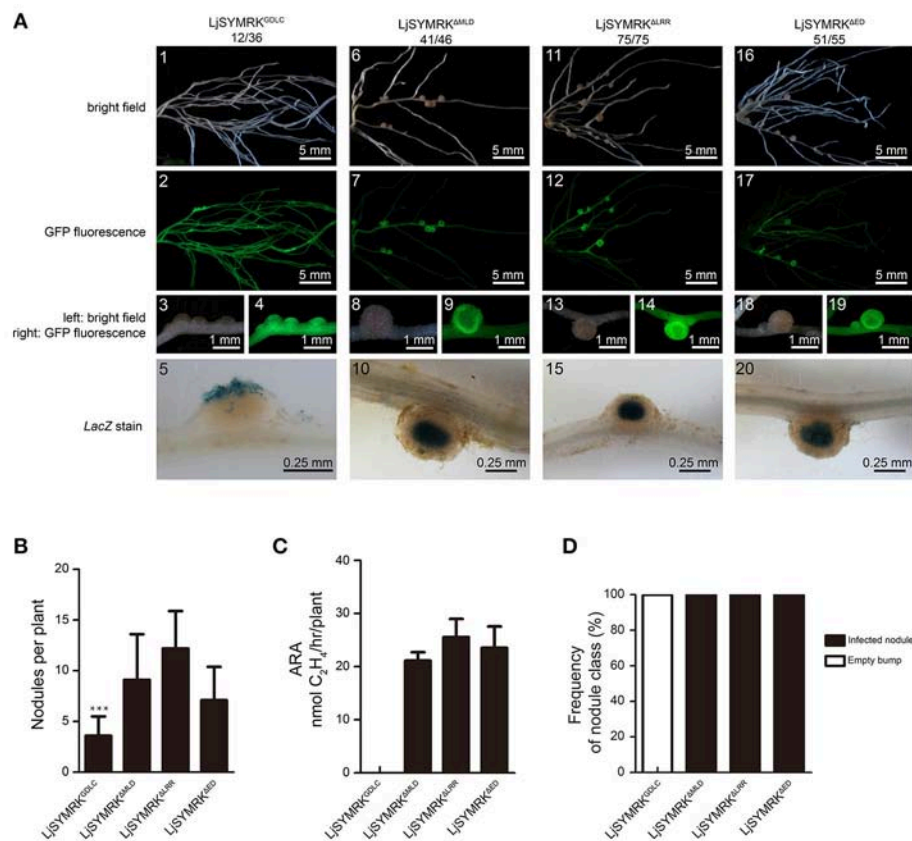
FIGURE 5 | Complementation of *Lotus symrk-409* mutants with LjSYMRK^{GDLC}, LjSYMRK$^{\Delta MLD}$, LjSYMRK$^{\Delta LRR}$, and LjSYMRK$^{\Delta ED}$. **(A)** Representative micrographs of different transgenic nodules. Positive transgenic roots identified using GFP signals were inoculated with *M. loti* strain NZP2235 containing a *LacZ* report gene. Nodules and bumps were stained with X-gal. Images showed nodules produced on *symrk-409* roots transgenically expressed LjSYMRK^{GDLC} **(A1–5)**, LjSYMRK$^{\Delta MLD}$ **(A6–10)**, LjSYMRK$^{\Delta LRR}$ **(A11–15)**, and LjSYMRK$^{\Delta ED}$ **(A16–20)**. Digital numbers on panel A indicate the number of nodulated plants per total positive roots. **(B)** Numbers of nodules and bumps produced on the transgenic roots 21 DPI with rhizobia. ***$P < 0.01$ (*t*-test). **(C)** Nitrogenase activities of transgenic nodules 21 DPI with rhizobia determined using acetylene reduction method. **(D)** Frequency of nodules and bumps generated on the transgenic roots.

DISCUSSION

SYMRK protein plays a crucial role in root nodule symbiosis in leguminous plants in response to compatible rhizobia. In this study, domain swap approach was used to study the function of different domains of SYMRK involved in RNS in *Lotus*. The broad goal of this study was to refine the molecular function of different domains of SYMRK in the contribution to root nodule symbiosis using domain swap approach in *Lotus japonicus*.

SYMRK is a key component in the symbiotic signaling pathway, which is essential for both RNS and AMS. In leguminous plants, such as in *Medicago*, MtDMI2 is required for both RNS and AMS, while in rice, OsSYMRK was only shown to be required for AMS and AtSYMRK was shown not to be involved in symbiosis in *Arabidopsis*. To test the roles of different domains of SYMRK in RNS, several chimeric SYMRK protein-encoding genes were constructed under the control of the *Ljubiquitin* promoter and multiple individual transgenic roots were used to test their functions in *Lotus*. The diverse function of SYMRK homologous proteins was further confirmed with the

finding that MtDMI2 and OsSYMRK but not AtSYMRK could complement *Lotus SYMRK$^{-/-}$* mutant to mediate RNS. The function of SYMRK proteins were tested in *Lotus SYMRK$^{-/-}$* plants in response to rhizobia. The partial complementation by MtDMI2 by forming some ineffective nodules suggests the slight different function of SYMRK between *Lotus* and *Medicago* might exist. Since *Medicago* forms indeterminate nodules while *Lotus* forms determinate nodules with rhizobia, it is possible that SYMRK-mediated RNS have a little difference within these two plant species. Domain swap approach by replacing the CD of LjSYMRK with CDs from other SYMRK homologs confirmed that AtSYMRK is not functional in Lotus to mediate RNS. While the ED of LjSYMRK seems to be not so important since replacing LjSYMRK^{ED} with ED from other SYMRK proteins or even from NFR1 and NFR5 is still functional to mediate RNS in *Lotus*. The critical function of SYMRK^{CD} in symbiosis was also reported that spontaneous nodule formation was observed in *Medicago* and *Lotus* when the kinase domain of AhSYMRK and the full length LjSYMRK were overexpressed, respectively (Ried et al., 2014; Saha et al., 2014). Chimeric protein

FIGURE 6 | Complementation of *Lotus symrk-409* mutants with LjSYMRK^{GDLC}-MtLRR, LjSYMRK^{GDLC}-OsLRR, and LjSYMRK^{GDLC}-AtLRR. **(A)** Representative micrographs of different transgenic nodules. Positive transgenic roots identified using GFP signals were inoculated with *M. loti* strain NZP2235 containing a *LacZ* report gene. Nodules and bumps were stained with X-gal. Images showed nodules produced on *symrk-409* roots transgenically expressed LjSYMRK^{GDLC}-MtLRR **(A1–5)**, LjSYMRK^{GDLC}-OsLRR **(A6–10)**, and LjSYMRK^{GDLC}-AtLRR **(A11–15)**. Digital numbers on **(A)** indicate the number of nodulated plants per total positive roots. **(B)** Numbers of nodules and bumps produced on the transgenic roots 21 DPI with rhizobia. **(C)** Nitrogenase activities of transgenic nodules 21 DPI with rhizobia determined using acetylene reduction method. **(D)** Frequency of nodules and bumps generated on the transgenic roots.

AtSYMRK^{ED}-LjSYMRK^{CD} but not full length of AtSYMRK could complement *Lotus SYMRK1$^{-/-}$* mutants indicates that the kinase domain of SYMRK is critical to induce nodule initiation even regardless of its ED or without rhizobia treatment when overexpressed in plants.

The ED of SYMRK contains two important domains, i.e., MLD and LRR linked by a GDPC motif. The well-established model about SYMRK reveals that cleavage of MLD is required for its association with NFR5 to mediate symbiotic signaling transduction in *Lotus* (Antolín-Llovera et al., 2014). It seems that the presence of MLD suppresses SYMRK-mediated symbiotic signaling since *Lotus symrk-14* mutants having a point mutation at the linker ("GDPC" motif was mutated into "GDLC") between MLD and LRR showed strong inhibition on bacterial infection but slight suppression on the development of nodules (Ried et al., 2014). The abnormal function of SYMRK-14 (or SYMRK^{GDLC}) in RNS was exaggerated with only ineffective nodules formed in *Lotus* when SYMRK^{GDLC} was overexpressed. However, overexpression of SYMRK^{GDLC} with deletion of either MLD or LRR domain only produced effective nodules, suggesting

that both MLD and LRR domain of SYMRK might play a negative rule but might work together in SYMRK^{GDLC}-mediated nodule organogenesis in *Lotus*. The same phenotype with effective nodule formation was also observed in Lotus when LjSYMRK^{GDLC} with its LRR replaced by the LRR domains from OsSYMRK or AtSYMRK but not from MtDMI2 was overexpressed, suggesting that LRR domains from OsSYMRK or AtSYMRK might be not functional in mediating RNS in Lotus. The negative role of LRR domain in SYMRK^{GDLC}-mediated response is further confirmed by point mutations at five conserved serine residues, leaving us a possibility that the negative function of LRR might be regulated by an unknown phosphorylation.

The ED of SYMRK seems to have a fine-tune regulation on the activation of symbiotic signaling in *Lotus*. Replacement ED with EDs from OsSYMRK or AtSYMRK proteins or even with NFR1 and NFR5 could contribute symbiotic response when overexpressed in *Lotus* plants. Since these proteins located far from LjSYMRK based on the phylogenetic tree, it is possible that these chimeric proteins with nonfunctional EDs still keep some function as LjSYMRK$^{\Delta ED}$ to induce symbiotic response. However, the chimeric protein MtDMI2^{ED}-LjSYMRK^{CD} could not completely complement Lotus *SYMRK$^{-/-}$* plants, indicating that the functional difference between *Lotus* and *Medicago* possibly due to that different nodule types formed on these two plants. The abovementioned hypothesis is further confirmed by the finding that LjSYMRK^{GDLC}-MtLRR but not LjSYMRK^{GDLC}-AtLRR or LjSYMRK^{GDLC}-OsLRR is not functional to mediate RNS in *Lotus*. The whole ED from LjSYMRK might be nonfunctional when LRR domain was replaced with LRR from AtSYMRK or OsSYMRK. However, since the LRR from MtDMI2 might have difference to mediate RNS, the MLD of LjSYMRK seems not to match with the LRR from MtDMI2 which finally made the chimeric protein LjSYMRK^{GDLC}-MtLRR lose its function in RNS. The data indicated that MLD and LRR must have a fine-tune regulation on the activation of symbiotic response.

In conclusion, the intracellular domain of SYMRK is critical while the ED of SYMRK plays a relatively minor role in mediating RNS in *Lotus*. GDPC motif that connects MLD and LRR domain required for MLD cleavage in response to RNS is responsible for nodule organogenesis. Both MLD and LRR domains might have a fine-tune regulation of SYMRK-mediated nodule organogenesis in *Lotus*. This progress in defining the molecular functions of SYMRK in RNS. However, identification of the direct targets of SYMRK and crystal structure of SYMRK are the future topics for understanding the molecular mechanisms mediated by SYMRK in NRS.

MATERIALS AND METHODS

Homology Analysis

SYMRK homologous proteins were identified based on the sequence similarity to *Lotus* SYMRK with the protein blast suite at the National Center for Biotechnology Information database (https://blast.ncbi.nlm.nih.gov) used. The phylogenetic tree was built by MEGA 5.1 software (Tamura et al., 2011).

FIGURE 7 | Complementation of *Lotus symrk-409* mutants with constructs of point mutations on the LRR domain of LjSYMRK^GDLC. **(A)** Representative micrographs of different transgenic nodules. Positive transgenic roots identified using GFP signals were inoculated with *M. loti* strain NZP2235 containing a *LacZ* report gene. Nodules and bumps were stained with X-gal. Images showed nodules produced on *symrk-409* roots transgenically expressed LjSYMRK^GDLC-S422A **(A1–5)**, LjSYMRK^GDLC-H437A **(A6–10)**, LjSYMRK^GDLC-P448A **(A11–15)**, LjSYMRK^GDLC-S450A **(A16–20)**, LjSYMRK^GDLC-S451A **(A21–25)**, LjSYMRK^GDLC-S455A **(A26–30)**, LjSYMRK^GDLC-Y460A **(A31–35)**, LjSYMRK^GDLC-E469A **(A36–40)**, and LjSYMRK^GDLC-S470A **(A41–45)**. Digital numbers on **(A)** indicate the number of nodulated plants per total positive roots. **(B)** Numbers of nodules and bumps produced on the transgenic roots 21 DPI with rhizobia. **(C)** Nitrogenase activities of transgenic nodules 21 DPI with rhizobia determined using acetylene reduction method. **P < 0.01 (*t*-test). **(D)** Frequency of nodules and bumps generated on the transgenic roots.

Plant Growth and Rhizobia Inoculation

Lotus seeds (B-129 Gifu and *symrk* mutant) and hair root transgenic seedlings were grown in vermiculite and perlite mixed (2:1 volume ratio) supplied a 1/2 B&D medium and placed in a growth chamber set at 22°C for 16-h-light/8-h-dark cycle. Plants were inoculated with *Mesorhizobium loti* strain NZP2235 expressing beta-galactosidase (LacZ) after 5 days. Rhizobia grown in liquid TY medium adding tetracycline at OD_{600} of 1.0. They were then pelleted and resuspended in 1/2 B&D medium containing 0.5 mM KNO_3 at OD_{600} of 0.02.

Medicago hair root transgenic seedlings were grown in vermiculite and perlite mixed (2:1 volume ratio) supplied a 1/2 FM medium and placed in a growth chamber set at 22°C for 16-h-light/8-h-dark cycle. Plants were inoculated with *Sinorhizobium meliloti* strain Sm2011 expressing LacZ after 5 days. Rhizobia grown in liquid TY medium adding tetracycline at OD_{600} of 1.0. They were then pelleted and resuspended in 1/2 FM medium containing 0.5 mM KNO_3 at OD_{600} of 0.02.

Lotus symrk Mutant Identification

Genomic DNA was extracted from putative mutant plants and used for PCR amplification of 2 min soak at 95°C, followed by 35 cycles of 30 s at 94°C, 30 s at 58°C, and 50 s at 72°C, followed by 5 min soak at 72°C. Primers for genotyping of *symrk-409* insertion alleles are provided in Table S1, which were performed as described previously (Fukai et al., 2012; Urbanski et al., 2012).

RNA Isolation and Quantitative RT-PCR

Total RNA was extracted from roots at 7 DPI using the EASYspin Plant RNA Kit (Aidlab, China). Primescript RT Reagent Kit (TaKaRa, Japan) was used to remove gDNA of RNA samples and reverse transcribe RNA into cDNA. Quantitative real-time PCR was performed on an ABI ViiA^TM7 Real-Time PCR System (ABI, USA) using One-Step SYBR PrimeScript RT-PCR kit II (Takara, Japan). *Lotus ATP synthase* (Genbank ID: AW719841) and *ubiquitin* (Genbank ID: AW720576) were used as reference genes, which are stably expressed in all plant tissues. Primers for qRT-PCR are listed in Table S1.

Plasmid Construction for Mutant Complementation

The pUB-GFP vector was digested by restriction endonucleases *Xba*I and *Stu*I, then linked to a fragment or multi-fragments by Gibson assembly (Gibson et al., 2009). Primers for PCR are listed in Table S1.

Lotus Hairy Root Transformation

The pUB-GFP vectors carrying a variety of chimeric SYMRK constructs were transformed into *symrk-409* seedlings using *Agrobacterium rhizogenes* LBA1334 as described by Wang et al. (2013). The GFP marker was used for selection of transgenic hairy roots by a fluorescence stereo microscope (Nikon SMZ18, Japan).

Medicago Hairy Root Transformation

Transgenic root on *dmi2-1* seedlings were induced using *A. rhizogenes* MSU440 as described by Wang et al. (2016). Positive transgenic hairy roots were identified by GFP marker using a fluorescence stereo microscope (Nikon SMZ18, Japan).

X-Gal Staining

Plants were inoculated with rhizobia carrying a *lacZ* reporter gene. At 7 or 21 DPI, roots were vacuum infiltrated for 5 min in fixative solution (1.25% Glutaraldehyde dissolved in 0.1 M potassium phosphate buffer pH 7.4), and placed at room temperature for 40 min, washed twice 10 min in 0.1 M potassium phosphate buffer. Roots were vacuum infiltrated for 5 min with X-Gal staining solution [0.1 M phosphate buffer, 6.25 mM $K_4Fe(CN)_6$, 6.25 mM $K_3Fe(CN)_6$, 0.75% X-Gal in DMF] and kept overnight at 28°C in the dark. Roots were washed twice in 0.1 M potassium phosphate buffer for 5 min, then rinsed twice in ddH_2O for 5 min. Stained roots were examined using a fluorescence stereo microscope (Nikon SMZ18, Japan) and a fluorescence microscope (LEICA DM2500, Germany).

Detection of Nitrogenase Activity

Acetylene reduction activity (ARA) was used to detect nitrogenase activity of root nodules at 21 DPI. Four nodulated seedlings were transferred into test tubes with 2 ml acetylene added for additional growth at 28°C for 2 h. At least three repeats of each experiment were analyzed. Acetylene was surveyed using a GC-4000A gas chromatograph (Dongxi Company, China).

ACCESSION NUMBERS

Sequence data from this work can be found under the following GenBank accession numbers: AAV88623.1 for SrSYMRK, XP_004512550.1 for CaSYMRK, CAD22013.1 for MaSYMRK, CAD10812.1 for PsSYMRK, XP_003517193.1 for GmSYMRK, CAD10811.1 for MtSYMRK, AAM67418.1 for LjSYMRK, NP_001234869.1 for SlSYMRK, NP_001105860.1 for ZmSYMRK, XP_015646949.1 for OsSYMRK, XP_014752741.1 for BdSYMRK, XP_002460874.1 for SbSYMRK, XP_020677592.1 for DcSYMRK, NP_564904.1 for AtSYMRK, XP_020890981.1 for AlSYMRK, XP_010511724.1 for CsSYMRK, XP_009105317.1 for BrSYMRK, and XP_018445268.1 for RsSYMRK.

AUTHOR CONTRIBUTIONS

HL, YC, and ZZ designed the experiments and analyzed the results. HL, MC, TZ, and LD performed the experiments and analyzed the data. HL, YC, and ZZ wrote the paper.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2018.00697/full#supplementary-material

REFERENCES

Ané, J. M., Lévy, J., Thoquet, P., Kulikova, O., de Billy, F., Penmetsa, V., 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. doi: 10.1094/MPMI.2002.15.11.1108

Antolín-Llovera, M., Ried, M. K., and Parniske, M. (2014). Cleavage of the SYMBIOSIS RECEPTOR-LIKE KINASE ectodomain promotes complex formation with Nod Factor Receptor 5. *Curr. Biol.* 24, 422–427. doi: 10.1016/j.cub.2013.12.053

Arrighi, J. F., Barre, A., Ben Amor, B., Bersoult, A., Soriano, L. C., Mirabella, R., et al. (2006). The *Medicago truncatula* lysine motif-receptor-like kinase gene family includes NFP and new nodule-expressed genes. *Plant Physiol.* 142, 265–279. doi: 10.1104/pp.106.084657

Broghammer, A., Krusell, L., Blaise, M., Sauer, J., Sullivan, J. T., Maolanon, N., et al. (2012). Legume receptors perceive the rhizobial lipochitin oligosaccharide signal molecules by direct binding. *Proc. Natl. Acad. Sci. U.S.A.* 109, 13859–13864. doi: 10.1073/pnas.1205171109

Fisher, R. F., and Long, S. R. (1992). Rhizobium–plant signal exchange. *Nature* 357, 655–660. doi: 10.1038/357655a0

Fukai, E., Soyano, T., Umehara, Y., Nakayama, S., Hirakawa, H., Tabata, S., et al. (2012). Establishment of a *Lotus japonicus* gene tagging population using the exon-targeting endogenous retrotransposon *LORE1*. *Plant J.* 69, 720–730. doi: 10.1111/j.1365-313X.2011.04826.x

Genre, A., and Russo, G. (2016). Does a common pathway transduce symbiotic signals in plant-microbe interactions? *Front. Plant Sci.* 7:96. doi: 10.3389/fpls.2016.00096

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A. III, and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345. doi: 10.1038/nmeth.1318

Kawaharada, Y., Kelly, S., Nielsen, M. W., Hjuler, C. T., Gysel, K., Muszynski, A., et al. (2015). Receptor-mediated exopolysaccharide perception controls bacterial infection. *Nature* 523, 308–312. doi: 10.1038/nature14611

Kawaharada, Y., Nielsen, M. W., Kelly, S., James, E. K., Andersen, K. R., Rasmussen, S. R., et al. (2017). Differential regulation of the Epr3 receptor coordinates membrane-restricted rhizobial colonization of root nodule primordia. *Nat. Commun.* 8:14534. doi: 10.1038/ncomms14534

Kosuta, S., Held, M., Hossain, M. S., Morieri, G., MacGillivary, A., Johansen, C., et al. (2011). *Lotus japonicus symRK-14* uncouples the cortical and epidermal symbiotic program. *Plant J.* 67, 929–940. doi: 10.1111/j.1365-313X.2011.04645.x

Kouchi, H., Imaizumi-Anraku, H., Hayashi, M., Hakoyama, T., Nakagawa, T., Umehara, Y., et al. (2010). How many peas in a pod? Legume genes responsible for mutualistic symbioses underground. *Plant Cell Physiol.* 51, 1381–1397. doi: 10.1093/pcp/pcq107

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. doi: 10.1126/science.1090074

Madsen, E. B., Madsen, L. H., Radutoiu, S., Olbryt, M., Rakwalska, M., Szczyglowski, K., et al. (2003). A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals. *Nature* 425, 637–640. doi: 10.1038/nature02045

Maekawa, T., Maekawa-Yoshikawa, M., Takeda, N., Imaizumi-Anraku, H., Murooka, Y., and Hayashi, M. (2009). Gibberellin controls the nodulation signaling pathway in *Lotus japonicus*. *Plant J.* 58, 183–194. doi: 10.1111/j.1365-313X.2008.03774.x

Małolepszy, A., Mun, T., Sandal, N., Gupta, V., Dubin, M., Urbanski, D., et al. (2016). The *LORE1* insertion mutant resource. *Plant J.* 88, 306–317. doi: 10.1111/tpj.13243

Markmann, K., Giczey, G., and Parniske, M. (2008). Functional adaptation of a plant receptor-kinase paved the way for the evolution of intracellular root symbioses with bacteria. *PLoS Biol.* 6:e68. doi: 10.1371/journal.pbio.0060068

Markmann, K., and Parniske, M. (2009). Evolution of root endosymbiosis with bacteria: how novel are nodules? *Trends Plant Sci.* 14, 77–86. doi: 10.1016/j.tplants.2008.11.009

Miwa, H., Sun, J., Oldroyd, G. E. D., and Downie, J. A. (2006). Analysis of Nod-factor-induced calcium signaling in root hairs of symbiotically defective mutants of *Lotus japonicus*. *Mol. Plant Microbe Interact.* 19, 914–923. doi: 10.1094/MPMI-19-0914

Moling, S., Pietraszewska-Bogiel, A., Postma, M., Fedorova, E., Hink, M. A., Limpens, E., et al. (2014). Nod factor receptors form heteromeric complexes and are essential for intracellular infection in *Medicago* nodules. *Plant Cell* 26, 4188–4199. doi: 10.1105/tpc.114.129502

Oldroyd, G. E. (2013). Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nat. Rev. Microbiol.* 11, 252–263. doi: 10.1038/nrmicro2990

Oldroyd, G. E., and Downie, J. M. (2008). Coordinating nodule morphogenesis with rhizobial infection in legumes. *Annu. Rev. Plant Biol.* 59, 519–546. doi: 10.1146/annurev.arplant.59.032607.092839

Parniske, M. (2008). Arbuscular mycorrhiza: the mother of plant root endosymbioses. *Nat. Rev. Microbiol.* 6, 763–775. doi: 10.1038/nrmicro1987

Paul, A., Samaddar, S., Bhattacharya, A., Banerjee, A., Das, A., Chakrabarti, S., et al. (2014). Gatekeeper tyrosine phosphorylation is autoinhibitory for Symbiosis Receptor Kinase. *FEBS Lett.* 588, 2881–2889. doi: 10.1016/j.febslet.2014.06.056

Radutoiu, S., Madsen, L. H., Madsen, E. B., Felle, H. H., Umehara, Y., Grønlund, M., et al. (2003). Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *Nature* 425, 585–592. doi: 10.1038/nature02039

Remy, W., Taylor, T. N., Hass, H., and Kerp, H. (1994). Four hundred-million-year-old vesicular arbuscular mycorrhizae. *Proc. Natl. Acad. Sci. U.S.A.* 91, 11841–11843. doi: 10.1073/pnas.91.25.11841

Ried, M. K., Antolín-Llovera, M., and Parniske, M. (2014). Spontaneous symbiotic reprogramming of plant roots triggered by receptor-like kinases. *eLife* 3:e03891. doi: 10.7554/eLife.03891

Saha, S., Dutta, A., Bhattacharya, A., and DasGupta, M. (2014). Intracellular catalytic domain of symbiosis receptor kinase hyperactivates spontaneous nodulation in absence of rhizobia. *Plant Physiol.* 166, 1699–U1826. doi: 10.1104/pp.114.250084

Saha, S., Paul, A., Herring, L., Dutta, A., Bhattacharya, A., Samaddar, S., et al. (2016). Gatekeeper tyrosine phosphorylation of SYMRK is essential for synchronizing the epidermal and cortical responses in root nodule symbiosis. *Plant Physiol.* 171, 71–81. doi: 10.1104/pp.15.01962

Samaddar, S., Dutta, A., Sinharoy, S., Paul, A., Bhattacharya, A., Saha, S., et al. (2013). Autophosphorylation of gatekeeper tyrosine by symbiosis receptor kinase. *FEBS Lett.* 587, 2972–2979. doi: 10.1016/j.febslet.2013.07.050

Shiu, S.-H., and Bleecker, A. B. (2001). Receptor-like kinases from *Arabidopsis* form a monophyletic gene family related to animal receptor kinases. *Proc. Natl. Acad. Sci. U.S.A.* 98, 10763–10768. doi: 10.1073/pnas.181141598

Stracke, S., Kistner, C., Yoshida, S., Mulder, L., Sato, S., Kaneko, T., et al. (2002). A plant receptor-like kinase required for both bacterial and fungal symbiosis. *Nature* 417, 959–962. doi: 10.1038/nature00841

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739. doi: 10.1093/molbev/msr121

Urbanski, D. F., Małolepszy, A., Stougaard, J., and Andersen, S. U. (2012). Genome-wide *LORE1* retrotransposon mutagenesis and high-throughput insertion detection in *Lotus japonicus*. *Plant J.* 69, 731–741. doi: 10.1111/j.1365-313X.2011.04827.x

Wang, C., Yu, H., Luo, L., Duan, L., Cai, L., He, X., et al. (2016). NODULES WITH ACTIVATED DEFENSE 1 is required for maintenance of rhizobial endosymbiosis in *Medicago truncatula*. *New Phytol.* 212, 176–191. doi: 10.1111/nph.14017

Wang, C., Zhu, H., Jin, L., Chen, T., Wang, L., Kang, H., et al. (2013). Splice variants of the SIP1 transcripts play a role in nodule organogenesis in *Lotus japonicus*. *Plant Mol. Biol.* 82, 97–111. doi: 10.1007/s11103-013-0042-3

Yoro, E., Suzaki, T., Toyokura, K., Miyazawa, H., Fukaki, H., and Kawaguchi, M. (2014). A positive regulator of nodule organogenesis, NODULE INCEPTION, acts as a negative regulator of rhizobial infection in *Lotus japonicus*. *Plant Physiol.* 165, 747–758. doi: 10.1104/pp.113.233379

Young, N. D., Debellé, F., Oldroyd, G. E. D., Geurts, R., Cannon, S. B., Udvardi, M. K., et al. (2011). The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature* 480, 520–524. doi: 10.1038/nature10625

Zhang, X., Dong, W., Sun, J., Feng, F., Deng, Y., He, Z., et al. (2015). The receptor kinase CERK1 has dual functions in symbiosis and immunity signalling. *Plant J.* 81, 258–267. doi: 10.1111/tpj.12723

Evolutionary History of Plant LysM Receptor Proteins Related to Root Endosymbiosis

Clare Gough*, Ludovic Cottret, Benoit Lefebvre and Jean-Jacques Bono

LIPM, Université de Toulouse, INRA, CNRS, Castanet-Tolosan, France

LysM receptor-like kinases (LysM-RLKs), which are specific to plants, can control establishment of both the arbuscular mycorrhizal (AM) and the rhizobium-legume (RL) symbioses in response to signal molecules produced, respectively, by the fungal and bacterial symbiotic partners. While most studies on these proteins have been performed in legume species, there are also important findings that demonstrate the roles of LysM-RLKs in controlling symbiosis in non-legume plants. Phylogenomic studies, which have revealed the presence or absence of certain LysM-RLKs among different plant species, have provided insight into the evolutionary mechanisms underlying both the acquisition and the loss of symbiotic properties. The role of a key nodulation LysM-RLK, NFP/NFR5, in legume plants has thus probably been co-opted from an ancestral role in the AM symbiosis, and has been lost in most plant species that have lost the ability to establish the AM or the RL symbiosis. Another LysM-RLK, LYK3/NFR1, that controls the RL symbiosis probably became neo-functionalised following two rounds of gene duplication. Evidence suggests that a third LysM-RLK, LYR3/LYS12, is also implicated in perceiving microbial symbiotic signals, and this protein could have roles in symbiosis and/or plant immunity in different plant species. By focusing on these three LysM-RLKs that are widespread in plants we review their evolutionary history and what this can tell us about the evolution of both the RL and the AM symbioses.

Keywords: LysM domain, kinase domain, NFP, LYR3, LYK3, nodulation, arbuscular mycorrhization, lipo-chitooligosaccharide

INTRODUCTION

In plants, several LysM receptor-like kinases (LysM-RLKs) have been characterised as symbiotic receptor proteins. After an introduction about the evolutionary origins of LysM-RLKs, we focus on three LysM-RLKs, known as NFP/NFR5, LYK3/NFR1, and LYR3/LYS12 in *Medicago truncatula/Lotus japonicus*, and their hypothetical evolutionary histories related to root endosymbiosis establishment. Other LysM-RLKs described in the text are listed in the glossary.

EVOLUTIONARY ORIGINS OF LYSM-RLKS

Proteins incorporating three extracellular LysM domains, a transmembrane domain and an intracellular kinase domain (**Figure 1A**) are specific to plants, and result from evolutionary events that apparently predate plant colonization of the land (Delaux et al., 2015). Whether the LysM triplet was only formed once is difficult to know, but LysM1, LysM2, and LysM3 of each protein

usually resemble more the same domain of homologous LysM-RLK proteins than other domains of the same protein (Arrighi et al., 2006). LysM domains are of bacterial origin, but unlike prokaryotic proteins with repeated LysM domains, those of LysM-RLKs are always separated by conserved cysteine-X-cysteine motifs, responsible for the formation of disulphide bridges (Lefebvre et al., 2012).

Three types of kinase domain can be distinguished for LysM-RLKs, two of which are predicted to have kinase activity, while a third group can be considered to be pseudo-kinases (Arrighi et al., 2006). These all have a common evolutionary origin, since all kinase domains of plant RLKs form a monophyletic family within the superfamily of plant kinases (Shiu and Bleecker, 2001). Symbiotic LysM-RLKs have so far been identified as "LYK-type" (active kinase) or "LYR-type" (pseudo-kinase) proteins. The presence of both types in all higher plants, together with strong conservation of intron-exon structure, indicates their ancient evolutionary origin and ancient events of duplication. Canonical 3-D structures can be predicted for LYR-type kinase domains (Arrighi et al., 2006), indicating that they evolved from active kinase domains. From our sequence analyses, the first changes were probably loss of the glycine-rich loop and the "DFG" motif at the start of the activation loop, which are ubiquitously absent in such proteins. Subsequently, one evolutionary line leads to proteins such as NFP that have lost the activation loop in their kinase domains, and another leads to most other LYR-type proteins, which have a conserved activation loop. The loss of the activation loop in NFP is common to all plant species analysed here, as shown in the sequence logos in **Figure 1B**, which include the basal flowering plant *Amborella trichopoda* estimated to have split from other flowering plants 145-210 MYA(Aoki et al., 2004). These sequence logos also show good sequence conservation in the activation loop region among all the LYR3 proteins analysed (**Figure 1B**).

THE EVOLUTIONARY ACQUISITION AND LOSS OF SYMBIOTIC NFP PROTEINS

NFP proteins have been shown to play important roles in establishment of the arbuscular mycorrhizal (AM) symbiosis in the non-legume dicot plants *Parasponia andersonni* (Op den Camp et al., 2010) and *Solanum lycopersicum* (Buendia et al., 2016), and they control nodulation in model legume plants (Gough and Cullimore, 2011). There is a lot of evidence that NFP proteins first controlled the AM symbiosis that evolved approximately 400 MYA, and were subsequently recruited by legume plants so that bacteria could activate symbiotic signalling pathways that were adapted to allow nodulation (Geurts et al., 2016), which appeared approximately 65 MYA. The presence of a pseudo-kinase in a LysM-RLK (LYR-type) involved in both symbioses suggests signal transduction occurs via protein/protein interactions. At least in the model legumes *M. truncatula* and *L. japonicus*, it is likely that heterodimerisation of NFP/NFR5 proteins with the kinase-active LYK3/NFR1 LysM-RLK forms a receptor complex (Madsen et al., 2011; Pietraszewska-Bogiel et al., 2013; Moling et al., 2014).

Legume NFP proteins are essential for the perception of rhizobial Nod factors, which are lipo-chitooligosaccharide (LCO) signal molecules, consisting of an *N*-acetyl glucosamine backbone and an acyl chain. In acquiring the ability to synthesise LCOs, rhizobial bacteria mimicked the production of LCOs, the so-called Myc-LCOs, produced by AM fungi (Maillet et al., 2011). LysM domains of bacterial proteins recognise peptidoglycan that contains a backbone alternating *N*-acetyl glucosamine and *N*-acetyl muramic acid. Appropriate changes to the module of three LysM domains that enabled recognition of an LCO molecule must therefore have been selected during evolution, and, for the rhizobium-legume (RL) symbiosis, there were presumably further selections for recognition of Nod factor decorations that are important in host specificity. Studies have identified critical residues in LysM2 of both MtNFP and LjNFR5. For MtNFP, this residue (L154) is different in orthologues of closely-related legume species nodulated by rhizobia producing different Nod factor structures (Bensmihen et al., 2011). Using complementation tests, the switch to L154 in the *Pisum sativum* NFP ortholog enabled *M. truncatula* mutant plants to be nodulated (Bensmihen et al., 2011). In LjNFR5, the L118 residue might define specificity towards decorations present on the non-reducing end of Nod factors (Radutoiu et al., 2007), and Nod factor binding to an individual LysM2 of LjNFR5 was reported (Sorensen et al., 2014). The sequence logos we generated with LysM domains of NFP proteins show that these two LysM2 residues implicated in the biological function of NFP proteins are both variable, compatible with direct or indirect roles in specific ligand recognition (**Figure 1C**). Using a sliding window approach in *L. japonicus*, some evidence was also provided for positive selection in LjNFR5 LysM2 (Lohmann et al., 2010). No significant signatures of selection were found in a study of nucleotide diversity of NFP in 30 genotypes of *M. truncatula* (De Mita et al., 2007).

During legume evolution, NFP became indispensable for the RL symbiosis, but not for the AM symbiosis, at least in model legumes (Ben Amor et al., 2003; Madsen et al., 2003). This suggests differences in activation of AM signalling in legume and non-legume plants. In addition, AM signalling activation might be different in monocots, despite the AM symbiosis appearing well before the monocot-dicot split, since there is no observable AM phenotype for *Osnfr5* knock-out mutants. However, these mutants show reduced symbiotic gene expression, and a chimeric receptor consisting of the extracellular domain of LjNFR5 and the intracellular domain of OsNFR5 complements an *Ljnfr5* mutant for rhizobial symbiosis, indicating some conservation of function (Miyata et al., 2016).

NFP and LYR3 are in different clades of the *M. truncatula* LysM-RLK family (Bono et al., 2018). Nevertheless, *NFP* genes are often in tandem with a *LYR3* gene, suggesting an ancient common evolutionary history, and a possible symbiotic role of the ancestral gene. This idea is reinforced by the finding that LYR3 proteins from *M. truncatula, L. japonicus* (LYS12)*, P. sativum, Glycine max, and Phaseolus vulgaris* are high affinity LCO binding proteins (Fliegmann et al., 2013; Malkov et al., 2016). Furthermore, Malkov et al. (2016) identified residues that are good candidates to be involved in a high affinity LCO binding

FIGURE 1 | Schematic presentation of a LysM-RLK protein structure **(A)** and sequence logos generated using Clustal W (Larkin et al., 2007) and WebLogo software (Schneider and Stephens, 1990; Crooks et al., 2004), and approximately 60 Angiosperm protein sequences each of NFP or LYR3 **(B,C)** (**Supplementary Table S1** and **Supplementary Data Sheet S1**). The sequence logos show the extent of sequence conservation within **(B)** part of the intracellular domain that contains the activation loop in LYR3 proteins but not in NFP proteins; and **(C)** individual LysM domains (LysM2 for NFP proteins and LysM3 for LYR3 proteins). In **(B)** the conserved "F" at position 6 of the sequence logos corresponds to the "F" in the "DFG" motif of active kinases. The amino acid positions highlighted in **(C)** correspond to L118 in LjNFR5 LysM2; L154 in MtNFP LysM2 and Y228 in MtLYR3 LysM3, as discussed in the text. The LysM sequence logos also show that each LysM domain is a combination of highly conserved and highly variable residues, many of which are probably involved in tertiary structure and ligand binding specificity, respectively. Sequence logos of all the 3 LysM domains of NFP and LYR3 are shown in **Supplementary Figures S1**, **S2**.

site in LysM3, including a tyrosine residue, and their data suggest that loss of one or more of these residues in the non AM species *Lupinus angustifolius*, has led to loss of LCO binding. In our sequence logo analysis on LysM domains of LYR3 proteins (**Figure 1C**), this tyrosine residue is relatively well conserved across LYR3 proteins in plants, suggesting conservation of ligand binding properties. Although MtLYR3 can bind both Nod factors and Myc-LCOs, this protein is not indispensable for either symbiosis, but the LCO binding properties and reported interaction between MtLYR3 and MtLYK3 (Fliegmann et al., 2016) are suggestive of a symbiotic-related function.

To get further insights into the evolutionary histories of NFP and LYR3 proteins, we constructed phylogenetic trees with approximately 60 protein sequences each of NFP and LYR3 orthologues from 64 Angiosperm species (**Supplementary Table S1** and **Supplementary Data Sheet S1**). In the *Fabaceae* family, traces of the whole genome duplication event, dated at approximately 60 MYA (Cannon et al., 2015), can be seen (**Figure 2**). In some species the second copy of NFP (called LYR1 in *M. truncatula*) has been maintained (**Figure 2A**), while in other species, there is a second copy of LYR3 (called LYR3-2 in *Lupinus angustifolius*) (**Figure 2B**). No symbiotic role has been shown for any of these second copies, and divergence from consensus

sequences could suggest a process of pseudogenisation. The presence of two copies each of NFP and LYR3 in *Populus* species could be due to the Salicoid genome duplication ancestral to speciation in this family, approximately 58 MYA (Harikrishnan et al., 2015). Other more recent events of genome duplication or hybridisation, have led to polyploidy in *G. max* and *Triticum aestivum*, respectively, and consequent additional copies of NFP and LYR3. Other examples of duplications include the two significantly different copies of NFP in both *Cucumis sativus* and *Cucumis melo*, and the three divergent copies of LYR3 encoded in tandem in *Beta vulgaris*.

A particularly interesting example of NFP duplication is in Rosales plants, where the *NFP2* gene was recently identified in three Parasponia species, and with high sequence divergence compared to the *NFP1* gene in these plants (van Velzen et al., 2018). Importantly, Van Velzen et al. found that NFP2 is absent in closely related non-nodulating Trema species (and also in other non-nodulating Rosaceae spp. like *Prunus persica*, *Fragaria vesca* and *Malus domesticus*), and that legume NFP proteins correspond to Parasponia NFP2. These data strongly suggest an importance of NFP2 for the ability of Parasponia, but not closely related plants, to nodulate. Furthermore, the absence of NFP2, as well as other key symbiotic genes, in
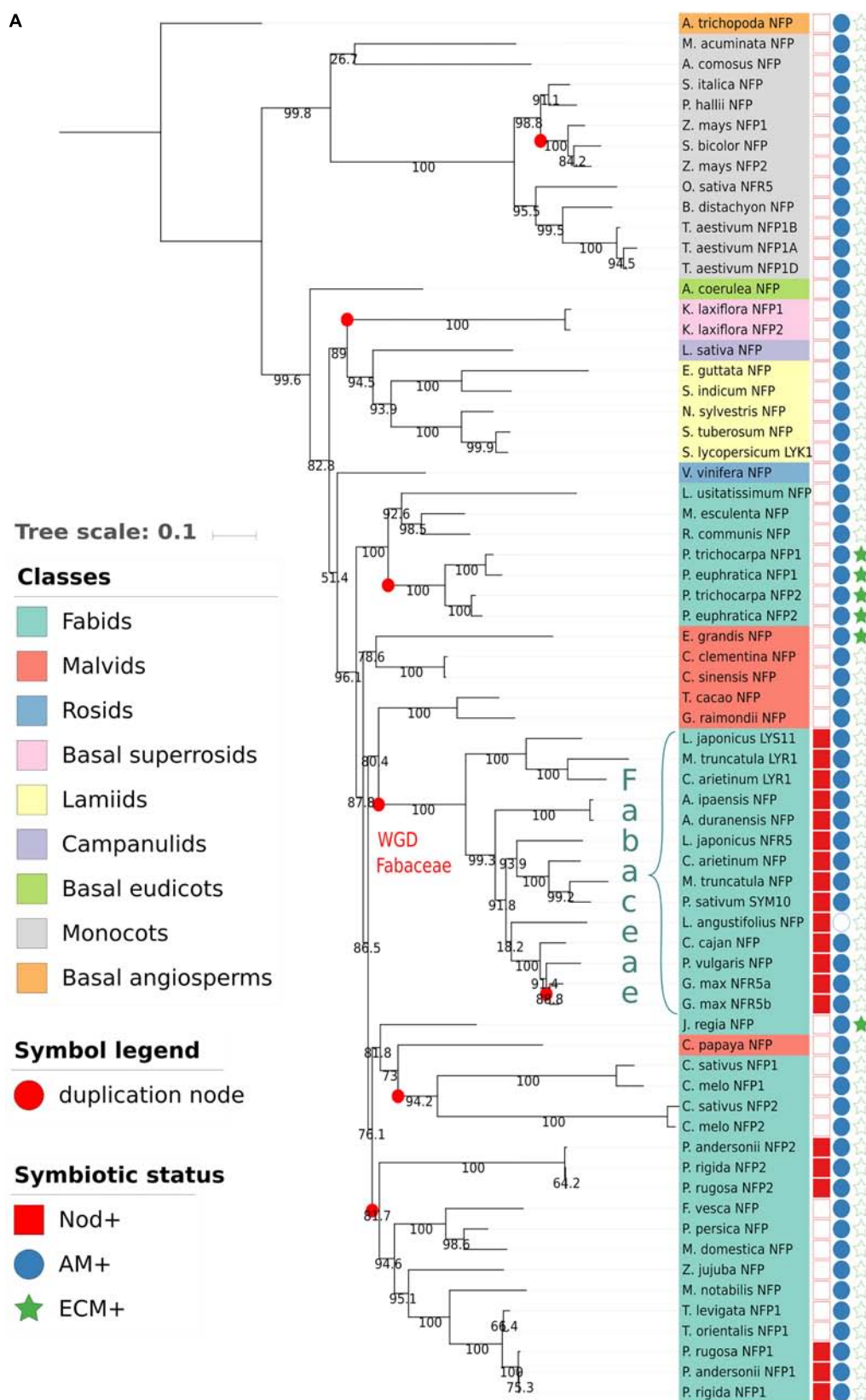
FIGURE 2 | Continued
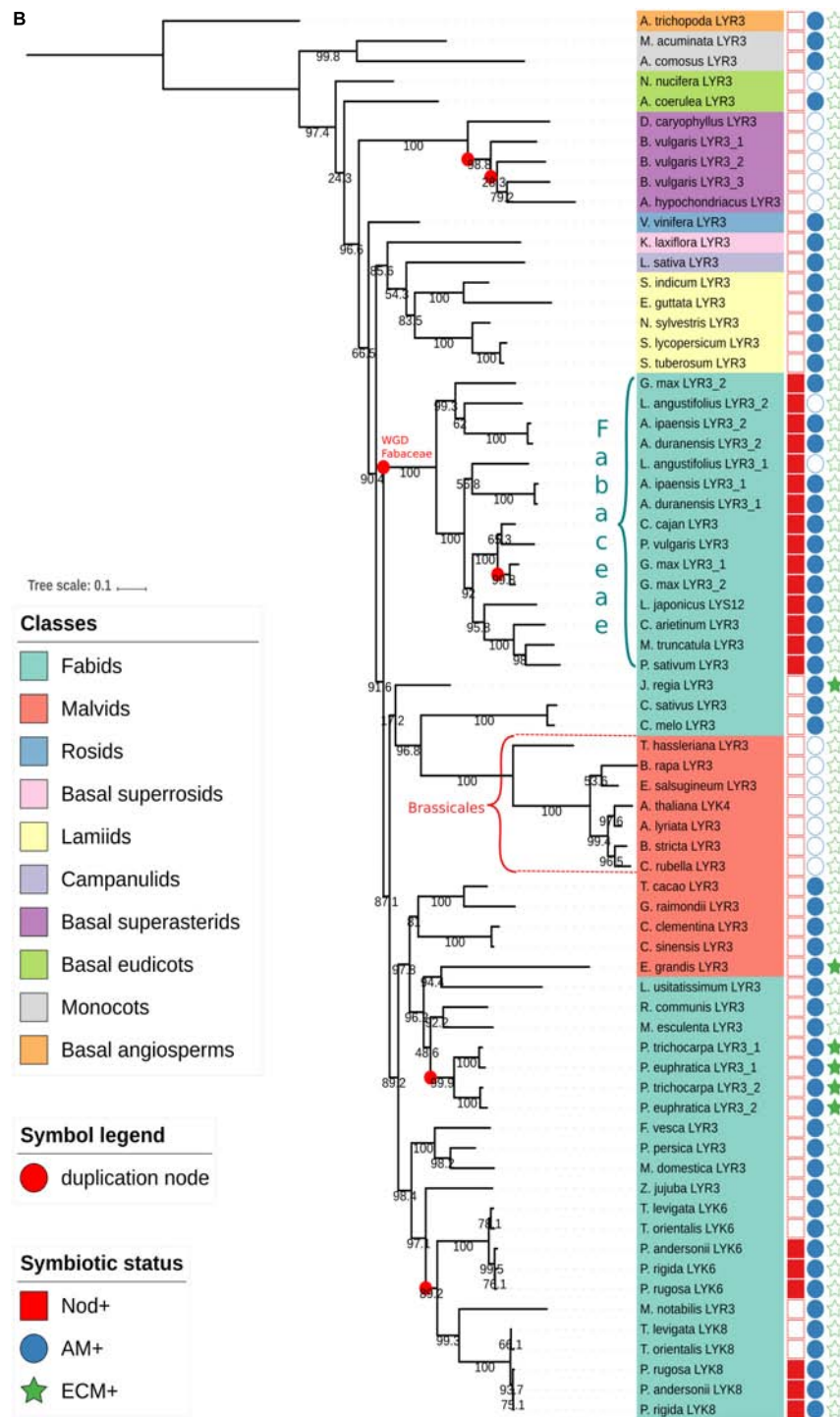
FIGURE 2 | Phylogenetic trees generated from approximately 60 Angiosperm sequences of **(A)** NFP proteins or **(B)** LYR3 proteins (**Supplementary Table S1** and **Supplementary Data Sheet S1**). The sequences were aligned by Mafft (v7.271) (Katoh et al., 2002) with the following parameters: –maxiterate 1000 –retree 1 –genafpair. Sites with more than 50% of gaps were pruned from the alignments. The phylogenetic trees were computed with IQ-Tree (Nguyen et al., 2015) with the following parameters: -nt AUTO -bb 1000 -alrt 1000. The trees and metadata were visualised with Itol (Letunic and Bork, 2007). Branches with alrt support values (Anisimova and Gascuel, 2006) less than 0.75 were collapsed. Support values are displayed on the internal branches. The Amborella trichopoda LYR3 sequence was considered as an outgroup for the NFP tree, and the Amborella trichopoda LYR3 sequence was considered as an outgroup for the LYR3 tree. These proteins were chosen as outgroups as Amborella trichopoda is a basal Angiosperm species. An interactive view for the NFP tree is available at this url: http://itol.embl.de/tree/14799102217475161526387035. An interactive view for the LYR3 tree is available at this url: http://itol.embl.de/tree/14799102217474321526387017

Trema species, led van Velzen et al. (2018) to challenge the long-standing hypothesis on the evolution of nitrogen-fixing symbioses, pointing towards massive loss of nodulation rather than many events of parallel acquisition following a single hypothetical predisposition event in the common ancestor of the nitrogen-fixing clade.

It is also noteworthy that LYR3 was probably duplicated in a common ancestor of extant Rosales plant species, and two, divergent copies (LYK6 and LYK8) have been retained in both Parasponia and Trema species but not in other closely related species (van Velzen et al., 2018). Of these two proteins, LYK6 is more divergent between Parasponia and Trema species, suggesting different selection pressures on LYK6, but not LYK8, between these nodulating and non-nodulating plants. Taken together, ancestral events of gene duplication of NFP and LYR3 have usually been followed either by loss of one copy or by strong divergence of the two copies. In the case of preferential retention of both copies, which could have led to neofunctionalisation, NFP1 and NFP2 in Parasponia species are excellent candidates for such a scenario.

Consistent with the symbiotic roles of NFP proteins, plant species such as Arabidopsis thaliana that have lost the ability to establish the AM symbiosis, have lost NFP (Delaux et al., 2014). Exceptions to this rule are non-mycorrhizal lupin species, which have retained NFP to control the RL symbiosis. Thus in most Brassicales species, only LYR3 is present. However, in Carica papaya, which is a mycorrhizal species, LYR3 is absent, but NFP was found as reported before (Delaux et al., 2014). Also, only NFP was clearly found in Poaceae species such as rice, maize, and sorghum, but we found NFP and LYR3 in the monocots Musa acuminata and Ananas comosus, and in Amborella trichopoda, which is basal to both monocots and dicots.

Considering the trees of NFP and LYR3 proteins, each one has two clades in which sequences are phylogenetically close to each other and that are significantly phylogenetically distant from the other sequences in the tree. For NFP, these two clades correspond to Fabaceae and monocot proteins (**Figure 2A**), and for LYR3, they correspond to Fabaceae and Brassicales proteins (**Figure 2B**). This suggests that these groups of proteins have undergone significantly different evolutionary events, leading to specific and well conserved functions or properties. Nodulation is a likely explanation for the Fabaceae NFP proteins, while this divergence of Brassicales LYR3 proteins is correlated with absence of the AM symbiosis. Therefore, it is possible that the loss of the AM symbiosis was the driving force for a change of function for LYR3 in these plants. Since the A. thaliana LYR3 protein (called AtLYK4) controls chitin oligomer (CO) perception for immunity (Wan et al., 2012), a significant change to Brassicales LYR3 proteins could have been in terms of the ligand binding site predicted on LysM2 (Tanaka et al., 2013). The lack of LCO binding to LYR3 proteins from lupin species (Malkov et al., 2016) could also have evolved following loss of the AM symbiosis in these plants. LYR3 proteins of other non-AM species (Dianthus caryophyllus, Beta vulgaris, Amaranthus hypochondriacus and Nelumbo nucifera) do not group with the Brassicales LYR3 proteins, probably because of multiple, independent losses of the AM symbiosis in higher plants.

THE EVOLUTION OF KINASE-ACTIVE LYSM-RLKS WITH SYMBIOTIC ROLES

LysM-RLK proteins with an active kinase domain also control symbiosis in legume and non-legume plants. OsCERK1 in Oryza sativa controls the AM symbiosis (Miyata et al., 2014; Zhang et al., 2015). OsCERK1 has a common ancestor with MtLYK3 and LjNFR1 that control nodulation in M. truncatula and L. japonicus, respectively (Gough and Cullimore, 2011). Phylogeny studies suggest that the common ancestor of MtLYK3 and LjNFR1 became neofunctionalised relatively recently following two rounds of legume-specific tandem duplications (De Mita et al., 2014). There is evidence of positive selection pressure for 3 amino acid residues in LysM1 of LYK3 (De Mita et al., 2014), compatible with data showing that MtLYK3 intervenes in controlling host range specificity (Limpens et al., 2003; Smit et al., 2007), and suggesting a role of LysM1 in Nod factor recognition. Similarly, Sulima et al. found evidence for strong selection pressure on LysM1 of MtLYK3, and also on LysM2 of a paralogous protein, MtLYK2 (Sulima et al., 2017). In P. sativum, evidence was found that PsLykX, which is encoded in a region syntenic to MtLYK2 and MtLYK3, is under positive selection in the extracellular region, mainly LysM1 again, compatible with a role in Nod factor recognition. This is reinforced by the higher allelic variation in PsLykX compared to highly homologous proteins, and one haplotype of PsLykX corresponds to nodulation specificity for a certain Nod factor structure (Sulima et al., 2017).

Chimeric protein studies indicate that the kinase "YAQ" sequence in LjNFR1 is important for activation of downstream symbiotic signalling leading to nodulation (Nakagawa et al., 2011). Phylogenetic studies suggest that the YAQ motif has an ancestral origin, predating the origin of nodulation (De Mita et al., 2014). Consistent with this and probably extending the role of the YAQ motif to the AM symbiosis, a very similar motif (YAR) is present in OsCERK1. Furthermore, the kinase domain of OsCERK1 combined with the LysM domains of LjNFR1 in complementation tests, can trigger nodulation signalling (Miyata et al., 2014). In legumes, PsLykX and MtLYK4, which are the results of recent duplications in M. truncatula and pea, respectively (Limpens et al., 2003; Sulima et al., 2017), are "YAQ-less" proteins that might heterodimerise with "YAQ-type" LysM-RLKs. The YAQ motif was also lost in AtCERK1 (De Mita et al., 2014), and this is reminiscent of the loss of NFP proteins and the divergence in LYR3 proteins in Brassicales plants.

In addition to LCOs, plant LysM proteins can recognise peptidoglycan and chitin (Antolin-Llovera et al., 2014), and these or other structurally-related, secreted or surface constituents of microbial symbionts are candidate ligands for symbiotic LysM-RLKs. COs are proposed to be symbiotic AM fungal signals (Genre et al., 2013), and OsCERK1 is involved in the recognition of such molecules, although no binding to chitin or COs has been observed (Shimizu et al., 2010; Shinya et al., 2012). For the RL symbiosis, the YAQ-less "LYK-type" LysM-RLK LjEPR3 was reported to bind rhizobial exopolysaccharide and control nodulation (Kawaharada et al., 2015). Interestingly, although there are candidate LjEPR3 orthologs in many plants,

Parasponia spp. are notable exceptions (van Velzen et al., 2018), indicating that this protein is not indispensable for symbiosis.

SYMBIOTIC LYSM-RLKS CAN ALSO CONTROL IMMUNITY

Relationships between symbiosis and plant immunity are becoming more documented (Rey and Jacquet, 2018), and interestingly, both MtNFP and OsCERK1 have dual roles in symbiosis and immunity (Ben et al., 2013; Rey et al., 2013, 2015; Miyata et al., 2014; Zhang et al., 2015). In the case of OsCERK1 this is likely linked to CO perception (Carotenuto et al., 2017), and in legume plants LjLYS6/MtLYK9/PsLYK9 is a good candidate to have analogous roles to OsCERK1 (De Mita et al., 2014; Bozsoki et al., 2017; Leppyanen et al., 2018; our unpublished data). For MtNFP, one hypothesis is that it evolved to interact with both symbiotic and immune-related receptor proteins (Gough and Jacquet, 2013). Recently, LjLYS12 was reported to intervene in plant immunity (Fuechtbauer et al., 2018), suggesting that this LYR3 protein can also have a dual role in symbiosis and immunity.

More generally, people have asked whether the roles of LysM-RLKs in plant immunity diverged from symbiotic perception mechanisms or vice versa (Liang et al., 2014). Given that certain immune-type responses might be co-opted to facilitate symbiosis establishment (Limpens et al., 2015), then ancestral proteins could have had dual symbiotic and immune roles such that *Brassicales* LysM-RLKs have become subfunctionalised in immunity, while other proteins could have become subfunctionalised for symbiosis or retained the dual role. Indeed, the sequence divergence and biological roles of AtCERK1 and AtLYK4, can be interpreted as examples of subfunctionalisation in immunity.

FUTURE CHALLENGES

Other symbiotic plant-fungal interactions include ectomycorrhiza (ECM), orchid mycorrhiza, ericoid mycorrhiza, and fine root endophytes have recently been described (Orchard et al., 2017). There are also other nitrogen-fixing symbioses, between Frankia bacteria and actinorhizal plants, and between certain legume plants and rhizobia that cannot produce Nod factors. Future studies should determine whether LysM-RLKs play roles in these diverse types of symbiosis. By analogy with the quenching role of the LysM protein Ecp6 of the fungal tomato pathogen *Cladosporium fulvum* (de Jonge et al., 2010) future investigations should also address whether any rhizobial or AM fungal LysM proteins have evolved symbiotic roles (Zeng et al., 2018). Evolutionary and structural studies are also needed to help decipher ligand binding properties of symbiotic LysM-RLKs, and more knowledge is needed on evolutionary events that have led to protein regulation, both quantitatively and at subcellular and tissue-specific levels.

SEQUENCE FILES

Glossary of LysM-RLK proteins discussed in the text:

MtNFP: *Medicago truncatula* Nod Factor Perception, controls all Nod factor responses and nodulation, as well as plant immunity

LjNFR5: the *Lotus japonicus* ortholog of MtNFP

SlLYK10: the *Solanum lycopersicum* ortholog of MtNFP, controls the AM symbiosis

OsNFR5: the *Oryza sativa* ortholog of MtNFP

MtLYR3: *Medicago truncatula* high affinity LCO binding protein with unknown biological function

LjLYS12: the *Lotus japonicus* ortholog of MtLYR3

AtLYK4: the *Arabidopsis thaliana* ortholog of MtLYR3, controls immunity

MtLYK3: *Medicago truncatula* LysM receptor-like kinase3, controls rhizobial infection and nodulation

LjNFR1: the *Lotus japonicus* ortholog of MtLYK3

PslykX: a *Pisum sativum* LysM-RLK potentially involved in nodulation specificity for Nod factor structure

AtCERK1: *Arabidopsis thaliana* Chitin Elicitor Receptor Kinase1, controls plant immunity

OsCERK1: the *Oryza sativa* ortholog of AtCERK1, controls the AM symbiosis and plant immunity

AUTHOR CONTRIBUTIONS

CG and BL provided the sequences. CG, LC, BL, and J-JB analysed the sequences and wrote the text. LC constructed the phylogeny trees. CG and J-JB constructed the weblogos.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2018.00923/full#supplementary-material

FIGURE S1 | Sequence logos generated using WebLogo software (Schneider and Stephens, 1990; Crooks et al., 2004), and approximately 60 Angiosperm protein sequences of NFP for each individual LysM domain of NFP proteins.

FIGURE S2 | Sequence logos generated using WebLogo software (Schneider and Stephens, 1990; Crooks et al., 2004), and approximately 60 Angiosperm protein sequences of LYR3 for each individual LysM domain of LYR3 proteins.

TABLE S1 | Lists of NFP and LYR3 sequences used in **Figures 1**, **2** and **Supplementary Figures S1**, **S2**.

DATA SHEET S1 | Protein sequences, whole protein, LysM and kinase domain alignments, and tree files.

REFERENCES

Anisimova, M., and Gascuel, O. (2006). Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst. Biol.* 55, 539–552. doi: 10.1080/10635150600755453

Antolin-Llovera, M., Petutsching, E. K., Ried, M. K., Lipka, V., Nurnberger, T., Robatzek, S., et al. (2014). Knowing your friends and foes - plant receptor-like kinases as initiators of symbiosis or defence. *New Phytol.* 204, 791–802. doi: 10.1111/nph.13117

Aoki, S., Uehara, K., Imafuku, M., Hasebe, M., and Ito, M. (2004). Phylogeny and divergence of basal angiosperms inferred from *APETALA3*- and *PISTILLATA*-like MADS-box genes. *J. Plant Res.* 117, 229–244. doi: 10.1007/s10265-004-0153-7

Arrighi, J. F., Barre, A., Ben Amor, B., Bersoult, A., Soriano, L. C., Mirabella, R., et al. (2006). The *Medicago truncatula* lysin motif-receptor-like kinase gene family includes NFP and new nodule-expressed genes. *Plant Physiol.* 142, 265–279. doi: 10.1104/pp.106.084657

Amor, B. B., Shaw, S. L., Oldroydm, G. E. D., Maillet, F., Penmetsa, R. V., Cook, D., et al. (2003). The NFP locus of *Medicago truncatula* controls an early step of Nod factor signal transduction upstream of a rapid calcium flux and root hair deformation. *Plant J.* 34, 495–506. doi: 10.1046/j.1365-313X.2003.01743.x

Ben, C., Toueni, M., Montanari, S., Tardin, M. C., Fervel, M., Negahi, A., et al. (2013). Natural diversity in the model legume *Medicago truncatula* allows identifying distinct genetic mechanisms conferring partial resistance to *Verticillium* wilt. *J. Exp. Bot.* 64, 317–332. doi: 10.1093/jxb/ers337

Bensmihen, S., de Billy, F., and Gough, C. (2011). Contribution of NFP LysM domains to the recognition of Nod factors during the *Medicago truncatula/Sinorhizobium meliloti* symbiosis. *PLoS One* 6:e26114. doi: 10.1371/journal.pone.0026114

Bono, J., Fliegmann, J., Gough, C., and Cullimore, J. (2018). "Expression and function of the *Medicago truncatula* lysin motif receptor-like kinase (LysM-RLK) gene family in the legume-rhizobia symbiosis," in *The Model Legume Medicago truncatula*, ed F. de Bruijn (Hoboken, NJ: John Wiley & Sons Inc.)

Bozsoki, Z., Cheng, J., Feng, F., Gysel, K., Vinther, M., Andersen, K. R., et al. (2017). Receptor-mediated chitin perception in legume roots is functionally separable from Nod factor perception. *Proc. Natl. Acad. Sci. U.S.A.* 114, E8118–E8127. doi: 10.1073/pnas.1706795114

Buendia, L., Wang, T., Girardin, A., and Lefebvre, B. (2016). The LysM receptor-like kinase SlLYK10 regulates the arbuscular mycorrhizal symbiosis in tomato. *New Phytol.* 210, 184–195. doi: 10.1111/nph.13753

Cannon, S. B., McKain, M. R., Harkess, A., Nelson, M. N., Dash, S., Deyholos, M. K., et al. (2015). Multiple polyploidy events in the early radiation of nodulating and nonnodulating legumes. *Mol. Biol. Evol.* 32, 193–210. doi: 10.1093/molbev/msu296

Carotenuto, G., Chabaud, M., Miyata, K., Capozzi, M., Takeda, N., Kaku, H., et al. (2017). The rice LysM receptor-like kinase *OsCERK1* is required for the perception of short-chain chitin oligomers in arbuscular mycorrhizal signaling. *New Phytol.* 214, 1440–1446. doi: 10.1111/nph.14539

Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190. doi: 10.1101/gr.849004

de Jonge, R., van Esse, H. P., Kombrink, A., Shinya, T., Desaki, Y., Bours, R., et al. (2010). Conserved fungal LysM effector Ecp6 prevents chitin-triggered immunity in plants. *Science* 329, 953–955. doi: 10.1126/science.1190859

De Mita, S., Ronfort, J., McKhann, H. I., Poncet, C., El Malki, R., and Bataillon, T. (2007). Investigation of the demographic and selective forces shaping the nucleotide diversity of genes involved in nod factor signaling in *Medicago truncatula*. *Genetics* 177, 2123–2133. doi: 10.1534/genetics.107.076943

De Mita, S., Streng, A., Bisseling, T., and Geurts, R. (2014). Evolution of a symbiotic receptor through gene duplications in the legume-rhizobium mutualism. *New Phytol.* 201, 961–972. doi: 10.1111/nph.12549

Delaux, P. M., Varala, K., Edger, P. P., Coruzzi, G. M., Pires, J. C., and Ane, J. M. (2014). Comparative phylogenomics uncovers the impact of symbiotic associations on host genome evolution. *PLoS Genet.* 10:e1004487. doi: 10.1371/journal.pgen.1004487

Delaux, P. M., Radhakrishnan, G. V., Jayaraman, D., Cheem, J., Malbreil, M., Volkening, J. D., et al. (2015). Algal ancestor of land plants was preadapted for symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* 112, 13390–13395. doi: 10.1073/pnas.1515426112

Fliegmann, J., Canova, S., Lachaud, C., Uhlenbroich, S., Gasciolli, V., Pichereaux, C., et al. (2013). Lipo-chitooligosaccharidic symbiotic signals are recognized by LysM receptor-like kinase LYR3 in the legume *Medicago truncatula*. *ACS Chem. Biol.* 8, 1900–1906. doi: 10.1021/cb400369u

Fliegmann, J., Jauneau, A., Pichereaux, C., Rosenberg, C., Gasciolli, V., Timmers, A. C., et al. (2016). LYR3, a high-affinity LCO-binding protein of *Medicago truncatula*, interacts with LYK3, a key symbiotic receptor. *FEBS Lett.* 590, 1477–1487. doi: 10.1002/1873-3468.12191

Fuechtbauer, W., Yunusov, T., Bozsoki, Z., Gavrin, A., James, E. K., Stougaard, J., et al. (2018). LYS12 LysM receptor decelerates *Phytophthora palmivora* disease progression in *Lotus japonicus*. *Plant J.* 93, 297–310. doi: 10.1111/tpj.13785

Genre, A., Chabaud, M., Balzergue, C., Puech-Pages, V., Novero, M., Rey, T., et al. (2013). Short-chain chitin oligomers from arbuscular mycorrhizal fungi trigger nuclear Ca2$^+$ spiking in *Medicago truncatula* roots and their production is enhanced by strigolactone. *New Phytol.* 198, 179–189. doi: 10.1111/nph.12146

Geurts, R., Xiao, T. T., and Reinhold-Hurek, B. (2016). What does it take to evolve a nitrogen-fixing endosymbiosis? *Trends Plant Sci.* 21, 199–208. doi: 10.1016/j.tplants.2016.01.012

Gough, C., and Cullimore, J. (2011). Lipo-chitooligosaccharide signaling in endosymbiotic plant-Microbe interactions. *Mol. Plant Microbe Interact.* 24, 867–878. doi: 10.1094/MPMI-01-11-0019

Gough, C., and Jacquet, C. (2013). Nod factor perception protein carries weight in biotic interactions. *Trends Plant Sci.* 18, 566–574. doi: 10.1016/j.tplants.2013.06.001

Harikrishnan, S. L., Pucholt, P., and Berlin, S. (2015). Sequence and gene expression evolution of paralogous genes in willows. *Sci. Rep.* 5:18662. doi: 10.1038/srep18662

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436

Kawaharada, Y., Kelly, S., Nielsen, M. W., Hjuler, C. T., Gysel, K., Muszynski, A., et al. (2015). Receptor-mediated exopolysaccharide perception controls bacterial infection. *Nature* 523, 308–312. doi: 10.1038/nature14611

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404

Lefebvre, B., Klaus-Heisen, D., Pietraszewska-Bogiel, A., Herve, C., Camut, S., Auriac, M. C., et al. (2012). Role of *N*-Glycosylation sites and *CXC* motifs in trafficking of *Medicago truncatula* nod factor perception protein to plasma membrane. *J. Biol. Chem.* 287, 10812–10823. doi: 10.1074/jbc.M111.281634

Leppyanen, I. V., Shakhnazarova, V. Y., Shtark, O. Y., Vishnevskaya, N. A., Tikhonovich, I. A., and Dolgikh, E. A. (2018). Receptor-like kinase LYK9 in *Pisum sativum*, L. Is the CERK1-like receptor that controls both plant immunity and AM symbiosis development. *Int. J. Mol. Sci.* 19:E8. doi: 10.3390/ijms19010008

Letunic, I., and Bork, P. (2007). Interactive tree of life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics.*23, 127–128. doi: 10.1093/bioinformatics/btl529

Liang, Y., Toth, K., Cao, Y. R., Tanaka, K., Espinoza, C., and Stacey, G. (2014). Lipochitooligosaccharide recognition: an ancient story. *New Phytol.* 204, 289–296. doi: 10.1111/nph.12898

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. doi: 10.1126/science.1090074

Limpens, E., van Zeijl, A., and Geurts, R. (2015). Lipochitooligosaccharides modulate plant host immunity to enable endosymbioses. *Annu. Rev. Phytopathol.* 53, 311–334. doi: 10.1146/annurev-phyto-080614-120149

Lohmann, G. V., Shimoda, Y., Nielsen, M. W., Jorgensen, F. G., Grossmann, C., Sandal, N., et al. (2010). Evolution and regulation of the lotus japonicus LysM

receptor gene family. *Mol. Plant Microbe Interact.* 23, 510–521. doi: 10.1094/MPMI-23-4-0510

Madsen, E. B., Antolin-Llovera, M., Grossmann, C., Ye, J. Y., Vieweg, S., Broghammer, A., et al. (2011). Autophosphorylation is essential for the *in vivo* function of the *Lotus japonicus* Nod factor receptor 1 and receptor-mediated signalling in cooperation with Nod factor receptor 5. *Plant J.* 65, 404–417. doi: 10.1111/j.1365-313X.2010.04431.x

Madsen, E. B., Madsen, L. H., Radutoiu, S., Olbryt, M., Rakwalska, M., Szczyglowski, K., et al. (2003). A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals. *Nature* 425, 637–640. doi: 10.1038/nature02045

Maillet, F., Poinsot, V., Andre, O., Puech-Pages, V., Haouy, A., Gueunier, M., et al. (2011). Fungal lipochitooligosaccharide symbiotic signals in arbuscular mycorrhiza. *Nature* 469, 58–63. doi: 10.1038/nature09622

Malkov, N., Fliegmann, J., Rosenberg, C., Gasciolli, V., Timmers, A. C. J., Nurisso, A., et al. (2016). Molecular basis of lipo-chitooligosaccharide recognition by the lysin motif receptor-like kinase LYR3 in legumes. *Biochem. J.* 473, 1369–1378. doi: 10.1042/BCJ20160073

Miyata, K., Hayafune, M., Kobae, Y., Kaku, H., Nishizawa, Y., Masuda, Y., et al. (2016). Evaluation of the role of the LysM receptor-like kinase, OsNFR5/OsRLK2 for AM symbiosis in rice. *Plant Cell Physiol.* 57, 2283–2290. doi: 10.1093/pcp/pcw144

Miyata, K., Kozaki, T., Kouzai, Y., Ozawa, K., Ishii, K., Asamizu, E., et al. (2014). The bifunctional plant receptor, OsCERK1, regulates both chitin-triggered immunity and arbuscular mycorrhizal symbiosis in rice. *Plant Cell Physiol.* 55, 1864–1872. doi: 10.1093/pcp/pcu129

Moling, S., Pietraszewska-Bogiel, A., Postma, M., Fedorova, E., Hink, M. A., Limpens, E., et al. (2014). Nod factor receptors form heteromeric complexes and are essential for intracellular infection in *Medicago* nodules. *Plant Cell* 26, 4188–4199. doi: 10.1105/tpc.114.129502

Nakagawa, T., Kaku, H., Shimoda, Y., Sugiyama, A., Shimamura, M., Takanashi, K., et al. (2011). From defense to symbiosis: limited alterations in the kinase domain of LysM receptor-like kinases are crucial for evolution of legume-*Rhizobium* symbiosis. *Plant J.* 65, 169–180. doi: 10.1111/j.1365-313X.2010.04411.x

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Op den Camp, R., Streng, A., De Mita, S., Cao, Q., Polone, E., Liu, W., et al. (2010). LysM-type mycorrhizal receptor recruited for rhizobium symbiosis in nonlegume *Parasponia*. *Science* 331, 909–912. doi: 10.1126/science.1198181

Orchard, S., Hilton, S., Bending, G. D., Dickie, I. A., Standish, R. J., Gleeson, D. B., et al. (2017). Fine endophytes (*Glomus tenue*) are related to Mucoromycotina, not Glomeromycota. *New Phytol.* 213, 481–486. doi: 10.1111/nph.14268

Pietraszewska-Bogiel, A., Lefebvre, B., Koini, M. A., Klaus-Heisen, D., Takken, F. L. W., Geurts, R., et al. (2013). Interaction of *Medicago truncatula* lysin motif receptor-like kinases, NFP and LYK3, produced in *Nicotiana benthamiana* induces defence- like responses. *PLoS One* 8:e65055. doi: 10.1371/journal.pone.0065055

Radutoiu, S., Madsen, L. H., Madsen, E. B., Jurkiewicz, A., Fukai, E., Quistgaard, E. M., et al. (2007). LysM domains mediate lipochitin-oligosaccharide recognition and *Nfr* genes extend the symbiotic host range. *EMBO J.* 26, 3923–3935. doi: 10.1038/sj.emboj.7601826

Rey, T., Chatterjee, A., Buttay, M., Toulotte, J., and Schornack, S. (2015). *Medicago truncatula* symbiosis mutants affected in the interaction with a biotrophic root pathogen. *New Phytol.* 206, 497–500. doi: 10.1111/nph.13233

Rey, T., and Jacquet, C. (2018). Symbiosis genes for immunity and vice versa. *Curr. Opin. Plant Biol.* 44, 64–71. doi: 10.1016/j.pbi.2018.02.010

Rey, T., Nars, A., Bonhomme, M., Bottin, A., Huguet, S., Balzergue, S., et al. (2013). NFP, a LysM protein controlling Nod factor perception, also intervenes in *Medicago truncatula* resistance to pathogens. *New Phytol.* 198, 875–886. doi: 10.1111/nph.12198

Schneider, T. D., and Stephens, R. M. (1990). Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* 18, 6097–6100. doi: 10.1093/nar/18.20.6097

Shimizu, T., Nakano, T., Takamizawa, D., Desaki, Y., Ishii-Minami, N., Nishizawa, Y., et al. (2010). Two LysM receptor molecules, CEBiP and OsCERK1, cooperatively regulate chitin elicitor signaling in rice. *Plant J.* 64, 204–214. doi: 10.1111/j.1365-313X.2010.04324.x

Shinya, T., Motoyama, N., Ikeda, A., Wada, M., Kamiya, K., Hayafune, M., et al. (2012). Functional characterization of CEBiP and CERK1 homologs in *Arabidopsis* and rice reveals the presence of different chitin receptor systems in plants. *Plant Cell Physiol.* 53, 1696–1706. doi: 10.1093/pcp/pcs113

Shiu, S. H., and Bleecker, A. B. (2001). Receptor-like kinases from *Arabidopsis* form a monophyletic gene family related to animal receptor kinases. *Proc. Natl Acad. Sci. U.S.A.* 98, 10763–10768. doi: 10.1073/pnas.181141598

Smit, P., Limpens, E., Geurts, R., Fedorova, E., Dolgikh, E., Gough, C., et al. (2007). Medicago LYK3, an entry receptor in rhizobial nodulation factor signaling. *Plant Physiol.* 145, 183–191. doi: 10.1104/pp.107.100495

Sorensen, K. K., Simonsen, J. B., Maolanon, N. N., Stougaard, J., and Jensen, K. J. (2014). chemically synthesized 58-mer lysm domain binds lipochitin oligosaccharide. *Chembiochem* 15, 2097–2105. doi: 10.1002/cbic.201402125

Sulima, A. S., Zhukov, V. A., Afonin, A. A., Zhernakov, A. I., Tikhonovich, I. A., and Lutova, L. A. (2017). Selection signatures in the first exon of paralogous receptor kinase genes from the *sym2* region of the *Pisum sativum*, L. Genome. *Front. Plant Sci.* 8:1957. doi: 10.3389/fpls.2017.01957

Tanaka, K., Nguyen, C., Liang, Y., Cao, Y., and Stacey, G. (2013). Role of LysM receptors in chitin-triggered plant innate immunity. *Plant Signal. Behav.* 8:e22598. doi: 10.4161/psb.22598

van Velzen, R., Holmer, R., Bu, F., Rutten, L., van Zeijl, A., Liu, W., et al. (2018). Comparative genomics of the nonlegume *Parasponia* reveals insights into evolution of nitrogen-fixing rhizobium symbioses. *Proc. Natl. Acad. Sci. U.S.A.* 115, E4700–E4709. doi: 10.1073/pnas.1721395115

Wan, J. R., Tanaka, K., Zhang, X. C., Son, G. H., Brechenmacher, L., Nguyen, T. H., et al. (2012). LYK4, a lysin motif receptor-like kinase, is important for chitin signaling and plant innate immunity in *Arabidopsis*. *Plant Physiol.* 160, 396–406. doi: 10.1104/pp.112.201699

Zeng, T., Holmer, R., Hontelez, J., Te Lintel-Hekkert, B., Marufu, L., de Zeeuw, T., et al. (2018). Host- and stage-dependent secretome of the arbuscular mycorrhizal fungus *Rhizophagus irregularis*. *Plant J.* 94, 411–425. doi: 10.1111/tpj.13908

Zhang, X. W., Dong, W. T., Sun, J. H., Feng, F., Deng, Y. W., He, Z. H., et al. (2015). The receptor kinase *CERK1* has dual functions in symbiosis and immunity signalling. *Plant J.* 81, 258–267. doi: 10.1111/tpj.12723

The Art of Self-Control – Autoregulation of Plant–Microbe Symbioses

Chenglei Wang, James B. Reid and Eloise Foo*

School of Natural Sciences, University of Tasmania, Hobart, TAS, Australia

Plants interact with diverse microbes including those that result in nutrient-acquiring symbioses. In order to balance the energy cost with the benefit gained, plants employ a systemic negative feedback loop to control the formation of these symbioses. This is particularly well-understood in nodulation, the symbiosis between legumes and nitrogen-fixing rhizobia, and is known as autoregulation of nodulation (AON). However, much less is understood about the autoregulation of the ancient arbuscular mycorrhizal symbioses that form between Glomeromycota fungi and the majority of land plants. Elegant physiological studies in legumes have indicated there is at least some overlap in the genes and signals that regulate these two symbioses but there are major gaps in our understanding. In this paper we examine the hypothesis that the autoregulation of mycorrhizae (AOM) pathway shares some elements with AON but that there are also some important differences. By reviewing the current knowledge of the AON pathway, we have identified important directions for future AOM studies. We also provide the first genetic evidence that *CLV2* (an important element of the AON pathway) influences mycorrhizal development in a non-legume, tomato and review the interaction of the autoregulation pathway with plant hormones and nutrient status. Finally, we discuss whether autoregulation may play a role in the relationships plants form with other microbes.

Keywords: autoregulation, nodulation, arbuscular mycorrhizae, CLAVATA, CLE peptide, tomato

INTRODUCTION

Mycorrhizal symbioses between plants and fungi are widespread and ancient, with evidence from fossils and extant basal plants indicating that such interactions evolved during the colonization of land by plants between 450 and 475 mya (Field et al., 2015; Martin et al., 2017). The mycorrhizal fungi invade and extend the host plant's root system, enabling enhanced nutrient uptake in exchange for fixed carbon. The most widespread are the arbuscular mycorrhizal (AM) associations that are characterized by the presence of a unique nutrient exchange unit, the fungal arbuscule. A significant amount of plant-derived carbon is invested, with estimates of between 4 and 20% of the carbon fixed by the plant transferred to the AM fungi (Bago et al., 2000; Douds et al., 2000). To prevent the energy cost to the plant outweighing the benefits of the interaction, plants might be expected to regulate AM colonization. Indeed, elegant physiological studies in flowering

plants have revealed powerful systemic regulation of AM colonization. In split root studies, pre-colonization of one side of the root system can suppress subsequent colonization of the other side of the root system, providing evidence for a root–shoot feedback system termed autoregulation of mycorrhiza (AOM) (Vierheilig et al., 2000a, 2008; Meixner et al., 2005). Although there is no direct evidence that such control mechanisms also occur in more basal plant lineages, it is fair to assume that autoregulation would be an important element in the evolution of mutually beneficial plant–mycorrhizal interactions to prevent a potentially parasitic relationship developing.

Significant progress has been made in the past few decades in our understanding of the plant genes and signals that regulate AM symbioses, assisted greatly by the cross-over with the more recently evolved nodulation symbioses that occur between nitrogen-fixing rhizobial bacteria and (almost exclusively) legumes. In particular, the identification of a common symbiotic signaling pathway, required for the formation of both nodules and AM, has revealed that elements of the ancient AM pathway were co-opted into the more recently evolved nodulation pathway (Delaux et al., 2013; Martin et al., 2017). In particular, the common symbiotic pathway includes genes essential for initial communication between the plant host and rhizobia or AM fungi. However, a missing element in these comparisons has been the autoregulation of nodulation (AON) pathway. Like AM, nodulation is under powerful systemic control and the identity of AON signals and transduction pathways are now becoming clear (for review see Reid et al., 2011b; Soyano and Kawaguchi, 2014). Studies in legumes suggest some key cross-overs in the autoregulation pathways, as plant mutants disrupted in the AON pathway display not only hypernodulation but also hypermycorrhizal colonization (Morandi et al., 2000; Shrihari et al., 2000) and split root studies indicate nodulation can systemically suppress AM and vice versa (Catford et al., 2003). In this article, we examine our current understanding of AOM and begin to extend this beyond legumes by providing evidence of a role for the key AON gene in legumes, *CLAVATA2 (CLV2)*, in the AOM pathway of tomato. We also consider the role of plant hormones in autoregulation and examine the potential for the autoregulation of other beneficial plant–microbe interactions.

GENES AND SIGNALS OF THE AUTOREGULATION PATHWAY(S)

The systemic regulation of AM colonization in split root studies has been observed in both legumes and non-legumes (Vierheilig et al., 2000a; Meixner et al., 2007). This negative regulation appears to require a certain threshold of the amount of root colonized by AM (Vierheilig, 2004). AOM does not appear to be due simply to the strength of the carbon sink (Vierheilig et al., 2000b, 2008), but is rather regulated by a more specific pathway. Nodulation can systemically suppress AM and vice versa and even the application of Nod factors, the rhizobial-derived signaling molecules, can suppress AM (Catford et al., 2003), suggesting a clear overlap between the AON and AOM

pathways. Similar conclusions were drawn from studies using a dual inoculation system in which both rhizobia and AM fungi were applied to the one root system (Sakamoto et al., 2013). This has also been explored in the non-legume barley, where it was found that inoculation with *Rhizobium* sp. could systemically inhibit AM formation. However, this inhibitory effect did not rely on Nod-factor production but was linked instead to the type 3 effector proteins of this rhizobial strain (Khaosaad et al., 2010).

A member of the Leucine-rich repeat (LRR) receptor kinase family that shares similarities with the CLAVATA1 (CLV1) protein (outlined below) has a central role in the AON in legumes (**Figure 1**). This *CLV1*-like gene is known as *NODULE AUTOREGULATION RECEPTOR KINASE (NARK)* in soybean, *HYPERNODULATION ABERRANT ROOT FORMATION 1 (HAR1)* in *Lotus japonicus*, *SUPER NUMERIC NODULES (SUNN)* in *Medicago truncatula* and *SYM29* in pea, and mutations in these genes result in hypernodulation (reviewed by Reid et al., 2011b). The CLV1-like protein appears to function as a shoot receptor for root-derived CLAVATA/ESR-related (CLE) peptides. In the root, events associated with nodulation generate specific rhizobial induced CLE peptides (Rh-CLE **Figure 1**) which in some cases appear to be arabinosylated via action of the enzyme ROOT DETERMINED NODULATION1 (RDN1) (e.g., *Mt*CLE12, *Gm*RIC1a, and *Gm*RIC1b) (Schnabel et al., 2011; Okamoto et al., 2013; Kassaw et al., 2017; Hastwell et al., 2018; Imin et al., 2018). The CLE peptides are then translocated to the shoot where they are perceived by a receptor complex that includes the CLV1-like protein (e.g., Okamoto et al., 2013). The perception of the root-derived signal(s) in the plant shoot generates a shoot to root signal that inhibits further nodulation. The shoot to root inhibitor is predicted to be a small molecular weight heat-stable molecule (Lin et al., 2010) that has been suggested to be cytokinin (Sasaki et al., 2014). Other elements of the AON pathway include the shoot acting CLV2 (Krusell et al., 2011), CORYNE (CRN) (Crook et al., 2016), and KLAVIER (KLV) (Miyazawa et al., 2010), all three of which encode LRR receptors that may also play a role in CLE perception, and TOO MUCH LOVE (TML), a root-acting F-Box protein that appears to act downstream of the shoot to root signal (Takahara et al., 2013).

In addition to its role in nodulation, the CLV1-like protein is also essential in the AOM pathway since *clv1-like* mutants across legume species (*sym29, sunn, nark,* and *har1*) also display hypermycorrhizal colonization (Morandi et al., 2000; Solaiman et al., 2000; Sakamoto and Nohara, 2009). It is important to note that the relatively small increase in AM in these mutant plants compared to wild type (WT) (between 20 and 50%) contrasts with the large (many fold) increase in nodule numbers of these lines compared with WT lines. Some have speculated that this relatively small increase may be due to the ability of the AM fungus to spread laterally in the root, meaning it is sometimes difficult to distinguish the effect of *clv1-like* mutations on subsequent AM infection (Schaarschmidt et al., 2013). One interpretation is that the AOM pathway may inhibit new AM infections at the epidermis but may not limit the spread of AM between cortical cells, although this has not been tested

FIGURE 1 | Proteins and signals that act in the shoot and/or root to autoregulate nodulation and mycorrhizal symbioses. Flat-ended lines indicate a negative influence, while arrows indicate a positive influence. Question marks and dotted lines indicate untested elements. CK, cytokinin; N, nitrogen; P, phosphorous; Rh, rhizobial; Myc, mycorrhizal; arb, arabinosylated.
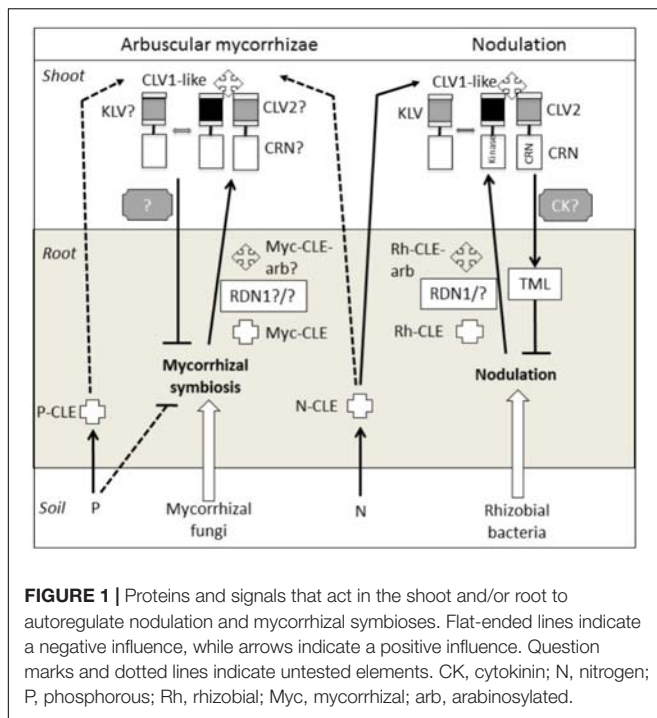
empirically. Such a hypothesis is consistent with the activation of AON during early stages of rhizobial infection (Li et al., 2009). Another way to examine the role of *CLV1-like* genes in AOM has been through the use of split root studies with mutant plants, which have shown that the *CLV1-like* gene in soybean (*NARK*) influences the suppression of subsequent mycorrhizal colonization (Meixner et al., 2005, 2007; Schaarschmidt et al., 2013). A role for NARK in the shoot control of AM was also observed in reciprocal grafts between the soybean *nark* mutant En6500 and WT plants (Sakamoto and Nohara, 2009) but not in a split root system with the same En6500 mutant (Meixner et al., 2007). Studies of orthologous mutants in other legumes species (e.g., *har1, sym29*) may be useful to clarify this inconsistency.

Apart from the requirement for the CLV1-like protein in AOM it is not yet clear if other AON genes encoding proteins that act in the root (RDN1, TML), shoot (CLV2, KLV, CRN) or as mobile signals (CLE) are also employed by the AOM pathway (**Figure 1**). One study examining the effect of a pea *clv2* mutant (*sym28)* found only a small but not significant increase in AM colonization compared with WT, although grafting and split root studies that would reveal if CLV2 plays a role in the systemic regulation of AM were not attempted (Morandi et al., 2000). This and other mutants disrupted in root and shoot acting elements of the AON pathway outlined above are available in a range of legumes, but to date their AM phenotypes and possible role in systemic AOM regulation has not been examined. Recent work in the legume *Lotus japonicas* indicated that mycorrhizal colonization and phosphate starvation generates CLE peptides distinct from those induced by nodulation or nitrogen (Funayama-Noguchi et al., 2011; Handa

et al., 2015), but the function of these CLE peptides has yet to be tested. In an attempt to understand downstream elements of the AOM pathway, Schaarschmidt et al. (2013) analyzed the soybean transcriptome in split root mycorrhizal studies using WT and *nark* lines. Two putative CCAAT-binding transcription factor genes were identified, *GmNF-YA1a* and *GmNF-YA1b,* that are down-regulated by AM in a NARK-dependent manner. Hairy root RNAi lines with reduced expression of *GmNF-YA1a/b* displayed reduced AM colonization and this occurred in both WT and *nark* backgrounds, consistent with *GmNF-YA1a/b* acting downstream of *NARK* to suppress AM.

The evolutionary origin of the AON/AOM genes is still emerging but is informed by the *CLAVATA-WUSCHEL* (*CLV-WUS*) shoot meristem pathway. This pathway is best understood in Arabidopsis and involves several LRR receptors that act locally in the shoot, including CLV1, CLV2 and CRN, to perceive a CLE peptide, CLV3, which in turn activates a feedback loop to maintain a defined stem cell population in the shoot apical meristem (Hazak and Hardtke, 2016). Arabidopsis lines disrupted in these genes displayed altered shoot meristem formation. In pea and *Lotus japonicas*, mutant studies have revealed CLV2 plays a dual role, acting in both shoot development and AON, as *clv2* mutants display hypernodulation and shoot fasciation (Krusell et al., 2011). In contrast, there appear to be specific CLE genes that act in the AON pathway (Mortier et al., 2010). Similarly, disruption of *CLV1-like* genes closely related to *CLV1* in pea, *Lotus, Medicago* and soybean (*SYM29, HAR1, SUNN,* and *NARK,* respectively) result in hypernodulation but not shoot fasciation (reviewed by Reid et al., 2011b). A possible explanation in soybean for divergence between *NARK* and *CLV1*, is that *CLV1* appears to have undergone a duplication, resulting in *NARK* and *CLV1A* (Yamamoto et al., 2000) but this does not appear to be the case in the other legumes examined (see Schnabel et al., 2005). *CLV1A* is more closely related to *AtCLV1* and recent studies have shown it influences shoot architecture but not nodulation (Mirzaei et al., 2017). Although many phylogenetic studies have examined the *CLV1, CLV2* and *CLE* gene families, these studies have been almost exclusively limited to angiosperms (e.g., Sun and Wang, 2011; Zan et al., 2013; Wei et al., 2015; Xu et al., 2015; Hastwell et al., 2017), preventing a more comprehensive understanding of the evolutionary history and possible functional diversification of these genes. One recent transcriptomic study failed to find evidence for the *CLAVATA-WUSCHEL* (*CLV*-WUS) pathway in the moss *Physcomitrella* and liverwort *Marchantia* (Frank and Scanlon, 2015). However, a more comprehensive examination of these gene families in basal plant lineages and in mycorrhizal vs. non-mycorrhizal species (see approach of Favre et al., 2014; Delaux et al., 2015) might provide an insight into their possible role in the AM program.

ROLE OF CLV2 IN AOM OF THE NON-LEGUME TOMATO

To fully understand the genes, signals and evolutionary history of AOM, we must go beyond legumes. Indeed, a non-legume system removes the possible complication of cross-talk between

the AON and AOM pathways to allow us to identify AOM components. Therefore, we employed the tomato *clv2-2* line, a CRISPR-Cas9 knock out line that targets the *CLV2* gene (Xu et al., 2015). As outlined above, this gene is essential for AON in legumes and also acts to control shoot apical meristem formation in legumes (Krusell et al., 2011) and non-legumes, including *clv2* tomato lines that display a weakly fasciated shoot and an increase in the number of floral organs (Xu et al., 2015).

We tested if the *CLV2* gene plays a role in AM development by examining the AM phenotype of the tomato *clv2-2* line. Compared with WT, *clv2* plants displayed a significant increase in AM colonization, including arbuscule frequency, compared with WT plants (**Figure 2A**). Although more frequent, the mycorrhizal structures observed in *clv2* mutants including arbuscules were similar in appearance to WT (**Figure 2B**). Therefore, we provide the first evidence that CLV2 in tomato, known to be important for AON in legumes, also acts as a negative regulator of AM. This provides the first genetic evidence for the AOM pathway in non-legumes.

ROLE OF PLANT HORMONES IN AOM

Autoregulation of mycorrhizae and AON are regulated by systemic signals and in addition to mobile CLE peptides, a range of studies have examined the role of plant hormones in AON and in some cases AOM. Double mutant studies in pea indicate gibberellins, brassinosteroids, and strigolactones are not required for the supernodulation phenotype and thus do not appear to act downstream of AON elements, CLV1-like, CLV2 or RDN1 (Ferguson et al., 2011; Foo et al., 2014b). In contrast, transcriptional studies showed that either jasmonic acid (JA) biosynthesis genes or JA regulated genes were systemically regulated by rhizobial colonization, and this was mediated by GmNARK in soybean. These results suggest the AON pathway influences JA signaling (Kinkema and Gresshoff, 2008). Several studies have also examined a role for auxin in autoregulation. In split root studies in soybean, significant auxin accumulation was observed in AM colonized roots but not uncolonized roots. However, this increase was not as great in the roots of a *nark* mutant (Meixner et al., 2005). This contrasts with nodulation studies with the orthologous mutant in *Medicago, sunn*, that suggest SUNN may be required to suppress auxin accumulation in the root following rhizobial challenge (van Noorden et al., 2006). In WT *Medicago*, challenge with rhizobia leads to the downregulation of auxin transport from the shoot to the root. In contrast, *sunn* mutants displayed elevated auxin levels in the infection zone of the root following inoculation with rhizobia. These studies suggest the AON pathway may modulate shoot to root auxin transport but this has yet to be investigated directly in AOM.

Cytokinin is an interesting case as it has been suggested to be a candidate for the shoot-derived inhibitor in AON, based on several lines of evidence from plants with altered cytokinin or CLE peptide biosynthesis, and measurement of cytokinin levels and transport (Sasaki et al., 2014). However, a key finding of this paper, that the LORE *ipt3-2* mutant allele, which is disrupted

in a key cytokinin biosynthesis gene, has increased nodulation could not be repeated in an independent study (Reid et al., 2017). In addition, another study indicated cytokinin may promote nodulation via the AON pathway. In *Medicago,* application of cytokinin directly to roots could induce the expression of the *MtCLE13* gene (Mortier et al., 2012) believed to encode the root to shoot AON signal. Given the ability for nodulation to suppress AM and vice versa outlined above, it is likely that the shoot-derived inhibitor is a common signal between the AON and AOM pathways. Studies with grafts between legume species certainly suggest that the shoot-derived inhibitor of nodulation is conserved across species (e.g., Lohar and VandenBosch, 2005; Ferguson et al., 2014; Foo et al., 2014a). However, unlike its clear role in nodulation there is less evidence to suggest cytokinin has an influence on AM. For example, the cytokinin receptor mutant in *M. truncatula*, *cre1,* did not display any alteration in AM development (Laffont et al., 2015). However, pharmacological studies suggest that cytokinin may promote AM development in pea (F. Guinel, personal communication). Clearly, questions remain around the role of CK in AON, in particular as the shoot-derived inhibitor, and studies directly testing its endogenous role in AOM are required.

DO NUTRIENT STATUS AND OTHER BENEFICIAL PLANT–MICROBE SYMBIOSES INTERACT WITH THE AUTOREGULATION PATHWAY(S)?

Forming symbioses with rhizobial or AM partners may only be beneficial to the plant under conditions of low mineral nutrient availability. In particular, legumes severely reduce nodulation when roots are exposed to elevated nitrogen levels and there are important roles for elements of the AON pathway in this nitrate-response (see Reid et al., 2011b). For example, the *clv1-like* mutants across legumes display a reduced ability to suppress nodulation in response to nitrate (Schnabel et al., 2005; Lim et al., 2011; Okamoto and Kawaguchi, 2015). This reduced response to nitrogen is also seen in *klv* and *rdn1* mutants (e.g., Jacobsen and Feenstra, 1984; Oka-Kira et al., 2005). However, this has not been comprehensively examined for all AON mutants across species. In addition, nitrate treatment induces the expression of specific CLE peptides, which in some cases are the same as those that are induced by rhizobia (Okamoto et al., 2009; Reid et al., 2011a). Whether the AOM pathway plays any role in nitrogen-regulation of AM symbioses has not been explored and, unlike the clearly suppressive effects on nodulation, it is not even clear if nitrogen is a promoter or inhibitor of AM (Correa et al., 2015). In contrast, phosphate has a strong suppressive influence on AM and this influence is systemic and does not require strigolactones (Breuillin et al., 2010; Foo et al., 2013b). Although phosphate induces expression of specific CLE peptides (Funayama-Noguchi et al., 2011), there is no direct evidence that the AOM pathway mediates the phosphate response of AM. Indeed phosphate regulation of AM is maintained in soybean and pea mutants disrupted in the *clv1-like* mutants, *nark* and *sym29* (Wyss et al.,
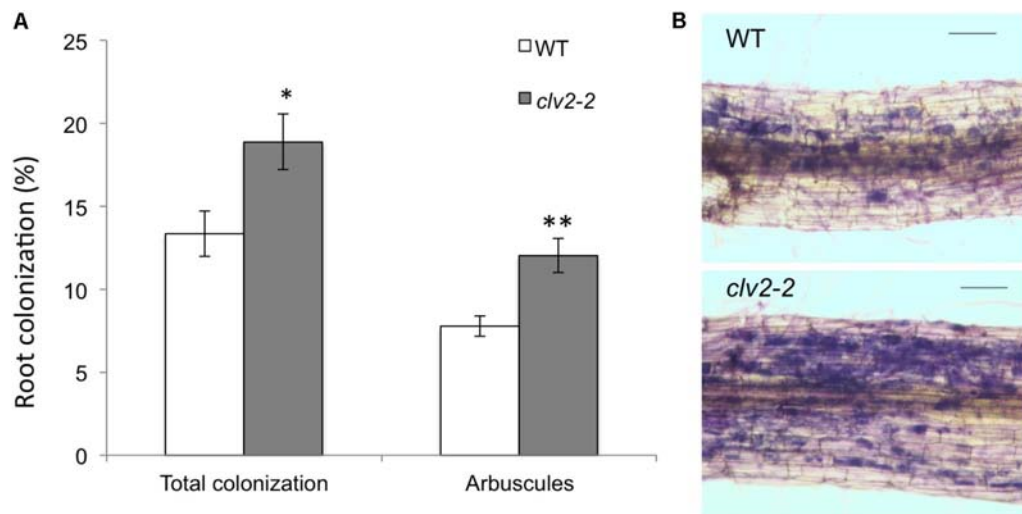
FIGURE 2 | Mycorrhizal colonization in tomato wild type (WT, M82) and *clv2-2* lines. **(A)** Percentage of the root colonized by all fungal structures and arbuscules. *n* = 12, values are means + SE *t*-tests were performed for each parameter, *$P < 0.05$, **$P < 0.01$. **(B)** Photograph of a typical length of colonized root (scale bar = 100 µm) after staining fungal structures (blue). Tomato seeds were germinated in potting mix for 2 weeks and 12 equal sized seedlings of each genotype were selected and transplanted to 2 L pots. The pots were premixed with vermiculite and gravel (1:1) plus mycorrhizal inoculum (1/5 volume of corn roots colonized with *Rhizophagus irregularis* and associated potting medium) and topped with vermiculite. The seedlings were grown in a glasshouse with the following condition: 18 h photoperiod, 25°C day/20°C night. Plants were nutriented three times a week with modified Long Ashton solution (5 mM N and 0.5 mM P) (Hewitt, 1966). Plants were harvested 6 weeks after transplanting. The root was cut into 1 to 1.5 cm segments and stained using the ink and vinegar method (Vierheilig et al., 1998). 25 root segments were selected per plant and mycorrhizal colonization scored using the intersect scoring method (McGonigle et al., 1990). Blind labeling was used to avoid any potential bias during the scoring process.

1990; Foo et al., 2013a). However, phosphate positively regulates nodule number and studies in pea have found this is disrupted in the *sym29* mutant (Foo et al., 2013a), suggesting a cross-over in the AON and phosphate response pathways. MicroRNAs of the 399 family have also been shown to play a role in the phosphate response and some were shown to be induced by phosphate-starvation in AM colonized *Medicago* (Branscheid et al., 2010). However, as overexpression of these miR399 genes in tobacco did not influence AM colonization, no clear role for these mobile microRNAs were established in the AM phosphate response (Branscheid et al., 2010).

In addition to nodulation and AM, plants form a range of other beneficial interactions with soil microbes. These include actinorhizal symbioses between members of the fabid clade and *Frankia* bacteria, ectomycorrhizal symbioses and interactions with fungal and bacterial endophytes. Systemic regulation of colonization has been demonstrated for actinorhizal associations (Wall and Huss-Danell, 1997), and for the interaction between Arabidopsis and the fungal enodphyte *Piriformospora indica* (Pedrotti et al., 2013). Indeed, it has been shown in some split root studies that plants infected with endophytes have a decreased level of AM colonization, although this was not found in all studies (Müller, 2003; Omacini et al., 2006; Mack and Rudgers, 2008). Phylogenetic studies have suggested that the common symbiotic pathway is conserved in AM, rhizobial and actinorhizal associations, although the role of these genes in ectomycorrhizae and endophyte relationships is not known (Martin et al., 2017). Arabidopsis is a particularly interesting case as it appears to have lost the majority of the common symbiotic pathway (Delaux et al.,

2014), consistent with the lack of AM colonization and suggesting that this pathway is not important for hosting fungal endophtyes. However, as these studies did not include AON/AOM genes the evolutionary origin of these pathways and their potential role across species is still not clear.

CONCLUSION AND FUTURE PERSPECTIVES

The similarities between the AOM and AON pathways and their shared genetic components is consistent with the AON pathway evolving at least in part from a pre-existing AOM pathway in early land plants. However, a lack of phylogenetic, genetic and physiological studies in non-legumes, including basal land plants, has hampered our understanding of the origin and diversification of the autoregulation pathway. In this paper, we show that in the non-legume tomato, the *CLV2* gene suppresses AM development, providing the first genetic evidence for an AOM gene in a non-legume. As found in legumes, this gene also plays a role in shoot apical meristem maintenance. However, the precise delineation in function of other AON/AOM elements such as the *CLV1* and *CLE* genes in shoot apical meristem maintenance is still not clear. Furthermore, it is likely that multiple systemic pathways regulate symbioses (Kassaw et al., 2015). For example, novel peptides and accompanying perception pathways with roles in nodulation and root development are now emerging (e.g., CEP1 and CRA2, Mohd-Radzman et al., 2016). Future studies could more systematically examine the role of AON genes and peptide

signals in AM development and take a phylogenetic approach to examine the evolutionary origin of symbiotic autoregulation.

AUTHOR CONTRIBUTIONS

EF conceived the project. CW carried out experiments. CW, JR, and EF wrote the manuscript.

ACKNOWLEDGMENTS

REFERENCES

Bago, B., Pfeffer, P. E., and Shachar-Hill, Y. (2000). Carbon metabolism and transport in arbuscular mycorrhizas. *Plant Physiol.* 124, 949. doi: 10.1104/pp. 124.3.949

Branscheid, A., Sieh, D., Pant, B. D., May, P., Devers, E. A., Elkrog, A., et al. (2010). Expression pattern suggests a role of MiR399 in the regulation of the cellular response to local Pi increase during arbuscular mycorrhizal symbiosis. *Mol. Plant Microbe Interact.* 23, 915–926. doi: 10.1094/MPMI-23-7-0915

Breuillin, F., Schramm, J., Hajirezaei, M., Ahkami, A., Favre, P., Druege, U., et al. (2010). Phosphate systemically inhibits development of arbuscular mycorrhiza in *Petunia hybrida* and represses genes involved in mycorrhizal functioning. *Plant J.* 64, 1002–1017. doi: 10.1111/j.1365-313X.2010.04385.x

Catford, J. G., Staehelin, C., Lerat, S., Piché, Y., and Vierheilig, H. (2003). Suppression of arbuscular mycorrhizal colonization and nodulation in split-root systems of alfalfa after pre-inoculation and treatment with Nod factors. *J. Exp. Bot.* 54, 1481–1487. doi: 10.1093/jxb/erg156

Correa, A., Cruz, C., and Ferrol, N. (2015). Nitrogen and carbon/nitrogen dynamics in arbuscular mycorrhiza: the great unknown. *Mycorrhiza* 25, 499–515. doi: 10.1007/s00572-015-0627-6

Crook, A. D., Schnabel, E. L., and Frugoli, J. A. (2016). The systemic nodule number regulation kinase SUNN in *Medicago truncatula* interacts with MtCLV2 and MtCRN. *Plant J.* 88, 108–119. doi: 10.1111/tpj.13234

Delaux, P.-M., Radhakrishnan, G. V., Jayaraman, D., Cheema, J., Malbreil, M., Volkening, J. D., et al. (2015). Algal ancestor of land plants was preadapted for symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* 112, 13390–13395. doi: 10.1073/pnas. 1515426112

Delaux, P.-M., Séjalon-Delmas, N., Bécard, G., and Ané, J.-M. (2013). Evolution of the plant–microbe symbiotic 'toolkit'. *Trends Plant Sci.* 18, 298–304. doi: 10.1016/j.tplants.2013.01.008

Delaux, P. M., Varala, K., Edger, P. P., Coruzzi, G. M., Pires, J. C., and Ane, J. M. (2014). Comparative phylogenomics uncovers the impact of symbiotic associations on host genome evolution. *PLoS Genet.* 10:e1004487. doi: 10.1371/journal.pgen.1004487

Douds, D. D., Pfeffer, P. E., and Shachar-Hill, Y. (2000). "Carbon partitioning, cost, and metabolism of arbuscular mycorrhizas," in *Arbuscular Mycorrhizas: Physiology and Function*, eds Y. Kapulnik and D. D. Douds (Dordrecht: Springer), 107–129. doi: 10.1007/978-94-017-0776-3_6

Favre, P., Bapaume, L., Bossolini, E., Delorenzi, M., Falquet, L., and Reinhardt, D. (2014). A novel bioinformatics pipeline to discover genes related to arbuscular mycorrhizal symbiosis based on their evolutionary conservation pattern among higher plants. *BMC Plant Biol.* 14:333. doi: 10.1186/s12870-014-0333-0

Ferguson, B. J., Foo, E., Ross, J. J., and Reid, J. B. (2011). Relationship between gibberellin, ethylene and nodulation in *Pisum sativum*. *New Phytol.* 189, 829–842. doi: 10.1111/j.1469-8137.2010.03542.x

Ferguson, B. J., Li, D., Hastwell, A. H., Reid, D. E., Li, Y., Jackson, S. A., et al. (2014). The soybean (*Glycine max*) nodulation-suppressive CLE peptide, GmRIC1, functions interspecifically in common white bean (*Phaseolus vulgaris*), but not in a supernodulating line mutated in the receptor PvNARK. *Plant Biotechnol. J.* 12, 1085–1097. doi: 10.1111/pbi.12216

Field, K. J., Pressel, S., Duckett, J. G., Rimington, W. R., and Bidartondo, M. I. (2015). Symbiotic options for the conquest of land. *Trends Ecol. Evol.* 30, 477–486. doi: 10.1016/j.tree.2015.05.007

Foo, E., Ferguson, B. J., and Reid, J. B. (2014a). Common and divergent roles of plant hormones in nodulation and arbuscular mycorrhizal symbioses. *Plant Signal. Behav.* 9, 1037–1045. doi: 10.4161/psb.29593

Foo, E., Ferguson, B. J., and Reid, J. B. (2014b). The potential roles of strigolactones and brassinosteroids in the autoregulation of nodulation pathway. *Ann. Bot.* 113, 1037–1045. doi: 10.1093/aob/mcu030

Foo, E., Yoneyama, K., Hugill, C., Quittenden, L. J., and Reid, J. B. (2013a). Strigolactones: internal and external signals in plant symbioses? *Plant Signal. Behav.* 8:e23168. doi: 10.4161/psb.23168

Foo, E., Yoneyama, K., Hugill, C. J., Quittenden, L. J., and Reid, J. B. (2013b). Strigolactones and the regulation of pea symbioses in response to nitrate and phosphate deficiency. *Mol. Plant.* 6, 76–87. doi: 10.1093/mp/sss115

Frank, M. H., and Scanlon, M. J. (2015). Transcriptomic evidence for the evolution of shoot meristem function in sporophyte-dominant land plants through concerted selection of ancestral gametophytic and sporophytic genetic programs. *Mol. Biol. Evol.* 32, 355–367. doi: 10.1093/molbev/msu303

Funayama-Noguchi, S., Noguchi, K., Yoshida, C., and Kawaguchi, M. (2011). Two CLE genes are induced by phosphate in roots of *Lotus japonicus*. *J. Plant Res.* 124, 155–163. doi: 10.1007/s10265-010-0342-5

Handa, Y., Nishide, H., Takeda, N., Suzuki, Y., Kawaguchi, M., and Saito, K. (2015). RNA-seq transcriptional profiling of an arbuscular mycorrhiza provides insights into regulated and coordinated gene expression in *Lotus japonicus* and *Rhizophagus irregularis*. *Plant Cell Physiol.* 56, 1490–1511. doi: 10.1093/pcp/pcv071

Hastwell, A. H., Corcilius, L., Williams, J., Gresshoff, P. M., Payne, R. J., and Ferguson, B. J. (2018). Triarabinosylation is required for nodulation-suppressive CLE peptides to systemically inhibit nodulation in *Pisum sativum*. *Plant Cell Environ*. doi: 10.1111/pce.13325 [Epub ahead of print].

Hastwell, A. H., De Bang, T. C., Gresshoff, P. M., and Ferguson, B. J. (2017). CLE peptide-encoding gene families in *Medicago truncatula* and *Lotus japonicus*, compared with those of soybean, common bean and Arabidopsis. *Sci. Rep.* 7:9384. doi: 10.1038/s41598-017-09296-w

Hazak, O., and Hardtke, C. S. (2016). CLAVATA 1-type receptors in plant development. *J. Exp. Bot.* 67, 4827–4833. doi: 10.1093/jxb/erw247

Hewitt, E. J. (1966). *Sand and Water Culture Methods Used in the Study of Plant Nutrition*. Wallingford: Centre for Agriculture and Bioscience International.

Imin, N., Patel, N., Corcilius, L., Payne, R. J., and Djordjevic, M. A. (2018). CLE peptide tri-arabinosylation and peptide domain sequence composition are essential for SUNN-dependent autoregulation of nodulation in *Medicago truncatula*. *New Phytol.* 218, 73–80. doi: 10.1111/nph.15019

Jacobsen, E., and Feenstra, W. J. (1984). A new pea mutant with efficient nodulation in the presence of nitrate. *Plant Sci. Lett.* 33, 337–344. doi: 10.1016/0304-4211(84)90025-7

Kassaw, T., Bridges, W. Jr., and Frugoli, J. (2015). Multiple autoregulation of nodulation (AON) signals identified through split root analysis of *Medicago truncatula* sunn and rdn1 Mutants. *Plants* 4, 209–224. doi: 10.3390/plants4020209

Kassaw, T., Nowak, S., Schnabel, E., and Frugoli, J. A. (2017). ROOT DETERMINED NODULATION1 is required for *M. truncatula* CLE12, but not CLE13 peptide signaling through the SUNN receptor kinase. *Plant Physiol.* 174, 2445–2456. doi: 10.1104/pp.17.00278

Khaosaad, T., Staehelin, C., Steinkellner, S., Hage-Ahmed, K., Ocampo, J. A., Garcia-Garrido, J. M., et al. (2010). The *Rhizobium* sp. strain NGR234 systemically suppresses arbuscular mycorrhizal root colonization in a split-root system of barley (*Hordeum vulgare*). *Physiol. Plant.* 140, 238–245. doi: 10.1111/j.1399-3054.2010.01396.x

Kinkema, M., and Gresshoff, P. M. (2008). Investigation of downstream signals of the soybean autoregulation of nodulation receptor kinase GmNARK. *Mol. Plant Microbe Interact.* 21, 1337–1348. doi: 10.1094/MPMI-21-10-1337

Krusell, L., Sato, N., Fukuhara, I., Koch, B. E. V., Grossmann, C., Okamoto, S., et al. (2011). The Clavata2 genes of pea and *Lotus japonicus* affect autoregulation of nodulation. *Plant J.* 65, 861–871. doi: 10.1111/j.1365-313X.2010.04474.x

Laffont, C., Rey, T., André, O., Novero, M., Kazmierczak, T., Debellé, F., et al. (2015). The CRE1 cytokinin pathway is differentially recruited depending on *Medicago truncatula* root environments and negatively regulates resistance to a pathogen. *PLoS One* 10:e0116819. doi: 10.1371/journal.pone.0116819

Li, D., Kinkema, M., and Gresshoff, P. M. (2009). Autoregulation of nodulation (AON) in *Pisum sativum* (pea) involves signalling events associated with both nodule primordia development and nitrogen fixation. *J. Plant Physiol.* 166, 955–967. doi: 10.1016/j.jplph.2009.03.004

Lim, C. W., Lee, Y. W., and Hwang, C. H. (2011). Soybean nodule-enhanced CLE peptides in roots act as signals in GmNARK-mediated nodulation suppression. *Plant Cell Physiol.* 52, 1613–1627. doi: 10.1093/pcp/pcr091

Lin, Y. H., Ferguson, B. J., Kereszt, A., and Gresshoff, P. M. (2010). Suppression of hypernodulation in soybean by a leaf-extracted, NARK- and Nod factor-dependent, low molecular mass fraction. *New Phytol.* 185, 1074–1086. doi: 10.1111/j.1469-8137.2009.03163.x

Lohar, D. P., and VandenBosch, K. A. (2005). Grafting between model legumes demonstrates roles for roots and shoots in determining nodule type and host/rhizobia specificity. *J. Exp. Bot.* 56, 1643–1650. doi: 10.1093/jxb/eri160

Mack, K. M. L., and Rudgers, J. A. (2008). Balancing multiple mutualists: asymmetric interactions among plants, arbuscular mycorrhizal fungi, and fungal endophytes. *OIKOS* 117, 310–320. doi: 10.1111/j.2007.0030-1299.15973.x

Martin, F. M., Uroz, S., and Barker, D. G. (2017). Ancestral alliances: plant mutualistic symbioses with fungi and bacteria. *Science* 356:eaad4501. doi: 10.1126/science.aad4501

McGonigle, T. P., Miller, M. H., Evans, D. G., Fairchild, G. L., and Swan, J. A. (1990). A new method which gives an objective measure of colonization of roots by vesicular—arbuscular mycorrhizal fungi. *New Phytol.* 115, 495–501. doi: 10.1111/j.1469-8137.1990.tb00476.x

Meixner, C., Ludwig-Muller, J., Miersch, O., Gresshoff, P., Staehelin, C., and Vierheilig, H. (2005). Lack of mycorrhizal autoregulation and phytohormonal changes in the supernodulating soybean mutant nts1007. *Planta* 222, 709–715. doi: 10.1007/s00425-005-0003-4

Meixner, C., Vegvari, G., Ludwig-Müller, J., Gagnon, H., Steinkellner, S., Staehelin, C., et al. (2007). Two defined alleles of the LRR receptor kinase GmNARK in supernodulating soybean govern differing autoregulation of mycorrhization. *Physiol. Plant.* 130, 261–270. doi: 10.1111/j.1399-3054.2007.00903.x

Mirzaei, S., Batley, J., El-Mellouki, T., Liu, S., Meksem, K., Ferguson, B. J., et al. (2017). Neodiversification of homeologous CLAVATA1-like receptor kinase genes in soybean leads to distinct developmental outcomes. *Sci. Rep.* 7:8878. doi: 10.1038/s41598-017-08252-y

Miyazawa, H., Oka-Kira, E., Sato, N., Takahashi, H., Wu, G. J., Sato, S., et al. (2010). The receptor-like kinase KLAVIER mediates systemic regulation of nodulation and non-symbiotic shoot development in *Lotus japonicus*. *Development* 137, 4317–4325. doi: 10.1242/dev.058891

Mohd-Radzman, N. A., Laffont, C., Ivanovici, A., Patel, N., Reid, D. E., Stougaard, J., et al. (2016). Different pathways act downstream of the peptide receptor CRA2 to regulate lateral root and nodule development. *Plant Physiol.* 171, 2536–2548. doi: 10.1104/pp.16.00113

Morandi, D., Sagan, M., Prado-Vivant, E., and Duc, G. (2000). Influence of genes determining supernodulation on root colonization by the mycorrhizal fungus Glomus mosseae in *Pisum sativum* and *Medicago truncatula* mutants. *Mycorrhiza* 10, 37–42. doi: 10.1007/s005720050285

Mortier, V., De Wever, E., Vuylsteke, M., Holsters, M., and Goormachtig, S. (2012). Nodule numbers are governed by interaction between CLE peptides and cytokinin signaling. *Plant J.* 70, 367–376. doi: 10.1111/j.1365-313X.2011.04881.x

Mortier, V., Den Herder, G., Whitford, R., Van De Velde, W., Rombauts, S., D'haeseleer, K., et al. (2010). CLE peptides control *Medicago truncatula* nodulation locally and systemically. *Plant Physiol.* 153, 222–237. doi: 10.1104/pp.110.153718

Müller, J. (2003). Artificial infection by endophytes affects growth and mycorrhizal colonisation of *Lolium perenne*. *Funct. Plant Biol.* 30, 419–424. doi: 10.1071/FP02189

Oka-Kira, E., Tateno, K., Miura, K. I., Haga, T., Hayashi, M., Harada, K., et al. (2005). klavier (klv), A novel hypernodulation mutant of *Lotus japonicus* affected in vascular tissue organization and floral induction. *Plant J.* 44, 505–515. doi: 10.1111/j.1365-313X.2005.02543.x

Okamoto, S., and Kawaguchi, M. (2015). Shoot HAR1 mediates nitrate inhibition of nodulation in *Lotus japonicus*. *Plant Signal. Behav.* 10:e1000138. doi: 10.1080/15592324.2014.1000138

Okamoto, S., Ohnishi, E., Sato, S., Takahashi, H., Nakazono, M., Tabata, S., et al. (2009). Nod factor/nitrate-induced CLE genes that drive HAR1-mediated systemic regulation of nodulation. *Plant Cell Physiol.* 50, 67–77. doi: 10.1093/pcp/pcn194

Okamoto, S., Shinohara, H., Mori, T., Matsubayashi, Y., and Kawaguchi, M. (2013). Root-derived CLE glycopeptides control nodulation by direct binding to HAR1 receptor kinase. *Nat. Commun.* 4:2191. doi: 10.1038/ncomms3191

Omacini, M., Eggers, T., Bonkowski, M., Gange, A. C., and Jones, T. H. (2006). Leaf endophytes affect mycorrhizal status and growth of co-infected and neighbouring plants. *Funct. Ecol.* 20, 226–232. doi: 10.1111/j.1365-2435.2006.01099.x

Pedrotti, L., Mueller, M. J., and Waller, F. (2013). *Piriformospora indica* root colonization triggers local and systemic root responses and inhibits secondary colonization of distal roots. *PLoS One* 8:e69352. doi: 10.1371/journal.pone.0069352

Reid, D. E., Ferguson, B. J., and Gresshoff, P. M. (2011a). Inoculation-and nitrate-induced CLE peptides of soybean control NARK-dependent nodule formation. *Mol. Plant Microbe Interact.* 24, 606–618. doi: 10.1094/MPMI-09-10-0207

Reid, D. E., Ferguson, B. J., Hayashi, S., Lin, Y. H., and Gresshoff, P. M. (2011b). Molecular mechanisms controlling legume autoregulation of nodulation. *Ann. Bot.* 108, 789–795. doi: 10.1093/aob/mcr205

Reid, D. E., Nadzieja, M., Novak, O., Heckmann, A. B., Sandal, N., and Stougaard, J. (2017). Cytokinin biosynthesis promotes cortical cell responses during nodule development. *Plant Physiol.* 175:pp.00832.2017. doi: 10.1104/pp.17.00832

Sakamoto, K., and Nohara, Y. (2009). Soybean (*Glycine max* [L.] Merr.) shoots systemically control arbuscule formation in mycorrhizal symbiosis. *Soil Sci. Plant Nutr.* 55, 252–257. doi: 10.1111/j.1747-0765.2009.00358.x

Sakamoto, K., Ogiwara, N., and Kaji, T. (2013). Involvement of autoregulation in the interaction between rhizobial nodulation and AM fungal colonization in soybean roots. *Biol. Fertil. Soils* 49, 1141–1152. doi: 10.1007/s00374-013-0804-8

Sasaki, T., Suzaki, T., Soyano, T., Kojima, M., Sakakibara, H., and Kawaguchi, M. (2014). Shoot-derived cytokinins systemically regulate root nodulation. *Nat. Commun.* 5:4983. doi: 10.1038/ncomms5983

Schaarschmidt, S., Gresshoff, P. M., and Hause, B. (2013). Analyzing the soybean transcriptome during autoregulation of mycorrhization identifies the transcription factors GmNF-YA1a/b as positive regulators of arbuscular mycorrhization. *Genome Biol.* 14:R62. doi: 10.1186/gb-2013-14-6-r62

Schnabel, E., Journet, E. P., De Carvalho-Niebel, F., Duc, G., and Frugoli, J. (2005). The *Medicago truncatula* SUNN gene encodes a CLV1-like leucine-rich repeat receptor kinase that regulates nodule number and root length. *Plant Mol. Biol.* 58, 809–822. doi: 10.1007/s11103-005-8102-y

Schnabel, E. L., Kassaw, T. K., Smith, L. S., Marsh, J. F., Oldroyd, G. E., Long, S. R., et al. (2011). The *ROOT DETERMINED NODULATION1* gene regulates nodule number in roots of *Medicago truncatula* and defines a highly conserved, uncharacterized plant gene family. *Plant Physiol.* 157, 328–340. doi: 10.1104/pp.111.178756

Shrihari, P. C., Sakamoto, K., Inubushi, K., and Akao, S. (2000). Interaction between supernodulating or non-nodulating mutants of soybean and two arbuscular mycorrhizal fungi. *Mycorrhiza* 10, 101–106. doi: 10.1007/s005720000064

Solaiman, Z. M., Senoo, K., Kawaguchi, M., Imaizumi-Anraku, H., Akao, S., Tanaka, A., et al. (2000). Characterization of mycorrhizas formed by glomus sp. on roots of hypernodulating mutants of *Lotus japonicus*. *J. Plant Res.* 113, 443–448. doi: 10.1007/PL00013953

Soyano, T., and Kawaguchi, M. (2014). "Systemic regulation of root nodule formation," in *Advances in Biology and Ecology of Nitrogen Fixation*, ed. T. Ohyama (Rijeka: InTech).

Sun, X., and Wang, G.-L. (2011). Genome-wide identification, characterization and phylogenetic analysis of the rice LRR-Kinases. *PLoS One* 6:e16079. doi: 10.1371/journal.pone.0016079

Takahara, M., Magori, S., Soyano, T., Okamoto, S., Yoshida, C., Yano, K., et al. (2013). Too much love, a novel Kelch repeat-containing F-box protein, functions in the long-distance regulation of the legume-Rhizobium symbiosis. *Plant Cell Physiol.* 54, 433–447. doi: 10.1093/pcp/pct022

van Noorden, G. E., Ross, J. J., Reid, J. B., Rolfe, B. G., and Mathesius, U. (2006). Defective long-distance auxin transport regulation in the *Medicago truncatula* super numeric nodules mutant. *Plant Physiol.* 140, 1494–1506. doi: 10.1104/pp.105.075879

Vierheilig, H. (2004). Further root colonization by arbuscular mycorrhizal fungi in already mycorrhizal plants is suppressed after a critical level of root colonization. *J. Plant Physiol.* 161, 339–341. doi: 10.1078/0176-1617-01097

Vierheilig, H., Coughlan, A. P., Wyss, U., and Piché, Y. (1998). Ink and vinegar, a simple staining technique for arbuscular-mycorrhizal fungi. *Appl. Environ. Microbiol.* 64, 5004–5007.

Vierheilig, H., Garcia-Garrido, J., Wyss, U., and Piché, Y. (2000a). Systemic suppression of mycorrhizal colonization of barley roots already colonized by AM fungi. *Soil Biol. Biochem.* 32, 589–595. doi: 10.1016/S0038-0717(99)00155-8

Vierheilig, H., Maier, W., Wyss, U., Samson, J., Strack, D., and Piché, Y. (2000b). Cyclohexenone derivative- and phosphate-levels in split-root systems and their role in the systemic suppression of mycorrhization in precolonized barley plants. *J. Plant Physiol.* 157, 593–599. doi: 10.1016/S0176-1617(00)80001-2

Vierheilig, H., Steinkellner, S., Khaosaad, T., and Garcia-Garrido, J. M. (2008). "The biocontrol effect of mycorrhization on soilborne fungal pathogens and the autoregulation of the AM symbiosis: one mechanism, two effects?" in *Mycorrhiza: State of the Art, Genetics and Molecular Biology, Eco-Function, Biotechnology, Eco-Physiology, Structure and Systematics*, ed. A. Varma (Berlin: Springer), 307–320.

Wall, L. G., and Huss-Danell, K. (1997). Regulation of nodulation in Alnus incana-Frankia symbiosis. *Physiol. Plant.* 99, 594–600. doi: 10.1111/j.1399-3054.1997.tb05362.x

Wei, Z., Wang, J., Yang, S., and Song, Y. (2015). Identification and expression analysis of the LRR-RLK gene family in tomato (Solanum lycopersicum) Heinz 1706. *Genome* 58, 121–134. doi: 10.1139/gen-2015-0035

Wyss, P., Mellor, R. B., and Wiemken, A. (1990). Mutants of soybean (*Glycine max*) unable to suppress nodulation in the presence of nitrate retain the ability to suppress mycorrhization in the presence of phosphate. *J. Plant Physiol.* 136, 507–509. doi: 10.1016/S0176-1617(11)80045-3

Xu, C., Liberatore, K. L., Macalister, C. A., Huang, Z., Chu, Y.-H., Jiang, K., et al. (2015). A cascade of arabinosyltransferases controls shoot meristem size in tomato. *Nat. Genet.* 47, 784–792. doi: 10.1038/ng.3309

Yamamoto, E., Karakaya, H. C., and Knap, H. T. (2000). Molecular characterization of two soybean homologs of Arabidopsis thaliana CLAVATA1 from the wild type and fasciation mutant11The sequence data reported in this paper have been submitted to the GenBank data libraries under the accession numbers AF197946 for GmCLV1A and AF197947 for GmCLV1B. *Biochim. Biophys. Acta* 1491, 333–340. doi: 10.1016/S0167-4781(00)00061-0

Zan, Y., Ji, Y., Zhang, Y., Yang, S., Song, Y., and Wang, J. (2013). Genome-wide identification, characterization and expression analysis of populusleucine-rich repeat receptor-like protein kinase genes. *BMC Genomics* 14:318. doi: 10.1186/1471-2164-14-318

Check for
updates

Tomato LysM Receptor-Like Kinase SlLYK12 Is Involved in Arbuscular Mycorrhizal Symbiosis

Dehua Liao†‡, Xun Sun‡, Ning Wang, Fengming Song and Yan Liang*

College of Agriculture and Biotechnology, Zhejiang University, Hangzhou, China

Arbuscular mycorrhiza (AM) is a widespread symbiotic relationship between plants and fungi (Glomeromycota), which improves the supply of water and nutrients to host plants. AM symbiosis is set in motion by fungal chitooligosaccharides and lipochitooligosaccharides, which are perceived by plant-specific LysM-type receptor kinases (LYK). In rice this involves OsCERK1, a LYK also essential for chitin triggered innate immunity. In contrast in legumes, the CERK1 homologous gene experienced duplication events resulting in subfunctionalization. However, it remains unknown whether this subfunctionalization is legume-specific, or has occurred also in other dicot plant species. We identified four CERK1 homologs in tomato (SlLYK1, SlLYK11, SlLYK12, and SlLYK13) and investigated their roles in chitin signaling and AM symbiosis. We found that knockdown of SlLYK12 in tomato significantly reduced AM colonization, whereas chitin-induced responses were unaffected. In contrast, knockdown of SlLYK1 resulted in reduced responses to chitin, but did not alter responses to AM fungi. Moreover, ectopic overexpression of SlLYK1 and SlLYK13 in *Nicotiana benthamiana* induced cell death, whereas SlLYK12 overexpression did not. Based on our results and comparison with rice OsCERK1, we hypothesize that OsCERK1 orthologs in tomato underwent gene duplication, leading to the subfunctionalization of immunity and symbiosis.

Keywords: innate immunity, chitin, chitin elicitor receptor kinase (CERK1), arbuscular mycorrhiza (AM) symbiosis, LysM receptor-like kinase (LYK), *Solanum lycopersicum*, SlLYK1, SlLYK12

INTRODUCTION

N-acetyl-D-glucosamine (GlcNAc)-containing molecules are important microbial signaling factors, and include chitin from pathogenic fungi, peptidoglycan from pathogenic bacteria, Nod factors from symbiotic rhizobia, and Myc factors from symbiotic arbuscular mycorrhizal (AM) fungi. These molecules mediate the initiation of either plant innate immune functions or symbiotic pathways (Liang et al., 2014; Zipfel and Oldroyd, 2017).

Chitin, a polymer of GlcNAc, is the major component of the fungal cell wall. When fungi infect plants, plants secrete chitinases to hydrolyze chitin, producing chitooligosaccharides among which those with degrees of polymerization (dp) between 6 and 8 elicit plant immune responses. Such immune responses include the elevation of cytosolic calcium, production of reactive oxygen species (ROS), induction of defense-related gene expression, callose deposition, and pathogen growth restriction (Boller and Felix, 2009; Dodds and Rathjen, 2010). In *Arabidopsis thaliana*, chitin is primarily recognized by LysM RECEPTOR KINASE5 (AtLYK5), which has an inactive

kinase domain (Cao et al., 2014). After perception of chitin, AtLYK5 forms a heterotetramer complex with CHITIN ELICITOR RECEPTOR KINASE1 (AtCERK1), activating the AtCERK1 intracellular kinase domain and downstream immune responses (Cao et al., 2014; Zipfel and Oldroyd, 2017; Desaki et al., 2018). In rice, CHITIN ELICITOR-BINDING PROTEIN (OsCEBiP), which has no kinase domain, play the major role in chitin perception, and transduces signals via a similar mechanism of complex formation with OsCERK1(Shimizu et al., 2010; Hayafune et al., 2014). Recently, a *Lotus japonicus* ortholog of CERK1, LjCERK6, required for chitin responses was identified (Bozsoki et al., 2017). Thus, CERK1 kinase activity is a key factor in the induction of chitin-induced immune responses.

In contrast to chitin oligomers, acylated chitooligosaccharides (so called lipochitooligosaccharides, LCOs) can act as signaling molecules triggering symbiosis; for example, Nod factors from rhizobia. Nod-LCOs are also recognized by the LysM-containing receptors, called NFR1-NFR5 (NF RECEPTOR1 and 5) complex in *L. japonicus*, and LYK3-NFP (NF PERCEPTION) complex in *Medicago truncatula* (Ben Amor et al., 2003; Limpens et al., 2003; Madsen et al., 2003; Radutoiu et al., 2003; Broghammer et al., 2012). Similar to the chitin recognition model in *Arabidopsis*, the kinase domain of NFR5/NFP is inactive, and the Nod-LCO signal is transduced via NFR1/LYK3 kinase activation (Kelly et al., 2017; Zipfel and Oldroyd, 2017; Desaki et al., 2018).

GlcNAc-containing molecules are also important signals (Myc factors) for AM symbiosis, a widespread symbiotic relationship occurring between fungi in the phylum Glomeromycota and 80% of terrestrial plant species under phosphorus- or nitrogen-limiting conditions (Schussler et al., 2001; Nouri et al., 2014). The AM symbiosis is characterized by the formation of arbuscule structures in the root cortex, the main site for nutrient exchange between the two symbiotic partners (Harrison, 2012). Myc factors are thought to be a mixture of short-chain chitin (dp = 3–5) and Myc-LCOs (Maillet et al., 2011; Genre et al., 2013). Perception of Myc factors activates the common symbiosis signaling pathway shared with Nod factor signaling in potential host plants, leading to the colonization of AM fungi in root epidermal cells (Oldroyd, 2013). Given the molecular similarities between Myc-LCO and Nod-LCO, as well as the shared common symbiosis signaling pathway, it has been proposed that Nod factor recognition may have evolved from Myc factor receptors (De Mita et al., 2014). Indeed, the ortholog of NFP/NFR5 in *Parasponia andersonii*, the only non-leguminous plant that can establish symbiosis with rhizobia, is required for both AM and rhizobial symbioses (Op den Camp et al., 2011). Similarly, knockdown of the ortholog of NFP/NFR5 in tomato (SlLYK10) affects AM colonization (Buendia et al., 2016), whereas the ortholog of NFP/NFR5 in rice is not required for AM symbiosis (Miyata et al., 2014). Interestingly, OsCERK1, the ortholog of NFR1/LYK3 in rice, is involved in both chitin and Myc factor signaling (Zhang et al., 2015; Miyata et al., 2016; Carotenuto et al., 2017). However, this dual function of CERK1 in AM symbiosis and chitin-triggered innate immunity is separated of two paralogous genes in legumes (Bozsoki et al., 2017). This raises the question whether such subfunctionalization upon gene duplication is specific for the legume family,

or may have occurred also in non-related dicot species; e.g., tomato.

In this study, we identified the orthologs of the CERK1 subclade in tomato, and investigated their function by knocking down their expression. We found that knockdown of *SlLYK12* significantly reduced AM colonization; however, the chitin responses of these plants were similar to those of controls. In contrast, knockdown of *SlLYK1* resulted in reduced responses to chitin, but normal responses to AM fungi. In addition, we found that ectopic overexpression of *SlLYK1* and *SlLYK13* in *Nicotiana benthamiana* caused cell death; however, *SlLYK12* overexpression did not. Taken together, these results suggest a hypothesis whereby an ancestor of *CERK1* with dual function in both immunity and symbiosis gave rise to multiple molecules during evolution through gene duplication in tomato, among which SlLYK1 and SlLYK12 were sub-functionalized for a role in immunity and symbiosis, respectively.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

Solanum lycopersicum L. cv Zheza 809 was used for all experiments. For virus-induced gene silencing (VIGS) experiments, seedlings were grown in a plant growth room at 22°C with a 16 h photoperiod. For *Rhizophagus irregularis* inoculation, plants were grown at 25°C with a 16 h photoperiod.

RNA Isolation and Quantitative RT-PCR (qRT-PCR)

Total RNA was isolated using TRIzol reagent (Invitrogen, Waltham, MA, United States). RNA samples were treated with DNase I to eliminate potential contamination with genomic DNA. qRT-PCR was performed on an Applied Biosystems Plus Real-Time PCR System (ABI, Foster city, CA, United States) using a SYBR premix Ex Taq kit (Takara, Mountain View, CA, United States). Primers are listed in Supplementary Table 1.

Gene Cloning and Plasmid Construction

The primers used for gene cloning are listed in Supplementary Table 1. Full length coding sequences for ectopic overexpression and fragments for VIGS experiments were amplified from cDNA. The amplified sequences were cloned into the pDONR/Zeo plasmid by BP cloning (Invitrogen, Waltham, MA, United States). After verification by sequencing, resultant plasmids were used for LR cloning into the destination plasmids pTRV2 for VIGS and pMDC83 for overexpression in *N. benthamiana*. All plasmids were introduced into *Agrobacterium tumefaciens* strain GV3101 by electroporation.

Gene Silencing Assay Using *Tobacco rattle virus* (TRV)

Agrobacteria carrying pTRV2-*GUS* (β-*GLUCURONIDASE*, a negative control), pTRV2-*NbPDS* (*PHYTOENE DESATURASE*, a positive control for monitoring the progress of gene silencing), pTRV2-*SlLYK1*, *12*, *13*, and pTRV1 were cultivated in YEP

medium (10 g/L yeast extract, 10 g/L peptone, 5 g/L NaCl, 50 μg/mL kanamycin, 50 μg/mL rifampicin, and 25 μg/mL gentamicin) for 36 h at 28°C. Cultures were passaged in fresh medium at a dilution of 1:100 and cultivated for a further 8 h. After adjusting the concentration to $OD_{600} = 1.5$, each pTRV2 construct was mixed with pTRV1 (1:1) in infiltration buffer (10 mM MgCl$_2$, 10 mM MES, 150 μM acetosyringone, pH 5.7). The agrobacterial mix was infiltrated into the abaxial surface of 10-day-old tomato seedlings. Gene silencing efficiency and specificity were determined 4 weeks after agrobacterial infiltration. At least six individual seedlings were analyzed for each construct.

Mycorrhizal Inoculation

Rhizophagus irregularis was purchased from the Institute of Plant Nutrition and Resources, Beijing Academy of Agriculture and Forestry Sciences, and propagated using *N. tabacum* as the host in 4 L pots containing sand. Plants were watered with a solution without added phosphorous every week. Four months later, the sand containing *R. irregularis* was harvested and dried to obtain mycorrhizal inoculums. For mycorrhizal inoculation, after VIGS infiltration, each tomato plant was transferred to a 4 L pot containing sand including 4 g of sand-based mycorrhizal inoculums at the base of the roots.

Detection of Mycorrhizal Colonization

Roots were cleared with 10% (w/v) KOH for 1 h at 90°C, acidified with 2% HCl for 5 min, and then stained with trypan blue (0.5 mM). Mycorrhizal colonization was observed under a light microscope (Nikon, Tokyo, Japan). The rate of root colonization was quantified using the grid line intersect method at 3 and 6 weeks after mycorrhizal inoculation, and calculated as the ratio of intersects with hyphopodia, intracellular hyphae, arbuscules and vesicles over all root intersects (100 intersects per plant) × 100 (Vierheilig et al., 1998).

Reactive Oxygen Species (ROS) Assay

Leaf disks (diameter, 0.5 cm) were punched and incubated in water for at least 8 h. After addition of water containing 1.25 μM L-012 chemiluminescent probe (Wako Chemicals USA, Richmond, VA, United States), 20 μg/mL horseradish peroxidase, and 500 nM chitooctaose (IsoSep, Tullinge, Sweden), chemiluminescent signals were immediately recorded using a Photek camera (HRPCS5; Photek Ltd., East Sussex, United Kingdom) for 30 min.

Detection of Cell Death in *N. benthamiana* Leaves

Cell death in *N. benthamiana* leaves was detected by trypan blue staining. Excised leaves were vacuum-infiltrated with trypan blue solution (2.6 mM) for 30 min, and incubated for a further 8 h. Leaves were then destained in a solution containing ethanol and glycerol at a ratio of 9:1 at 65°C for 30 min.

Antibodies and Immunoblot Analysis

Immunoblot analysis was performed as previously described (Liao et al., 2017) using anti-phospho-p44/p42 MAP kinase antibody (Cell Signaling Technology, Danvers, MA, United States).

Bioinformatics Analysis

The amino acid sequences of LYK genes are listed in Supplementary Table 2. Multiple sequence alignments were performed using the ClustalX program (version 1.83) with default gap penalties. An approximately maximum-likelihood tree was constructed using the FastTree program with default parameters[1].

RESULTS

SlLYK12 Is Required for AM Symbiosis

To identify orthologs of OsCERK1/AtCERK1 in tomato, we generated a phylogenetic tree using the full length amino acid sequences of *S. lycopersicum* LYKs. The results showed that four genes from *S. lycopersicum*, *SlLYK1*, *SlLYK11*, *SlLYK12*, and *SlLYK13*, were clustered into one clade with *AtCERK1*, similar to the tree constructed by Buendia et al. (2016) using intracellular region sequences. Previously generated RNA sequencing data indicate that the *SlLYK1* gene is expressed at similar levels in roots and leaves, and the *SlLYK12* and *SlLYK13* genes were each primarily expressed in roots and leaves, respectively (Supplementary Figure 1A) (Tomato Genome, 2012). We confirmed these results using qRT-PCR (Supplementary Figure 1B). The expression level of *SlLYK11* was much lower in both roots and leaves compared to the other three genes (Supplementary Figure 1). In addition, we were unable to amplify the predicted full-length coding sequence of the *SlLYK11* gene, even using cDNA extracted from chitooctaose (CO8)-treated leaves and AM-inoculated roots; therefore, we did not perform further analysis of the *SlLYK11* gene.

To study the function of *SlLYK1*, *SlLYK12*, and *SlLYK13*, we silenced these genes individually using a VIGS approach, which is a powerful tool for the study of AM symbiosis in the tomato (Buendia et al., 2016). As the sequence similarity between these three genes is 74%, we designed three sets of primers for each gene and each amplified region was fused to the VIGS vector. The transcript levels of *SlLYK1*, *SlLYK12*, and *SlLYK13* were detected by qRT-PCR in VIGS leaves 4 weeks after agrobacterial infiltration. The best sets of primers were chosen according to the specificity and efficiency of gene silencing; for example, *SlLYK1* gene expression was down-regulated 50% in VIGS-*SlLYK1* leaves compared to VIGS-*GUS* (β-*GLUCURONIDASE*) control, but the transcript levels of *SlLYK12* and *SlLYK13* did not show significant differences (**Figure 1A**). After confirming the silencing effectiveness in roots (**Figure 1B**), roots were inoculated with *R. irregularis* and grown for another 3 and 6 more weeks. To observe mycorrhizal colonization, roots were harvested and stained with trypan blue. Arbuscules could be observed in plants infiltrated with all constructs (**Figures 2A,B**). The rate of total root colonization including hyphopodia, intracellular hyphae, arbuscules, and vesicles was calculated using the grid line

---

[1]http://www.microbesonline.org/fasttree/
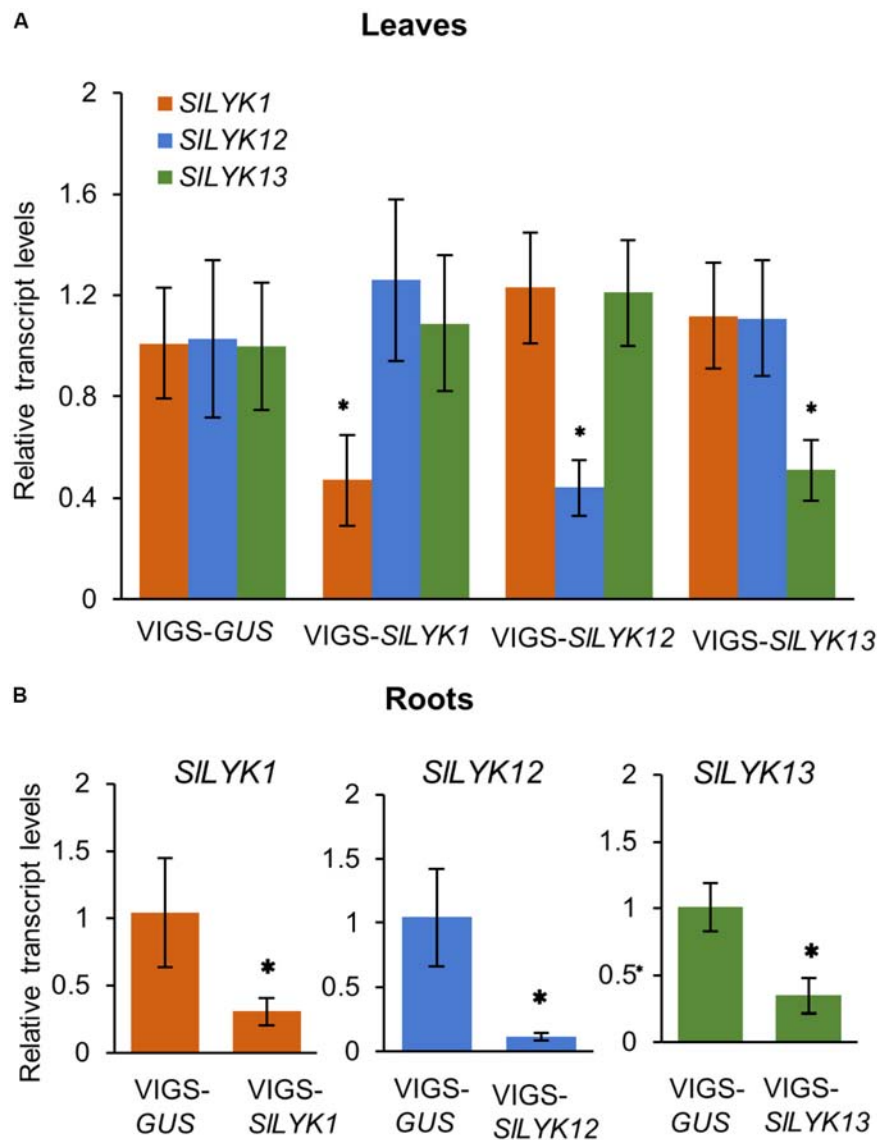
FIGURE 1 | Each endogenous gene of CERK1 family in tomato was specifically silenced by its respective VIGS construct. **(A)** The relative transcript levels of *SlLYK1*, *SlLYK12*, and *SlLYK13* in leaves infiltrated with VIGS-*SlLYK1*, *-12*, *-13*, or -*GUS* (control). **(B)** Relative transcript levels in roots. RNA was extracted from leaves and roots 4 weeks after leaf infiltration with *Agrobacterium tumefaciens* carrying the indicated constructs. Transcript levels were detected using qRT-PCR. Data are expressed as means ± SD from three biological replicates. Asterisks indicate significant differences from the VIGS-*GUS* control (Student's *t*-test: *$P \leq 0.05$).

intersect method (Vierheilig et al., 1998). We found that VIGS-*SlLYK12*- infiltrated plants exhibited a more than 50% reduction of mycorrhizal colonization, whereas those infiltrated with VIGS-*SlLYK1* and -*SlLYK13* did not exhibit significant differences from control plants (**Figures 2C,D**). Similarly, the percentage of hyphopodia and arbuscules were significantly reduced in VIGS-*SlLYK12*- infiltrated plants (**Figures 2E,F**). Consistent with these results, levels of the *SlLYK12* transcript were increased fourfold after *R. irregularis* inoculation, whereas those of *SlLYK1* and *SlLYK13* were not (**Figure 2G**).

To confirm that *SlLYK12* is required for the development of AM symbiosis, we determined the expression levels of a fungal housekeeping gene *GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE* (*RiGAPDH*), and three AM responsive genes in tomato, including *REDUCED ARBUSCULAR MYCORRHIZATION2* (*SlRAM2*, an early signaling gene in the common symbiosis pathway), *BLUE COPPER-BINDING PROTEIN1* (*SlBCP1*, a gene induced in arbuscule-containing regions), and *PHOSPHATE TRANSPORTER4* (*SlPT4*, a late arbuscule developmental gene) (Liu et al., 2007; Buendia et al., 2016). Compared with control plants, the expression levels of all four genes were significantly reduced in VIGS-*SlLYK12*-infiltrated plants, suggesting that *SlLYK12* affects the development of AM symbiosis (**Figure 2H**).
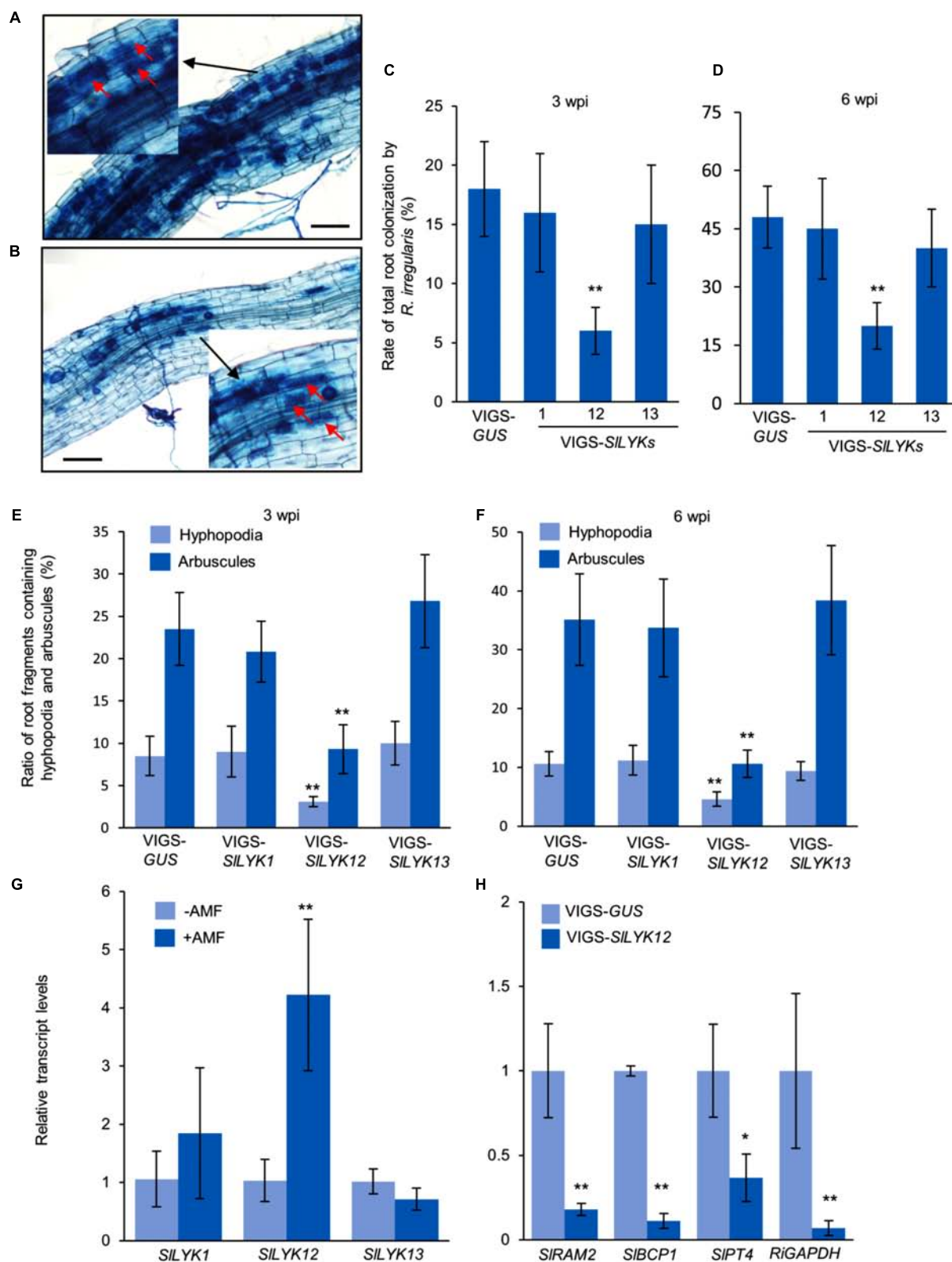
FIGURE 2 | SILYK12 is required for AM symbiosis. **(A,B)** Images of roots inoculated with *Rhizophagus irregularis* after infiltration with VIGS-*GUS* control **(A)** and VIGS-*SILYK12* constructs **(B)**. **(C,D)** Rate of total root colonization by *R. irregularis* 3 and 6 weeks post inoculation (wpi), respectively. **(E,F)** Ratio of root fragments
(Continued)

SlLYK1 Is Required for Chitin Responses

To determine whether *SlLYK1*, *SlLYK12*, and *SlLYK13* are required for chitin signaling, we analyzed their expression after CO8 treatment in 4-week-old leaves and roots. Our results indicated that *SlLYK1* gene expression is upregulated sevenfold after CO8 treatment in leaves and threefold in roots. *SlLYK13* is also slightly upregulated in leaves after CO8 treatment, whereas *SlLYK12* did not show significant upregulation in either leaves or roots (**Figures 3A,B**). These results suggest that *SlLYK12* might not have a role in chitin signaling. Next, we examined the responses of leaves infiltrated with VIGS-*SlLYK1*, -*SlLYK12*, -*SlLYK13*, and VIGS control to CO8 elicitation. First, ROS generation was measured after CO8 treatment, using a luminol-based chemiluminescence detection system. We found that VIGS-*SlLYK1*-inoculated leaves showed reduced ROS levels after CO8 treatment compared to the VIGS control, whereas plants infiltrated with VIGS-*SlLYK12* and -*SlLYK13* did not exhibit lower ROS levels (**Figure 3C**). Second, transcript levels of *SlWRKY53* (the ortholog of *AtWRKY53*), a typical chitin response gene (Cao et al., 2014), were detected by qRT-PCR. The results indicated that induction of *SlWRKY53* expression triggered by CO8 was significantly reduced in VIGS-*SlLYK1*-infiltrated plants (**Figure 3D**). Third, MAP kinase activity was analyzed by immunoblot assay. CO8 treatment causes MAP kinase phosphorylation, as demonstrated using an anti-p42/p44-MAPK antibody in VIGS control plants; however, levels of phosphorylation were significantly reduced in VIGS-*SlLYK1*-inoculated plants (**Figure 3E**). Similar to the results in leaves, CO8-triggered *SlWRKY53* gene expression and MAPK phosphorylation in roots was only reduced in VIGS-*SlLYK1* infiltrated plants (**Figures 3F,G**). Together, these results suggest that SlLYK1 plays a role in chitin signaling.

SlLYK13 Is Involved in Cell Death

The results described above (sections "SlLYK12 Is Required for AM Symbiosis" and "SlLYK1 Is Required for Chitin Responses") suggest that SlLYK1 is required for chitin signaling, while SlLYK12 is involved in AM symbiosis; hence, we wished to determine the function of SlLYK13. AtCERK1 has a chitin-independent role in cell death (Petutschnig et al., 2014), and ectopic overexpression of *AtCERK1* in *N. benthamiana* leaves results in symptoms of cell death (Pietraszewska-Bogiel et al., 2013); therefore, we analyzed the signs of cell death in *N. benthamiana* leaves ectopically overexpressing *SlLYK1*, *SlLYK12*, and *SlLYK13*. To this end, we fused the cDNA fragments encoding the proteins of interest to the 5′ of the *GREEN FLUORESCENT PROTEIN* (*GFP*) cDNA driven by the CaMV

35S promoter, respectively, and the resulting constructs were transiently expressed in *N. benthamiana* leaves using the same amount of agrobacteria. A positive control, overexpression of *AtCERK1-GFP*, caused leaf chlorosis and tissue collapse in the entire infiltrated region 3 days after agrobacterial infiltration. Compared with *AtCERK1-GFP*, overexpression of *SlLYK13-GFP* resulted in similar levels of cell death (**Figure 4A**). Overexpression of *SlLYK1-GFP* did not result in obvious tissue collapse; however, dead tissues (dark blue in color) were visible after trypan blue staining (**Figure 4A**). This symptom of cell death was never observed on overexpression of *SlLYK12-GFP* (**Figure 4A**).

We then examined whether SlLYK1- and SlLYK13-induced cell death were dependent on their kinase activities, by generating constructs with inactive kinase domains; *SlLYK1* (K355E) and *SlLYK13* (K328E). Our results showed that SlLYK1 and SlLYK13 lacking kinase activity did not cause symptoms of cell death (**Figure 4B**). All constructs were verified by confocal microscopy observation of the green fluorescence signal and immunoblot analysis to determine protein size (**Figures 4C,D**). Since tissue collapse leads to degradation of target proteins, SlLYK1- and SlLYK13-GFP proteins could not be detected by either method; however, SlLYK12-, SlLYK1(K355E)-, and SlLYK13(K328E)-GFP all showed green fluorescence signals in the cell periphery as expected for proteins predicted to localize in the plasma membrane (**Figure 4C**). Taken together, our results suggest that SlLYK13 and SlLYK1 have redundant functions in cell death, while overexpression of *SlLYK13* could cause more severe symptoms of cell death.

SlLYK12 Is Subfunctionalized for AM Symbiosis in Tomato

To decipher the evolutionary relationships among LYK family proteins correlated with their functions in immunity and symbiosis, we constructed a phylogenetic tree by analyzing the protein sequences of CERK1 homologs from six Leguminosae species (*Glycine max*, *G. soja*, *Cajanus cajan*, *Lupinus angustifolius*, *M. truncatula*, and *L. japonicus*), three Solanaceae species (*Capsicum annuum*, *S. lycopersicum*, and *Solanum tuberosum*), three Cruciferae species (*A. thaliana*, *Brassica napus*, and *Brassica rapa*), and two Gramineae species (*Oryza sativa* and *Zea mays*). Finally, full length amino acid sequences encoded by 48 genes from 14 species were used to construct the phylogenetic tree (**Figure 5**).

Similar to other reported phylogenetic trees of LysM receptor proteins or kinases (Arrighi et al., 2006; Zhang et al., 2007, 2009; Lohmann et al., 2010; De Mita et al., 2014), CERK1 homologs
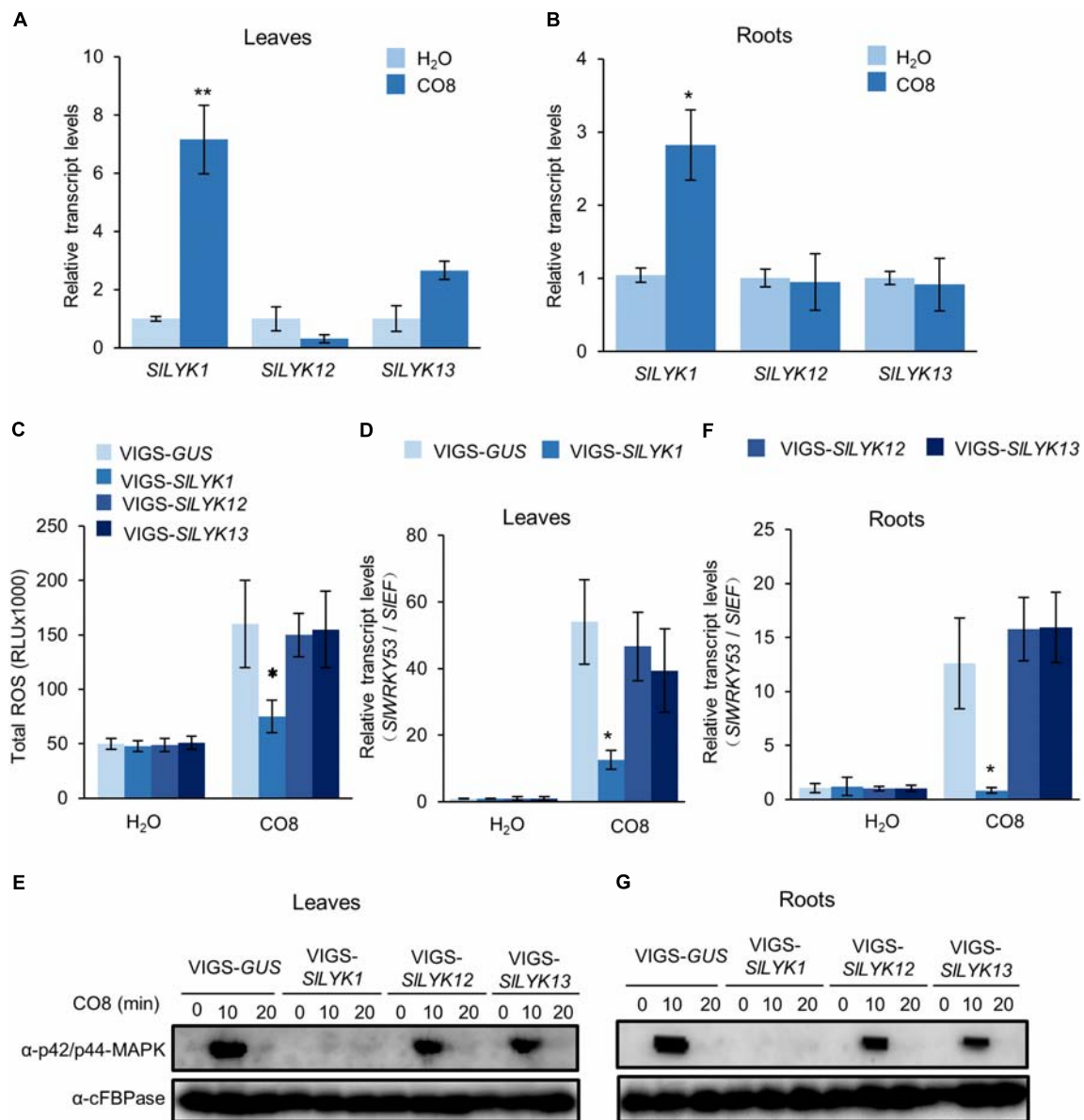
FIGURE 3 | SlLYK1 is required for chitin responses. **(A,B)** Relative transcript levels of *SlLYK1*, *SlLYK12*, and *SlLYK13* after chitooctaose (CO8) treatment in leaves **(A)** and roots **(B)**. RNA was extracted from 2-week-old wild type leaves and roots 30 min after CO8 treatment. Transcript levels were detected by qRT-PCR. CO8-induced immune responses were analyzed in leaves **(C–E)** and roots **(F,G)** 4 weeks after infiltration with VIGS-*SlLYK1*, *-12*, *-13*, and VIGS-*GUS*. **(C)** CO8-induced reactive oxygen species (ROS) accumulation. ROS was measured using a chemiluminescence assay. Signals were recorded for 30 min and ROS were quantified as the total amount of light emitted (RLU). Data are expressed as means ± SD (*n* = 8). **(D,F)** CO8-induced *SlWRKY53* (*Solyc08g008280*) gene expression in leaves **(D)** and roots **(F)**. RNA was extracted 30 min after CO8 treatment, and gene expression was detected by qRT-PCR. Asterisks indicate significant differences from the VIGS-*GUS* control (Student's *t*-test: *$P \leq 0.05$, **$P \leq 0.01$). **(E,G)** CO8-induced MAP kinase phosphorylation in leaves **(E)** and roots **(G)**. After CO8 treatment, MAP kinase phosphorylation was detected by immunoblot using the α-P42/P44 MAPK antibody and α-cFBPase (CYTOSOLIC FRUCTOSE-1,6-BISPHOSPHATASE) as a loading control. The experiment was repeated twice with similar results.

from monocotyledonous and dicotyledonous species were assigned to two different groups (**Figure 5**). Monocotyledonous maize and rice only have one or two CERK1 paralogs, whereas dicotyledonous species other than the Cruciferae have evolved several homologs, suggesting that the *CERK1* family has experienced duplication events in dicotyledonous species during their evolution. The CERK1 homologs in Solanaceae

and Leguminosae species are clustered into three clades: a Leguminosae-specific clade (blue), containing the Nod factor receptor NFR1/LYK3 and chitin receptor LjCERK6; a Solanaceae-specific clade (olive), including the immune receptors SlLYK1 and SlLYK13; and a mixed clade (red). SlLYK12, the potential Myc-factor receptor, was assigned to the mixed clade, which contains LYKs from both Solanaceae and Leguminosae species,
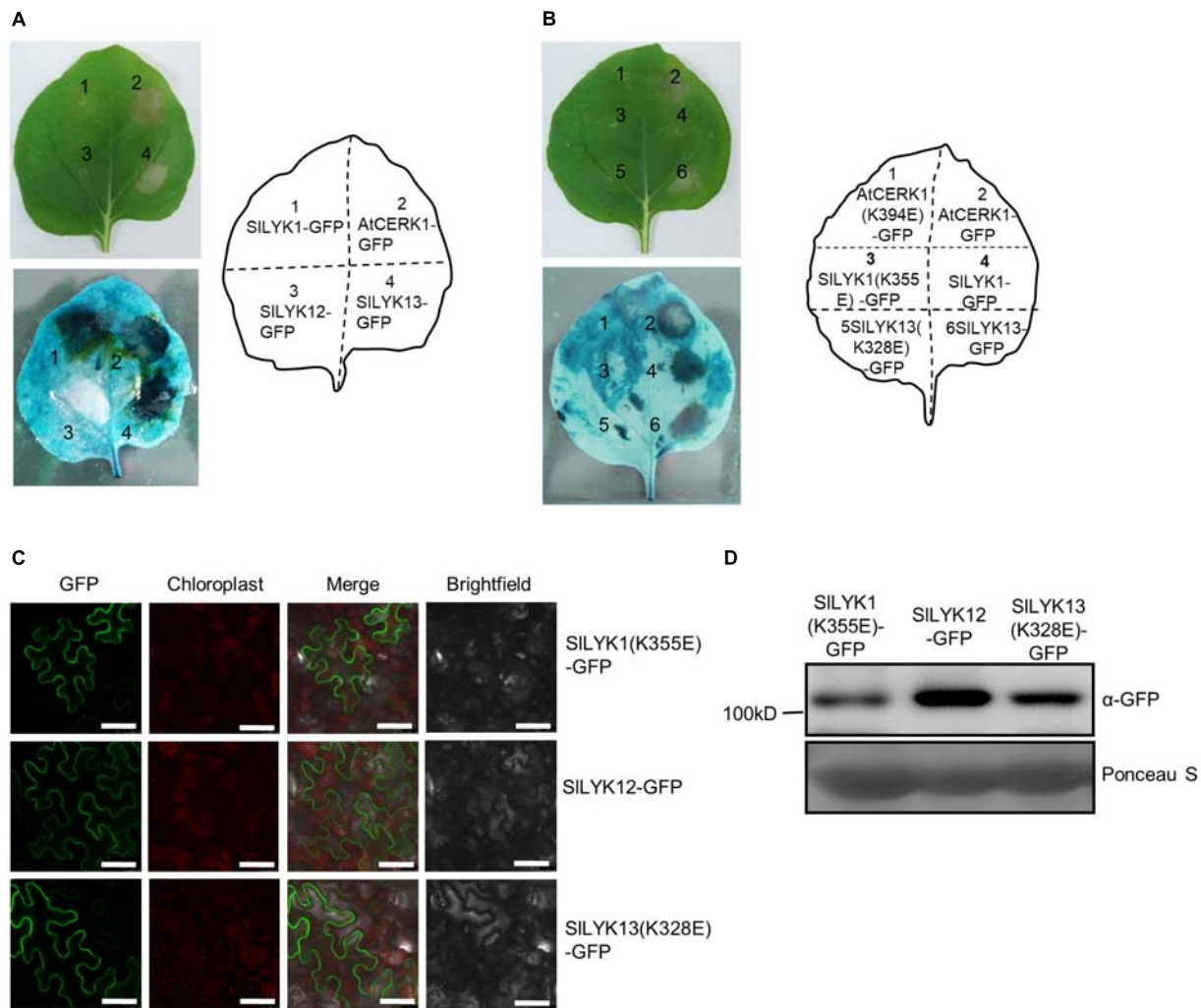
FIGURE 4 | Overexpression of *SlLYK1* and *SlLYK13* induces cell death in *Nicotiana benthamiana* leaves. **(A)** Detection of cell death in *N. benthamiana* leaves. C terminus of AtCERK1, SlLYK1, SlLYK12, and SlLYK13 proteins were fused to green fluorescence protein (GFP) and then transiently expressed in *N. benthamiana*. Images were taken before (above) and after trypan blue staining (below) 3 days after infiltration. **(B)** Kinase inactivation of SlLYK1(K355E), SlLYK13(K328E), and AtCERK1(K394E) abolished the symptoms of cell death. **(C)** GFP observation. Images were taken using confocal microscopy 3 days after infiltration. Bars, 50 μM. **(D)** Size and abundance of the SlLYK1(K355E)-GFP, SlLYK12-GFP, and SlLYK13(K328E)-GFP proteins. Total proteins were extracted from leaves expressing the *SlLYK1(K355E)-GFP*, *SlLYK12-GFP*, and *SlLYK13(K328E)-GFP* constructs. Immunoblot analysis was performed using anti-GFP antibody. Ponceau S staining was used as a protein loading control. The experiment was repeated twice with similar results.

suggesting that an ancestral gene duplication event occurred before the divergence of the Solanaceae and Leguminosae species. As OsCERK1 has a dual role in immunity and symbiosis, we hypothesize that an ancestor molecule in dicotyledons, which was responsible for both immunity and symbiosis, underwent gene duplication, leading to the sub-functionalization for a role in immunity and symbiosis, respectively.

DISCUSSION

In this study, we identified the orthologs of CERK1 in tomato, and investigated their function using a VIGS approach. Unlike rice OsCERK1, which has dual roles in chitin and AM symbiosis

(Miyata et al., 2014; Zhang et al., 2015), we found that no single tomato CERK1 ortholog is responsible for both functions; rather, SlLYK1 mainly affects chitin signaling, while SlLYK12 is required for AM symbiosis. Therefore, we hypothesize that a gene duplication event and functional divergence occurred in an ancient ancestor of tomato. This subfunctionalization could be because of tissue-specific expression patterns. SlLYK12 is mainly expressed in the roots where AM symbiosis is established, whereas SlLYK1 showed equal expression levels in both leaves and roots. In addition, the expression of these genes was specifically induced by AM symbiosis and chitin, respectively.

Functional studies of LysM-RLP and LysM-RLK suggest that these receptors can function as hetero-oligomers (Kelly et al., 2017; Zipfel and Oldroyd, 2017; Desaki et al., 2018); for
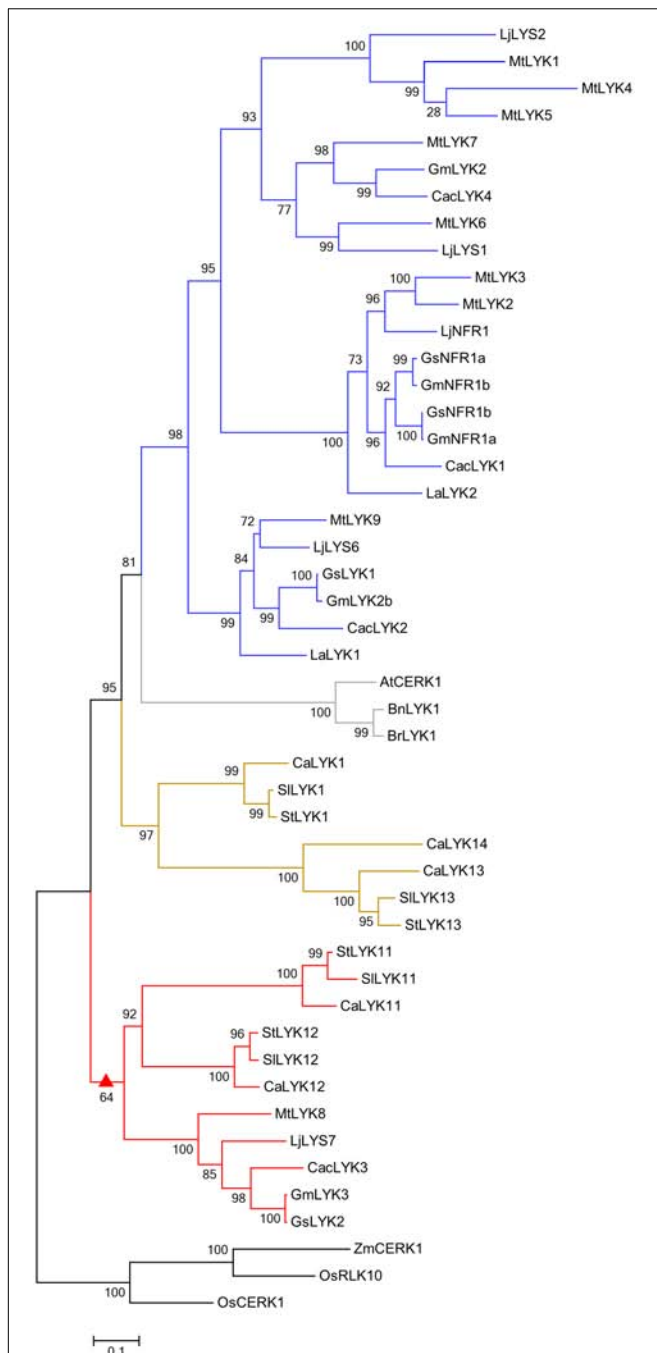
FIGURE 5 | Phylogenetic tree of CERK1 homologs. An unrooted phylogenetic tree of CERK1 protein homologs was constructed using the approximately maximum-likelihood method. Species of origin: At, *Arabidopsis thaliana*; Bn, *Brassica napus*; Br, *Brassica rapa*; Ca, *Capsicum annuum*; Cac, *Cajanus cajan*; Gm, *Glycine max*; Gs, *Glycine soja*; La, *Lupinus angustifolius*; Lj, *Lotus japonicus*; Mt, *Medicago truncatula*; Os, *Oryza sativa*; Sl, *Solanum lycopersicum*; St, *Solanum tuberosum*; Zm, *Zea mays*. Branches are labeled with their respective bootstrap values. The solid triangle represents the common ancestor sub-functionalized for a role in AM symbiosis, the blue color represents the Leguminosae-specific subclade, the gray color represents the Cruciferae-specific subclade, the olive color represents the Solanaceae-specific subclade, and the red color represents the mixed subclade containing members from Leguminosae and Solanaceae.

example, the hetero-oligomer of AtLYK5-AtCERK1 recognizes chitin in *Arabidopsis* (Cao et al., 2014), OsCEBiP-OsCERK1 recognizes chitin in rice (Shimizu et al., 2010; Hayafune et al., 2014), AtLYM1/3-AtCERK1 recognizes peptidoglycans in *Arabidopsis* (Willmann et al., 2011), OsLYP4/6-OsCERK1 recognizes peptidoglycan and chitin in rice (Liu et al., 2013; Ao et al., 2014), and LjNFR5-LjNFR1 recognizes Nod factor in *L. japonicus* (Broghammer et al., 2012). Therefore, it is very likely that SlLYK12 pairs with SlLYK10, the ortholog of NFR5 in tomato (Buendia et al., 2016). Plants with silenced SlLYK10 showed reduced AM colonization (Buendia et al., 2016). However, whether SlLYK10 associates with SlLYK12 awaits biochemical confirmation.

SlLYK1 and SlLYK13 were reported to have redundant roles in bacteria-mediated immune responses (Zeng et al., 2012). SlLYK1 (previously referred to as Bit9) and SlLYK13 both interact with the bacterial effector, avrPtoB, and plants with silenced *SlLYK1*, *SlLYK11*, *SlLYK12*, and *SlLYK13* were more susceptible to *Pseudomonas syringae* (Zeng et al., 2012). In this study, we found that ectopic expression of *SlLYK1* and *SlLYK13* in *N. benthamiana* induced cell death; however, SlLYK13 could trigger a stronger cell death response than SlLYK1. In addition, plants with silenced *SlLYK13* showed unaltered responses to chitin, suggesting that its involvement in cell death may be separate from chitin signaling. Indeed, in *Arabidopsis*, the cell death phenotype mediated by AtCERK1(L124F) is completely independent of chitin signaling (Petutschnig et al., 2014), but ectopic expression of *AtCERK1* in *N. benthamiana* strongly induces cell death. Therefore, tomato SlLYK1 and SlLYK13 may have undergone sub-functionalization after gene duplication during evolution.

CERK1 homologs in Solanaceae and Leguminosae species are clustered into three clades: a legume-specific clade (NFR1-MtLYK3-LjCERK6), a Solanaceae-specific clade (SlLYK1), and a mixed clade (SlLYK12-LjLYS7-MtLYK8). Consistent with the function of SlLYK12 in AM symbiosis, the SlLYK12-LjLYS7-MtLYK8 clade does not contain genes from *Lupinus angustifolius*, a legume species which cannot form AM symbiosis, but can establish a symbiotic relationship with rhizobia (Oba et al., 2001; Schulze et al., 2006), suggesting that the SlLYK12-LjLYS7-MtLYK8 clade may mediate the host specificity of AM symbiosis. In this clade, LjLYS7 and MtLYK8 are the closest orthologs of SlLYK12 in *L. japonicus* and *M. truncatula*, which were also predicted to be involved in symbiotic perception in endomycorrhizae in other phylogenetic analysis (De Mita et al., 2014). Given the common pathways used for AM symbiosis and legume-rhizobium symbiosis, it has been hypothesized that Nod factor receptors evolved from a Myc factor receptor (Lohmann et al., 2010; De Mita et al., 2014). According to this theory, it is reasonable to predict that the putative Myc factor receptor should be clustered with NFR1 in a single group. Surprisingly, we found that the NFR1 clade is more closely related to the SlLYK1-immunity clade than to the SlLYK12-AM symbiosis clade. Consistent with this notion, LjCERK6, the closest paralog of NFR1, was recently identified as involved in chitin recognition in *L. japonicus*, but not AM symbiosis (Bozsoki et al., 2017). All these studies suggest that NFR1 might not have evolved directly from a receptor for AM symbiosis, rather, it may

have evolved from an ancestor with a dual function, which underwent gene duplication in legumes and the paralogous gene underwent neofunctionalization to become a Nod factor receptor. Alternatively, it is interesting to note that the YAQ motif that is proposed to be associated with a role in symbiosis in the CERK1 family is conserved in SlLYK1 and SlLYK12 (Nakagawa et al., 2011; De Mita et al., 2014), so it is possible that SlLYK1 might have a symbiotic role. However, this role might not have been observed because of redundancy with SlLYK12 or because of incomplete silencing. Indeed, PsLYK9, the ortholog of LjCERK6 in *Pisum sativum*, is required for plant immunity and could be involved in Myc factor perception (Leppyanen et al., 2018). Therefore, the evolutionary origin of Nod factor receptors awaits future experimental investigation.

AUTHOR CONTRIBUTIONS

YL designed the research and wrote the manuscript. DL and XS performed most of the experiments. NW contributed to analysis of gene expression. FS designed primers for VIGS constructs and provided technique support for VIGS approach.

REFERENCES

Ao, Y., Li, Z., Feng, D., Xiong, F., Liu, J., Li, J. F., et al. (2014). OsCERK1 and OsRLCK176 play important roles in peptidoglycan and chitin signaling in rice innate immunity. *Plant J.* 80, 1072–1084. doi: 10.1111/tpj.12710

Arrighi, J. F., Barre, A., Ben Amor, B., Bersoult, A., Soriano, L. C., Mirabella, R., et al. (2006). The *Medicago truncatula* lysin motif-receptor-like kinase gene family includes NFP and new nodule-expressed genes. *Plant Physiol.* 142, 265–279. doi: 10.1104/pp.106.084657

Ben Amor, B., Shaw, S. L., Oldroyd, G. E., Maillet, F., Penmetsa, R. V., Cook, D., et al. (2003). The NFP locus of *Medicago truncatula* controls an early step of Nod factor signal transduction upstream of a rapid calcium flux and root hair deformation. *Plant J.* 34, 495–506. doi: 10.1046/j.1365-313X.2003. 01743.x

Boller, T., and Felix, G. (2009). A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annu. Rev. Plant Biol.* 60, 379–406. doi: 10.1146/annurev.arplant.57. 032905.105346

Bozsoki, Z., Cheng, J., Feng, F., Gysel, K., Vinther, M., Andersen, K. R., et al. (2017). Receptor-mediated chitin perception in legume roots is functionally separable from Nod factor perception. *Proc. Natl. Acad. Sci. U.S.A.* 114, E8118–E8127. doi: 10.1073/pnas.1706795114

Broghammer, A., Krusell, L., Blaise, M., Sauer, J., Sullivan, J. T., Maolanon, N., et al. (2012). Legume receptors perceive the rhizobial lipochitin oligosaccharide signal molecules by direct binding. *Proc. Natl. Acad. Sci. U.S.A.* 109, 13859–13864. doi: 10.1073/pnas.1205171109

Buendia, L., Wang, T., Girardin, A., and Lefebvre, B. (2016). The LysM receptor-like kinase SlLYK10 regulates the arbuscular mycorrhizal symbiosis in tomato. *New Phytol.* 210, 184–196. doi: 10.1111/nph.13753

Cao, Y., Liang, Y., Tanaka, K., Nguyen, C. T., Jedrzejczak, R. P., Joachimiak, A., et al. (2014). The kinase LYK5 is a major chitin receptor in *Arabidopsis* and forms a chitin-induced complex with related kinase CERK1. *eLife* 3:e03766. doi: 10.7554/eLife.03766

Carotenuto, G., Chabaud, M., Miyata, K., Capozzi, M., Takeda, N., Kaku, H., et al. (2017). The rice LysM receptor-like kinase OsCERK1 is required for the perception of short-chain chitin oligomers in arbuscular mycorrhizal signaling. *New Phytol.* 214, 1440–1446. doi: 10.1111/nph.14539

De Mita, S., Streng, A., Bisseling, T., and Geurts, R. (2014). Evolution of a symbiotic receptor through gene duplications in the legume-rhizobium mutualism. *New Phytol.* 201, 961–972. doi: 10.1111/nph.12549

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2018.01004/full#supplementary-material

Desaki, Y., Miyata, K., Suzuki, M., Shibuya, N., and Kaku, H. (2018). Plant immunity and symbiosis signaling mediated by LysM receptors. *Innate Immun.* 24, 92–100. doi: 10.1177/1753425917738885

Dodds, P. N., and Rathjen, J. P. (2010). Plant immunity: towards an integrated view of plant-pathogen interactions. *Nat. Rev. Genet.* 11, 539–548. doi: 10.1038/ nrg2812

Genre, A., Chabaud, M., Balzergue, C., Puech-Pages, V., Novero, M., Rey, T., et al. (2013). Short-chain chitin oligomers from arbuscular mycorrhizal fungi trigger nuclear Ca2+ spiking in *Medicago truncatula* roots and their production is enhanced by strigolactone. *New Phytol.* 198, 190–202. doi: 10.1111/nph.12146

Harrison, M. J. (2012). Cellular programs for arbuscular mycorrhizal symbiosis. *Curr. Opin. Plant Biol.* 15, 691–698. doi: 10.1016/j.pbi.2012.08.010

Hayafune, M., Berisio, R., Marchetti, R., Silipo, A., Kayama, M., Desaki, Y., et al. (2014). Chitin-induced activation of immune signaling by the rice receptor CEBiP relies on a unique sandwich-type dimerization. *Proc. Natl. Acad. Sci. U.S.A.* 111, E404–E413. doi: 10.1073/pnas.1312099111

Kelly, S., Radutoiu, S., and Stougaard, J. (2017). Legume LysM receptors mediate symbiotic and pathogenic signalling. *Curr. Opin. Plant Biol.* 39, 152–158. doi: 10.1016/j.pbi.2017.06.013

Leppyanen, I. V., Shakhnazarova, V. Y., Shtark, O. Y., Vishnevskaya, N. A., Tikhonovich, I. A., and Dolgikh, E. A. (2018). Receptor-Like Kinase LYK9 in Pisum sativum L. is the CERK1-Like Receptor that controls both plant immunity and AM symbiosis development. *Int. J. Mol. Sci.* 19:E8. doi: 10.3390/ ijms19010008

Liang, Y., Toth, K., Cao, Y., Tanaka, K., Espinoza, C., and Stacey, G. (2014). Lipochitooligosaccharide recognition: an ancient story. *New Phytol.* 204, 289–296. doi: 10.1111/nph.12898

Liao, D., Cao, Y., Sun, X., Espinoza, C., Nguyen, C. T., Liang, Y., et al. (2017). *Arabidopsis* E3 ubiquitin ligase PLANT U-BOX13 (PUB13) regulates chitin receptor LYSIN MOTIF RECEPTOR KINASE5 (LYK5) protein abundance. *New Phytol.* 214, 1646–1656. doi: 10.1111/nph.14472

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. doi: 10.1126/science.1090074

Liu, B., Li, J. F., Ao, Y., Li, Z., Liu, J., Feng, D., et al. (2013). OsLYP4 and OsLYP6 play critical roles in rice defense signal transduction. *Plant Signal. Behav.* 8:e22980. doi: 10.4161/psb.22980

Liu, J., Maldonado-Mendoza, I., Lopez-Meyer, M., Cheung, F., Town, C. D., and Harrison, M. J. (2007). Arbuscular mycorrhizal symbiosis is accompanied by local and systemic alterations in gene expression and an increase in disease

resistance in the shoots. *Plant J.* 50, 529–544. doi: 10.1111/j.1365-313X.2007.03069.x

Lohmann, G. V., Shimoda, Y., Nielsen, M. W., Jorgensen, F. G., Grossmann, C., Sandal, N., et al. (2010). Evolution and regulation of the *Lotus japonicus* LysM receptor gene family. *Mol. Plant Microbe Interact.* 23, 510–521. doi: 10.1094/MPMI-23-4-0510

Madsen, E. B., Madsen, L. H., Radutoiu, S., Olbryt, M., Rakwalska, M., Szczyglowski, K., et al. (2003). A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals. *Nature* 425, 637–640. doi: 10.1038/nature02045

Maillet, F., Poinsot, V., Andre, O., Puech-Pages, V., Haouy, A., Gueunier, M., et al. (2011). Fungal lipochitooligosaccharide symbiotic signals in arbuscular mycorrhiza. *Nature* 469, 58–63. doi: 10.1038/nature09622

Miyata, K., Hayafune, M., Kobae, Y., Kaku, H., Nishizawa, Y., Masuda, Y., et al. (2016). Evaluation of the role of the LysM receptor-like kinase, OsNFR5/OsRLK2 for AM symbiosis in rice. *Plant Cell Physiol.* 57, 2283–2290. doi: 10.1093/pcp/pcw144

Miyata, K., Kozaki, T., Kouzai, Y., Ozawa, K., Ishii, K., Asamizu, E., et al. (2014). The bifunctional plant receptor, OsCERK1, regulates both chitin-triggered immunity and arbuscular mycorrhizal symbiosis in rice. *Plant Cell Physiol.* 55, 1864–1872. doi: 10.1093/pcp/pcu129

Nakagawa, T., Kaku, H., Shimoda, Y., Sugiyama, A., Shimamura, M., Takanashi, K., et al. (2011). From defense to symbiosis: limited alterations in the kinase domain of LysM receptor-like kinases are crucial for evolution of legume-Rhizobium symbiosis. *Plant J.* 65, 169–180. doi: 10.1111/j.1365-313X.2010.04411.x

Nouri, E., Breuillin-Sessoms, F., Feller, U., and Reinhardt, D. (2014). Phosphorus and nitrogen regulate arbuscular mycorrhizal symbiosis in *Petunia hybrida*. *PLoS ONE* 9:e90841. doi: 10.1371/journal.pone.0090841

Oba, H., Tawaraya, K., and Wagatsuma, T. (2001). Arbuscular mycorrhizal colonization in *Lupinus* and related genera. *Soil Sci. Plant Nutr.* 47, 685–694. doi: 10.1080/00380768.2001.10408433

Oldroyd, G. E. (2013). Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nat. Rev. Microbiol.* 11, 252–263. doi: 10.1038/nrmicro2990

Op den Camp, R., Streng, A., De Mita, S., Cao, Q., Polone, E., Liu, W., et al. (2011). LysM-type mycorrhizal receptor recruited for rhizobium symbiosis in nonlegume Parasponia. *Science* 331, 909–912. doi: 10.1126/science.1198181

Petutschnig, E. K., Stolze, M., Lipka, U., Kopischke, M., Horlacher, J., Valerius, O., et al. (2014). A novel *Arabidopsis* CHITIN ELICITOR RECEPTOR KINASE 1 (CERK1) mutant with enhanced pathogen-induced cell death and altered receptor processing. *New Phytol.* 204, 955–967. doi: 10.1111/nph.12920

Pietraszewska-Bogiel, A., Lefebvre, B., Koini, M. A., Klaus-Heisen, D., Takken, F. L., Geurts, R., et al. (2013). Interaction of *Medicago truncatula* lysin motif receptor-like kinases, NFP and LYK3, produced in *Nicotiana benthamiana* induces defence-like responses. *PLoS One* 8:e65055. doi: 10.1371/journal.pone.0065055

Radutoiu, S., Madsen, L. H., Madsen, E. B., Felle, H. H., Umehara, Y., Gronlund, M., et al. (2003). Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *Nature* 425, 585–592. doi: 10.1038/nature02039

Schulze, J., Temple, G., Temple, S. J., Beschow, H., and Vance, C. P. (2006). Nitrogen fixation by white lupin under phosphorus deficiency. *Ann. Bot.* 98, 731–740. doi: 10.1093/aob/mcl154

Schussler, A., Schwarzott, D., and Walker, C. (2001). A new fungal phylum, the Glomeromycota: phylogeny and evolution. *Mycol. Res.* 105, 1413–1421. doi: 10.1017/S0953756201005196

Shimizu, T., Nakano, T., Takamizawa, D., Desaki, Y., Ishii-Minami, N., Nishizawa, Y., et al. (2010). Two LysM receptor molecules, CEBiP and OsCERK1, cooperatively regulate chitin elicitor signaling in rice. *Plant J.* 64, 204–214. doi: 10.1111/j.1365-313X.2010.04324.x

Tomato Genome, C. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485, 635–641. doi: 10.1038/nature11119

Vierheilig, H., Coughlan, A. P., Wyss, U., and Piche, Y. (1998). Ink and vinegar, a simple staining technique for arbuscular-mycorrhizal fungi. *Appl. Environ. Microbiol.* 64, 5004–5007.

Willmann, R., Lajunen, H. M., Erbs, G., Newman, M. A., Kolb, D., Tsuda, K., et al. (2011). *Arabidopsis* lysin-motif proteins LYM1 LYM3 CERK1 mediate bacterial peptidoglycan sensing and immunity to bacterial infection. *Proc. Natl. Acad. Sci. U.S.A.* 108, 19824–19829. doi: 10.1073/pnas.1112862108

Zeng, L., Velasquez, A. C., Munkvold, K. R., Zhang, J., and Martin, G. B. (2012). A tomato LysM receptor-like kinase promotes immunity and its kinase activity is inhibited by AvrPtoB. *Plant J.* 69, 92–103. doi: 10.1111/j.1365-313X.2011.04773.x

Zhang, X., Dong, W., Sun, J., Feng, F., Deng, Y., He, Z., et al. (2015). The receptor kinase CERK1 has dual functions in symbiosis and immunity signalling. *Plant J.* 81, 258–267. doi: 10.1111/tpj.12723

Zhang, X. C., Cannon, S. B., and Stacey, G. (2009). Evolutionary genomics of LysM genes in land plants. *BMC Evol. Biol.* 9:183. doi: 10.1186/1471-2148-9-183

Zhang, X. C., Wu, X., Findley, S., Wan, J., Libault, M., Nguyen, H. T., et al. (2007). Molecular evolution of lysin motif-type receptor-like kinases in plants. *Plant Physiol.* 144, 623–636. doi: 10.1104/pp.107.097097

Zipfel, C., and Oldroyd, G. E. (2017). Plant signalling in symbiosis and immunity. *Nature* 543, 328–336. doi: 10.1038/nature22009

Check for updates

Impact of Plant Peptides on Symbiotic Nodule Development and Functioning

Attila Kereszt[1], Peter Mergaert[2], Jesús Montiel[1†], Gabriella Endre[1] and Éva Kondorosi[1*]

[1] Institute of Plant Biology, Biological Research Centre, Hungarian Academy of Sciences, Szeged, Hungary, [2] Institute of Integrative Biology of the Cell, UMR 9198, CNRS – CEA – Université Paris-Sud, Gif-sur-Yvette, France

Ribosomally synthesized peptides have wide ranges of functions in plants being, for example, signal molecules, transporters, alkaloids, or antimicrobial agents. Legumes are an unprecedented rich source of peptides, which are used to control the symbiosis of these plants with the nitrogen-fixing *Rhizobium* bacteria. Here, we discuss the function and the evolution of these peptides playing an important role in the formation or functioning of the symbiotic organs, the root nodules. We distinguish peptides that can be either cell-autonomous or secreted short-range or long-range signals, carrying messages in or between plant cells or that can act as effectors interacting with the symbiotic bacteria. Peptides are further classified according to the stage of the symbiotic process where they act. Several peptide classes, including RALF, DLV, ENOD40, and others, control *Rhizobium* infection and the initiation of cell divisions and the formation of nodule primordia. CLE and CEP peptides are implicated in systemic and local control of nodule initiation during autoregulation of nodulation and in response to the nutritional demands of the plant. Still other peptides act at later stages of the symbiosis. The PSK peptide is thought to be involved in the suppression of immunity in nodules and the nodule-specific cysteine-rich, GRP, and SNARP (LEED..PEED) peptide families are essential in the functioning of the nitrogen fixing root nodules. The NCRs and possibly also the GRP and SNARPs are targeted to the endosymbionts and play essential roles in the terminal differentiation of these bacteria.

Keywords: legume-rhizobium symbiosis, nodule development, signaling peptides, NCR, CLE, CEP, GRP

INTRODUCTION

Ribosomally synthesized peptides with biological functions are arbitrarily (and loosely) defined as gene-encoded small proteins of 2 to about 100 amino acids. Research on peptide-mediated signaling processes and other peptide functions in plants has gained momentum in the last decade [for a review, see Tavormina et al. (2015)]. Crucial importance of peptides has been demonstrated in embryogenesis (Costa et al., 2014), fertilization (Okuda et al., 2009; Higashiyama, 2010; Mecchia et al., 2017; Ge et al., 2017), cell expansion (Haruta et al., 2014; Murphy and De Smet, 2014), cell differentiation (Butenko et al., 2003; Hunt et al., 2009; Matsuzaki et al., 2010; Sugano et al., 2010; Lee et al., 2015; Santiago et al., 2016; Doblas et al., 2017; Nakayama et al., 2017), immunity (Constabel et al., 1995; Pearce et al., 2001; Huffaker et al., 2006; Hou et al., 2014; Stegmann et al., 2017), nutrition (Tabata et al., 2014; Ohkubo et al., 2017), as well as in other processes

(Whitford et al., 2012; Takahashi et al., 2018). The field of *Rhizobium*-legume symbiosis is not lagging behind when it comes to discoveries of peptides with key roles in the nodulation process (Djordjevic et al., 2015). In this review, we summarize these peptide signals and peptide effectors and the present knowledge on their identified or predicted functions in symbiosis (**Figure 1**).

In response to nitrogen starvation, plants from the Leguminosae family can establish symbiosis with their *Rhizobium* partners resulting in the development of root nodules and within the nodule plant cells, the conversion of the bacteria into nitrogen fixing bacteroids. For the initiation of the symbiosis and finding the appropriate *Rhizobium* bacterium in the soil, legume plants excrete from their roots flavonoids and isoflavonoids acting as inducers of nodulation genes in their symbiont. Activation of nodulation genes leads to the production of bacterial signal molecules, the Nod factors, which induce nodule organogenesis in the host plant and are required for the infection process as well. Already at this stage, several plant peptides affect the plant susceptibility to infection and nodulation and participate in the regulation of these first developmental and differentiation steps. Moreover, the plant invests only in the number of nodules required to satisfy its nitrogen needs. The nodule numbers are negatively controlled by autoregulation of nodulation and by high nitrate using CLV3/ESR-related (CLE) peptides exerting a systemic negative regulation on nodulation *via* root- and shoot-derived signaling. Members of the C-terminally encoded peptide (CEP) family, on the contrary, exert a positive effect on nodulation in response to low nitrogen availability.

While these peptide signals have general and conserved roles and some of their members were recruited to serve in the
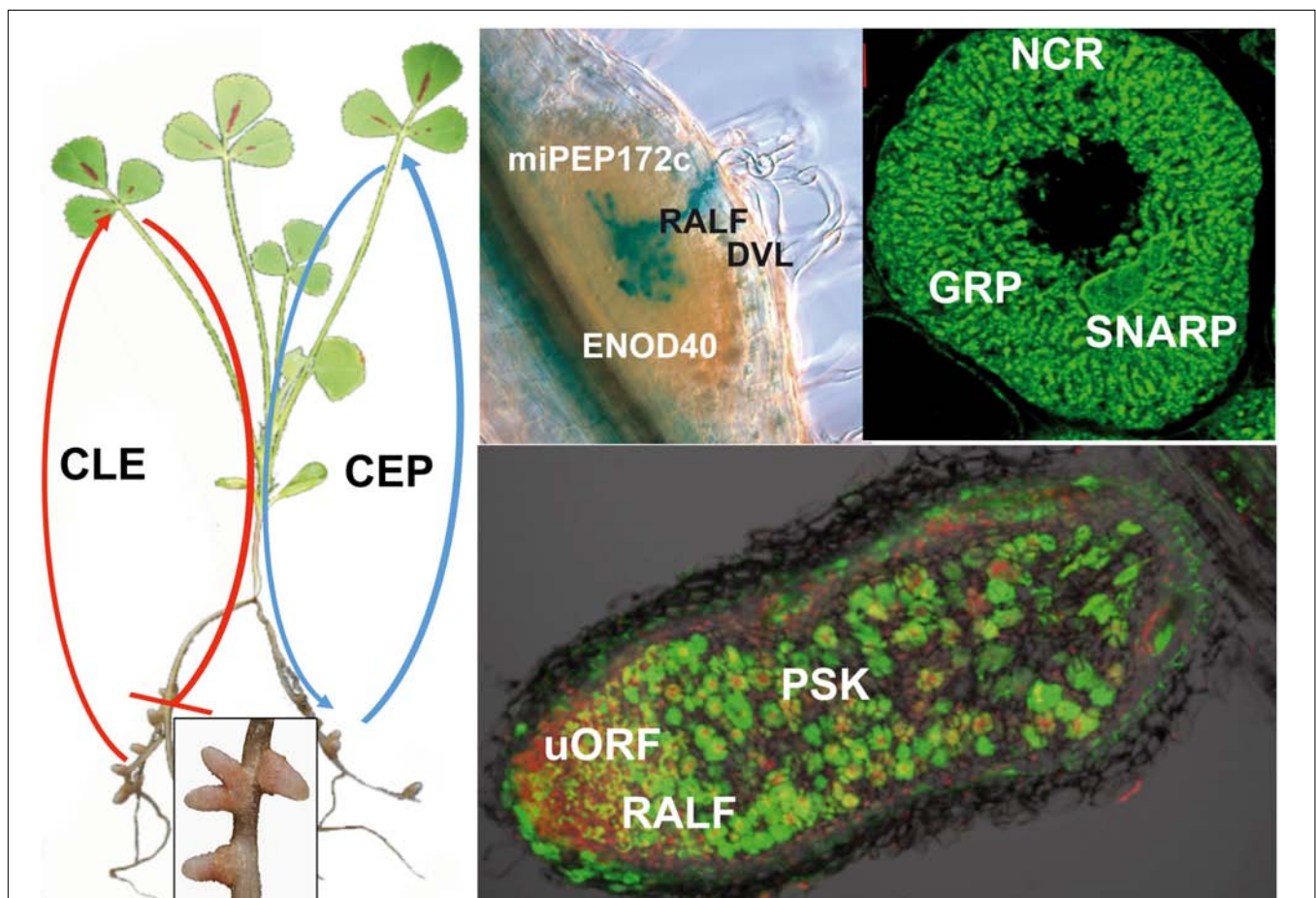
FIGURE 1 | Peptides contributing to nodule formation. The left part of the figure presents a nodulated *M. truncatula* plant. The inset shows an enlarged image of nodules. The upper pictures show a section of a nodule primordium showing a network of infection threads stained in blue **(Left)**, and a symbiotic nodule cell densely packed with differentiated bacteroids **(Right)**. The lower image is a section of a mature nodule showing the rhizobia in green. The red staining shows plant cell nuclei, highlighting primarily the nodule meristem. Peptides involved in nodulation are indicated at their presumed site of action. CLE peptides are produced in response to already initiated nodules or by high nitrate. They are systemic signals received in the shoot by the HAR1/SUNN/NARK receptor-like kinases, which in return produce a shoot-derived signal that inhibits further nodulation in the roots. CEP peptides on the other hand are produced in the roots and their production is enhanced by low nitrogen. The peptides are perceived by the CRA2 receptor-like kinase and stimulate nodulation through a systemic mechanism. The RALF and DLV peptides negatively affect the infection process in early stages of nodule development. During later stages, RALF has a negative, while PSK has positive effect on the process. ENOD40 and miPEP172c affect nodule primordium formation. uORFp1 participates in the control of the meristem maintenance. NCR peptides and potentially also the GRP and SNARP peptides control bacteroid differentiation and functioning.

nitrogen fixing symbiosis as well, other large secreted symbiotic peptide families, which evolved specifically in certain legume lineages, are linked to irreversible, or terminal, differentiation of the bacteroids. These bacteroids are not able to switch from the symbiotic state to free living one and are thus non-cultivable. They are also morphologically different from the free living cells, often exhibiting remarkable cell growth and change in cell shape, having an amplified genome and altered cell envelope with increased membrane permeability. Such terminal bacteroid differentiation occurs in several but not all branches of the Leguminosae family and, so far, has only been studied in legumes of the Inverted Repeat Lacking Clade (IRLC) from the Papilionoideae subfamily and in the phylogenetically distant *Aeschynomene* genus of Dalbergoid legumes. In these legumes, these peptides direct terminal bacteroid differentiation, which is indispensable for nitrogen fixation. In other legumes, where the fate of the bacteroids is reversible and endosymbionts can return to the free-living life, there is no change in morphology, size, DNA content, and membrane properties of the bacteria, and these legumes lack genes coding for the abovementioned symbiotic peptide families in their genome. The large majority of these peptides are nodule specific cysteine-rich (NCR) peptides, but there are also glycine-rich peptides (GRPs) as well as small nodulin acidic RNA-binding peptides (SNARPs or also named LEED..PEED according to a conserved amino acid motif).

PLANT PEPTIDES INVOLVED IN INFECTION AND NODULE ORGANOGENESIS

Regulators Encoded by Peptide-Coding Genes

Rapid Alkanization Factor (RALF) Family

The *Medicago truncatula MtRALFL1* gene was identified in a transcriptome screen for early Nod factor-induced genes using a double supernodulating mutant line (Combier et al., 2008b). The gene encodes a member of the RALF cysteine-rich and secreted peptide family, which comprises 15 members in *M. truncatula* (Silverstein et al., 2007; de Bang et al., 2017). Peptides of the RALF family are known in plants to control immunity as well as cell expansion, notably in root hair and pollen tube growth (Murphy and De Smet, 2014; Haruta et al., 2014; Ge et al., 2017; Mecchia et al., 2017; Stegmann et al., 2017). The secretion of RALF peptides indicates that they function in cell-to-cell signaling what they indeed do through the interaction with cell membrane located receptors like the receptor-like kinase FERONIA. Notably, FERONIA and other RALF-binding receptors are related receptor-like kinases that have an ectodomain (ligand-binding extracellular domain of the receptor) composed of two malectin-like domains (Li et al., 2016). The RALF peptides are bound by the malectin-containing ectodomains but the molecular details of the interaction need to be further clarified.

The *MtRALFL1* gene is induced by Nod factor treatment of *M. truncatula* roots although this induction was only observed

in the particular genetic background of the supernodulating *sunn-2 sickle* double mutant (Combier et al., 2008b). This *M. truncatula* line has a higher sensitivity to Nod factors and, therefore, allowed to detect responses that are not visible in a wild-type background. The involvement of the MtRALFL1 peptide in nodulation was further supported by overexpression of *MtRALFL1* in *M. truncatula* transgenic roots, which resulted in a drastic reduction of nodule number and an abnormally high number of aborted infection threads. Moreover, the few nodules initiated on the transgenic roots did not develop into mature, nitrogen fixing organs. Thus, MtRALFL1 controls infection thread formation and possibly other stages of nodule development (Combier et al., 2008b). Interestingly, infection threads have a polar growth mode similar to root hairs and pollen tubes whose growth is also affected by RALF peptides in other plant species. Furthermore, the nodulation receptor-like kinase NORK [also known as "doesn't make infections 2" (DMI2) or "symbiosis receptor-like kinase" (SymRK)], which is part of the Nod factor receptor complex regulating infection thread formation, contains in its ectodomain a malectin-like domain (Endre et al., 2002; Antolín-Llovera et al., 2014). Although the overall ectodomain structure of NORK differs from the ectodomain of FERONIA and related receptor-like kinases by the presence of an additional leucine-rich repeat domain, it is tempting to speculate that the MtRALFL1 peptide targets the Nod factor receptor complex.

Medicago DEVIL (MtDVL1) Non-secretory Peptide

Another characterized regulatory peptide in *M. truncatula* is called MtDVL1. The *MtDVL1* gene was identified in the same transcriptome screen as the above described *MtRALFL1* gene (Combier et al., 2008b). MtDVL1 is homologous to the family of ROTUNDIFOLIA FOUR (ROT4) or DEVIL (DVL) peptides in Arabidopsis, which are conserved in plants (Guo et al., 2015). These peptides are non-secretory and are thought to function cell-autonomously (Ikeuchi et al., 2011). Understanding the biological function and mode of action of this family of peptides is limited. Loss-of-function or knock-down mutants show no noticeable phenotypes, likely due to gene redundancy in the family, but overexpression of several members of the family in Arabidopsis produces phenotypes suggesting the implication of DVL peptides in plant development (Narita et al., 2004; Wen et al., 2004; Ikeuchi et al., 2011; Valdivia et al., 2012; Guo et al., 2015). The function of the *MtDVL1* gene in symbiosis was also assessed by overexpression studies in transgenic *M. truncatula* roots, which led to a strong increase in the number of abortive infections in the root cortex and in line with this, a significant reduction of nodule formation suggesting that the MtDVL1 peptide has a negative regulatory role in nodulation and particularly in rhizobial infection (Combier et al., 2008b).

Phytosulfokine (PSK) Peptides

Phytosulfokines are five-amino acid peptides containing two sulfated tyrosine residues. These peptides are produced as preproproteins, which are secreted as sulfated precursors in the cell apoplast where they are further processed to the PSK peptide by a subtilisin serine protease (Srivastava et al., 2008;

Komori et al., 2009). The PSK peptides are recognized by a receptor-like kinase and act primarily as growth-promoting factors (Matsubayashi et al., 2002) but also participate in the immune response of plants by either attenuating pattern-triggered immunity against biotrophs or promoting immunity against necrotrophic pathogens (Igarashi et al., 2012; Zhang et al., 2018). In *Lotus japonicus*, five *PSK* genes have been identified, two of which showed a nodule-specific expression found mainly in the rhizobium-infected symbiotic cells (Wang et al., 2015). Overexpression of one of the *PSK* genes in transgenic *L. japonicus* roots but also the external application of the PSK-α peptide to roots enhanced nodulation. *PSK* overexpression did not increase infection events or the number of initiated nodule primordia and, therefore, the enhanced nodulation was attributed to a stimulation of nodule growth from primordia by PSK. Moreover, PSK overexpression resulted in the downregulation of the jasmonate signal transduction pathway. Thus, the nodule-specific PSK peptides might be also important in nodules to suppress host defense responses against the rhizobia (Wang et al., 2015).

Short Peptides Encoded by sORFs (sPEPs)

The peptides described so far are derived from genes whose major open reading frame (ORF) encodes the peptide or a peptide precursor. Recently, an additional source of peptides was recognized in short ORFs (sORFs) located in RNA molecules, which have another primary function (Andrews and Rothnagel, 2014; Hellens et al., 2016; Makarewich and Olson, 2017). These RNAs can be transcripts previously annotated as long non-coding RNAs (lncRNAs), primary microRNA transcripts (pri-miRNAs), or protein-encoding mRNAs. The translation of some of these sORFs has been demonstrated experimentally by translational GUS fusions, ribosome profiling, overexpression or mutational analysis of the transcript *in planta*, or by the immunological detection of the peptides. The biological activity of some sPEPs has been demonstrated by *in planta* application or *in vitro* biochemical activities. Examples of sORFs/sPEPs have been identified as regulators of nodulation.

Peptides Encoded by sORFs in pri-miRNAs (miPEPs)

A class of newly identified peptide regulators, discovered thus far in plants only, is encoded by microRNA (miRNA) genes. The miRNAs are 21–24 nt regulatory RNA molecules and function *via* base-pairing with complementary sequences in target mRNAs, mediating cleavage or inhibition of translation of the target. The miRNAs are transcribed as large pri-miRNAs, which are then processed into mature miRNAs. Based on case studies in Arabidopsis and *M. truncatula*, certain pri-miRNAs were reported to contain in their 5′ part functional sORFs encoding the so-called miPEPs (Lauressergues et al., 2015). Evidence for the production of the Arabidopsis miPEP165a and the *M. truncatula* miPEP171b was obtained by translational fusions with GUS and by western blots and immunolocalization with specific antibodies produced against the corresponding synthetic peptides. It was further shown by overexpression of the corresponding sORFs or external application of synthetic peptides to plants that miPEPs enhance specifically the transcription of their cognate primary

transcript. They thereby form a positive feedback loop and increase the level of the corresponding miRNA, amplify the original effect, and reduce even more the expression of the miRNA target genes. Although miPEPs have been experimentally characterized only in a few cases, a survey of the sequences of plant pri-miRNAs indicates that they generally contain sORFs suggesting that miPEPs are commonly encoded by pri-miRNAs.

A miPEP was recently shown to control nodulation in soybean (Couzigou et al., 2016). Several miRNAs are known to regulate different stages of the nodulation process [reviewed in Couzigou and Combier (2016)]. One of these miRNAs is the soybean miR172c, which targets *NNC1* gene coding for an APETALA 2 transcription factor. The NNC1 transcription factor, which regulates the nodule-specific gene *ENOD40* (see below), negatively affects nodulation, and thus the miR172c expression stimulates nodulation by reducing NNC1 activity (Wang et al., 2014). Similarly as the earlier characterized pri-miRNAs of *M. truncatula* and Arabidopsis, the primary transcript of soybean miR172c encodes the miPEP172c. Intriguingly, it was found that watering soybean plants with a solution containing synthetic miPEP172c peptide resulted in an increase of nodule numbers. This enhanced nodulation correlated with a higher expression of the miR172c primary transcript and several marker genes of nodulation while the *NNC1* gene expression was significantly reduced (Couzigou et al., 2016).

Peptides Encoded by Upstream ORFs (uORFs)

Regulatory sPEPs can also be encoded by sORFs located within 5′ leader sequences of protein encoding mRNAs. These sORFs are commonly referred to as uORFs. uORFs are ubiquitous and have been identified in most eukaryotes. Close to 50% of mRNAs contain uORFs (Andrews and Rothnagel, 2014; Hellens et al., 2016). Ribosome profiling and proteomics in mammalian and human cells revealed that many of the uORFs are indeed translated but only few have been functionally analyzed. A common function of these characterized uORFs is to attenuate the translation of their associated downstream coding ORF by stalling of the ribosomes on the 5′ leader sequence (Andrews and Rothnagel, 2014).

Such a uORF and its encoded sPEP have an essential role in the nodule meristem maintenance in *M. truncatula*. The NF-YA1 transcription factor in *M. truncatula* (also known as MtHAP2-1) controls nodule meristem function (Combier et al., 2006). The *NF-YA1* gene is expressed in the meristem and its spatial expression profile is finely regulated by two mechanisms that repress expression of *NF-YA1* in the adjacent infection zone of the nodule (zone II). In younger nodules, the *NF-YA1* gene is negatively regulated in zone II by the miR169 miRNA (Combier et al., 2006). In older nodules, a second mechanism takes over the negative regulation in zone II. An alternatively spliced mRNA of the *NF-YA1* gene becomes expressed in this zone. The alternative splicing of the first intron of the *NF-YA1* gene results in a transcript with a long 5′ leader sequence containing an uORF encoding the 62 amino acids peptide called uORF1p (Combier et al., 2008a). The translation of this peptide was supported by translational fusions of uORF1p with GUS. The functionality of the uORF1p peptide was investigated with

its overexpression, which reduced the expression of the *NF-YA1* gene resulting in the development of aberrant nodules lacking tissue differentiation. Moreover, specific binding of the uORF1p peptide was demonstrated *in vitro* to the 5′ region of the *NF-YA1* transcript. Together, the data indicate that the synthesis of the uORFp1 peptide in zone II reduces in *trans* the mRNA levels of its cognate *NF-YA1* gene and thus in combination with the miR169, it restricts the expression of the gene to the nodule meristem. Interestingly, this in-*trans*-mode-of-action of the uORFp1 peptide on mRNA levels is unique because other described uORFs are *cis*-acting and inhibit translation of the downstream ORF by ribosome stalling (Andrews and Rothnagel, 2014).

Short Peptide of Early Nodulin ENOD40

sORFs are also present in RNAs annotated as lncRNAs, which are transcripts that do not encode a longer protein. In most cases, it is difficult to predict the significance of the sORFs in these lncRNAs (Andrews and Rothnagel, 2014). Nevertheless, a few examples of sPEPs encoded by these transcripts have been reported in plants (Tavormina et al., 2015). One of these sPEP and lncRNA encoding plant genes is the well-studied but still enigmatic *ENOD40* gene of legumes. *ENOD40* expression is induced in the incipient nodule primordium, and its expression is under the control of the early symbiosis signaling cascade (Frugier et al., 2008). Both *M. truncatula* and *L. japonicus* have two *ENOD40* gene copies and their downregulation or overexpression in transgenic plants result in a reduced or enhanced initiation of nodule primordia, respectively (Charon et al., 1999; Kumagai et al., 2006; Wan et al., 2007). Thus, the genes play a key role in the initiation of nodules and the establishment of the nodule primordium. The *ENOD40* gene lacks a long ORF but several sORFs are present, some of which are conserved among plant species, notably two sORFs encoding the peptides ENOD40-I (13 amino acids) and ENOD40-II (27 amino acids) (Sousa et al., 2001). The translatability of the sORFs was suggested by *in vivo* translational GUS fusions expressed in *Medicago sativa* roots. Moreover, a soybean ENOD40 peptide could be detected by western blotting (Röhrig et al., 2002). Ballistic microtargeting of the *ENOD40* gene into cortical root cells of *M. sativa* was found to induce cell divisions, which correspond to the early steps of nodule initiation (Charon et al., 1997; Sousa et al., 2001). *ENOD40* variants were used in this assay to demonstrate that both the ENOD40-I and ENOD40-II peptides as well as a structured RNA region of the transcript are involved in the elicitation of cortical cell divisions (Sousa et al., 2001). Another reported cellular activity of the *M. truncatula ENOD40* gene is the re-localization of the RNA-binding protein MtRBP1 from the nucleus into the cytoplasm (Campalans et al., 2004). The re-localization of MtRBP1 is dependent on the *ENOD40* RNA. However, the *ENOD40*-encoded peptides do not seem to be involved in this activity because mutant *ENOD40* genes where the translational initiation ATG codons of the peptides were mutated, were still able to induce the cytoplasmic localization of MtRBP1. Moreover, Laporte et al. (2010) reported binding of the *ENOD40* RNA molecule also with the SNARP peptides, which are further described in detail below. One can speculate that forming *ENOD40* RNA–protein interactions may be related to the facilitation of the translation of small proteins like the SNARPs. The picture is even more complex, as it was also found that the soybean ENOD40 sPEPs, named ENOD40-A and ENOD40-B, covalently bind to sucrose synthase, thereby stimulating its sucrose cleavage activity and protein stability. This suggests that the ENOD40 peptides are involved in the control of sucrose use in incipient nodules (Röhrig et al., 2002, 2004). Because the activity of sucrose synthase in plant tissues and organs correlates with the sink strength of these tissues, the ability to attract sucrose, and because sucrose synthase is involved in nodule initiation and essential for effective nitrogen fixation in nodules, it was suggested that the ENOD40 peptides may increase the carbon sink strength in pre-dividing root cortical cells and in mature nodule tissues (Röhrig et al., 2002, 2004).

Thus altogether, it seems that the *ENOD40* genes of legumes act both as a structured RNA molecule and by encoding sPEPs (Bardou et al., 2011). However, how the reported activities of the RNA molecule or the sPEPs are related to the key *ENOD40* functions in the activation of cell divisions in the root cortex and nodule primordium initiation remains a conundrum that requires more investigations.

PLANT PEPTIDES REGULATING NODULE NUMBER

CLE Peptides, Signals in the Autoregulation of Nodulation

The plant has to maintain the balance between the gains and costs of the formation and functioning of the symbiotic nodule, which are energy and carbon demanding processes. That is why legumes control the number of nodules formed on their roots. Both available nitrogen (mainly nitrate) sources and newly forming nodules restrict the initiation and progression of nodule development in the zone susceptible for rhizobial infection (Pierce and Bauer, 1983). Kosslak and Bohlool (1984) demonstrated with split-root inoculation experiments the existence of a long-distance signaling mechanism called autoregulation of nodulation (AON), which prevents the formation of new nodules on the whole root system after the initiation of nodule development at the first inoculation site [for review, see Reid et al. (2011) and Mortier et al. (2012)]. Using nodule excision experiments (Nutman, 1952; Caetano-Anolles and Gresshoff, 1990) and various split-root approaches including the use of bacterial and plant mutants as well as Nod factors revealed that this systemic regulatory signal is generated very rapidly after root hair curling and before the initiation of visible cortical and pericycle cell divisions, while its strength increases with progression of development (Kosslak and Bohlool, 1984; Mathews et al., 1989; Olsson et al., 1989; Caetano-Anolles and Gresshoff, 1990, 1991; van Brussel et al., 2002; Li et al., 2009). After the isolation of plant mutants defective in nitrate- and autoregulation of nodulation (Carroll et al., 1985; Park and Buttery, 1989; Wopereis et al., 2000; Penmetsa et al., 2003), it was demonstrated with the help of grafting experiments

(Delves et al., 1986) that the shoot plays an important role in AON. It was proposed that in the developing nodule, the graft-transmissible signal cue (Q) is produced and is transported to the shoot where it induces the synthesis of the shoot-derived inhibitor (SDI), which suppresses nodulation (Delves et al., 1986; Caetano-Anolles and Gresshoff, 1991). In addition, the nitrate-tolerant phenotype of the mutants indicated that nitrate- and autoregulation pathways share genetic components. Map-based cloning of the mutated genes (HAR1 in L. japonicus, NARK in Glycine max, SUNN in M. truncatula) identified a receptor kinase (Nishimura et al., 2002; Searle et al., 2003; Schnabel et al., 2005) that is the closest legume homolog of the Arabidopsis Clavata1 (CLV1) receptor (Clark et al., 1997). Also by genetic analysis, the receptor-like kinases KLAVIER and the membrane protein CLV2 were later identified as likely co-receptors of HAR1 in L. japonicus (Miyazawa et al., 2010; Krusell et al., 2011). Arabidopsis CLV1 functions in a protein complex controlling stem cell proliferation by short-distance signaling in shoot apices. As CLV1 is known to act as a receptor for the shoot apical meristem regulator CLV3 of the CLV3/ESR-related (CLE) small secreted peptide family (Fletcher et al., 1999; Cock and McCormick, 2001), it was immediately hypothesized that the ligand of HAR1, the signal Q, might be a CLE peptide (Nishimura et al., 2002). Indeed, it was shown that a few genes from the gene family encoding CLE peptides in legumes have Nod factor-dependent nodule-enhanced or even -specific and/or nitrate-dependent expression, and their ectopic expression led to systemic HAR1/SUNN/NARK-mediated repression of nodulation by interference with the nodulation signaling pathway (Okamoto et al., 2009; Mortier et al., 2010; Reid et al., 2011; Saur et al., 2011). Interestingly, the L. japonicus CLE-RS2 gene plays a role in both nitrate- and nodulation dependent regulation, and its expression is under dual regulation by the Nod factor signal transduction pathway during nodulation via the NIN transcription factor and independently from it by nitrate exposition via the NIN-like transcription factor NRSYM1 (Okamoto et al., 2009; Soyano et al., 2014; Nishida et al., 2018). In contrast, soybean responds to nodulation by the expression of the GmRIC1 and GmRIC2 genes, while nitrate induces the expression of GmNIC1, another CLE encoding gene (Reid et al., 2011). The translated products of the CLE genes undergo extensive post-translational modifications and proteolytic processing resulting in 13 amino-acid long mature peptides with hydroxylated prolines in the fourth and the seventh positions (Okamoto et al., 2013). In addition, the hydroxyproline in the seventh position is glycosylated with three arabinosyl residues (Okamoto et al., 2013) by the activity of the RDN1 and RDN1-related proteins (Schnabel et al., 2011; Kassaw et al., 2017). The glycosylation by the tri-arabinosyl oligosaccharide is absolutely required for the activity of the peptides because mono-arabinosylated peptides or peptides only with hydroxy-prolines have no biological activity (Imin et al., 2018). The hypothesis that a CLE peptide is signal Q that travels from the root to the shoot to bind to HAR1 has been proven by showing that the triple-arabinosylated peptide CLE-RS2 is transported through the xylem of L. japonicus and binds directly to HAR1 (Okamoto et al., 2013). On the other hand, the nature of the SDI, which is produced after CLE-mediated

activation of its shoot receptor is less clearly defined but, in L. japonicus, it involves cytokinin production in the shoot via the CLE-HAR1/SUNN/NARK-activated isopentenyltransferase gene IPT3, suggesting that SDI is a cytokinin derivative or that the shoot-derived cytokinins generate a secondary signal (Sasaki et al., 2014). In line with this hypothesis, it was shown by petiole feeding of soybean leaf extracts from AON-induced plants that SDI is a small molecular weight (<1 kDa) molecule that is heat-stable and resistant against the activity of proteases and RNases (Lin et al., 2010). The inhibition of nodulation in the roots by SDI further requires the F-box protein Too Much Love (TML), whose target molecules are not yet known (Takahara et al., 2013; Sasaki et al., 2014).

CEP Peptides, Positive Effectors of Nodulation Efficiency

Another class of post-translationally modified peptides, the CEP molecules, was also found to regulate systemically nodule formation but unlike the CLE peptides, having a positive effect on nodule number. These peptides are involved in controlling other developmental processes as well, such as lateral root development and nitrate transporter deployment. All these functions are related to assuring the adequate nitrogen supply for plants, and by this means, CEPs can be central molecules coordinating these processes (Taleski et al., 2018). Several CEP genes were induced by low-nitrogen conditions in M. truncatula (Imin et al., 2013). Among them, MtCEP1 was shown to positively influence the number of nodules on M. truncatula roots and at the same time to negatively control lateral root formation. Both overexpression of MtCEP1 and adding synthetic MtCEP1 peptide resulted in increased nodule number and size, as well as more efficient nitrogen-fixation, and even partially tolerating high-nitrogen levels, which typically strongly suppresses nodulation (Imin et al., 2013). MtCEP1 peptide treatment also increased the root competency for nodule development as well as infection thread formation. MtCEP1 could also alleviate the inhibitory effects of increased ethylene-precursor levels on nodulation without affecting the ethylene production (Mohd-Radzman et al., 2016). This work has revealed an interface between MtCEP1 and the phytohormone-mediated signaling that regulates nodulation efficiency and plant susceptibility to infection in M. truncatula. Genetic evidence presented by Mohd-Radzman et al. (2016) also proved that the positive effect of MtCEP1 on nodulation is dependent on COMPACT ROOT ARCHITECTURE 2 (CRA2) (Huault et al., 2014) through—at least in part—the ethylene signal-transduction pathway including MtEIN2/SKL (ETHYLENE INSENSITIVE2; SICKLE) (Penmetsa and Cook, 1997; Oldroyd and Downie, 2008). CRA2, which is the putative receptor of MtCEP1, was shown to act positively on root nodule formation systemically from the shoot (Huault et al., 2014), however, independently from the AON regulation. Thus, CRA2 and MtCEP1 represent a new systemic circuit of regulation on nodulation.

The functional, 15 amino acid long CEP family members are processed from non-functional prepropeptides and decorated with similar post-translational modifications as CLE peptides.

CEP genes encode prepropeptides with an N-terminal secretion signal sequence, a variable domain, one or more conserved CEP domains, and one or more flanking variable regions (Ogilvie et al., 2014). MtCEP1 has two conserved CEP domains, D1 and D2. CEPs are frequently hydroxylated at various proline residues and the pattern of hydroxylation has an influence on their biological activity (Delay et al., 2013; Imin et al., 2013; Mohd-Radzman et al., 2015). MtCEP1 D1 peptide variants were also identified with tri-arabinosylation at proline in the 11th position (Mohd-Radzman et al., 2015). A recent work by Patel et al. (2018) analyzed the secreted peptidome of *Medicago* hairy root cultures and xylem sap and found completely new versions of the MtCEP peptides. Some of them possessed unexpected N- and C-terminal extensions that suggested roles for endo- and exoproteases in CEP peptide maturation. These authors determined not only the structure of these molecules with various length and modifications but also chemically synthesized different MtCEP1 D1 variants to test their biological activities. The peptides with N-terminal extensions were unable to increase root nodule number, while the variant with only one amino acid C-terminal extension had biological activity. Unexpectedly, tri-arabinosylated MtCEP1 D1 derivatives had a reduced capacity to increase nodule numbers. Thus, this post-translational modification seems to have a different effect on the biological activity of CLE and CEP peptides. It remains to be determined whether these modifications affect the perception of the CEPs by their receptor and what are the elements involved in the transduction of the CEP signal. Furthermore, the exact biological meaning of tri-arabinosylation of the CEP peptides needs further analysis. The intriguing opposing effect of this post-translational modification on nodule-inhibiting CLE and nodule-stimulating CEP suggests that arabinosylation of peptides plays key regulatory roles in the peptides' activity controlling nodule numbers that can integrate the AON and CRA2/CEP regulatory circuits.

NODULE-SPECIFIC PEPTIDES TARGETED TO THE SYMBIOSOME

Nodule-Specific Peptides Governing Terminal Bacteroid Differentiation

Beijerinck (1888) described the bacteroids in *Vicia faba* nodules as "derived from bacteria by a metamorphic process, that have lost their ability to reproduce... They are derived from normal *Bacillus radicicola* (probably a *Rhizobium leguminosarum* sp.) by a stepwise loss in their power of reproduction. Bacteria that are still capable of growth on gelatin plates can be isolated in large numbers from the very young root nodules, as well as from the actively growing regions of older root nodules." This remarkably precise account, made 130 years ago, is one of the first descriptions of bacteria housed in legume nodules and it drew already the attention to the striking differentiation process of the bacteria in nodules. Since then, this process has been on and off (but mostly off) the scientific agenda of researchers in the field, and it is only since the last 15 years, with the

discovery of the NCR peptides that we see a renewed interest. Beijerinck made drawings of large, often *Y*-shaped terminally differentiated bacteroids. Such bacteroids were observed in the nodules of *Vicia*, *Pisum*, and *Medicago* species belonging to the IRLC and were thought to be characteristic features of the indeterminate nodules. Later works revealed that (i) terminal bacteroid differentiation is not universal in the legume family and depends on the genetic repertoire of the host plant (Mergaert et al., 2006); (ii) it is not a general characteristic of the indeterminate nodules (Ishihara et al., 2011); (iii) the ability to direct bacterial differentiation into swollen (most probably terminally differentiated) bacteroids evolved independently in five out of the six investigated subclasses of the Papilionoideae subfamily (Oono et al., 2010); (iv) terminally differentiated bacteroids fix nitrogen more efficiently than unaltered ones (Sen and Weaver, 1984; Oono and Denison, 2010); (v) that the process is in large part determined by nodule-specific peptides called NCRs and possibly other secreted peptides. Comparative nodule transcriptome analysis of two model legumes, *M. truncatula* and *L. japonicus* hosting terminally differentiated and unaltered bacteroids, respectively, in their nodules identified three gene families in *M. truncatula* encoding secreted peptides that are missing from the *L. japonicus* transcriptome (Kevei et al., 2002; Mergaert et al., 2003; Laporte et al., 2010; Trujillo et al., 2014). These families, called the NCRs, the GRPs, and the SNARPs are described below.

The Nodule-Specific NCRs of *M. truncatula*

The Extremely Large *NCR* Gene Family

The genes in the largest family with over 700 members in *M. truncatula* code for peptides termed nodule-specific cysteine-rich (NCR) peptides (Mergaert et al., 2003). Nearly all *NCR* genes are exclusively expressed in the infected cells of the nodules (Mergaert et al., 2003; Guefrachi et al., 2014). The gene products are characterized by four or six cysteines in conserved positions in the otherwise extremely divergent mature peptide sequence and by a relatively conserved signal peptide sequence. The structure of the NCR peptides resembles that of defensins, innate immunity effectors in plants, which have the capacity to target and kill infecting microbes. Many NCRs have indeed antimicrobial activity; however, they are different from defensins in many aspects (Maróti et al., 2011; Maróti and Kondorosi, 2014). Unlike defensins, NCR peptides have no role in immunity; they only have a function in symbiosis and are targeted to the bacteroids as was shown by immunological methods (Van de Velde et al., 2010) and by detecting over 200 peptides in the bacteroid proteome (Durgo et al., 2015; Marx et al., 2016). The plethora of NCR peptides, evolving with gene duplication and diversifying selection, reflects likely multiple interactions with bacterial targets and many diverse modes of actions.

NCR Peptides Control Terminal Bacteroid Differentiation

Treating rhizobia with NCR peptides or ectopic expression of NCR genes in legumes devoid of NCRs provoked symptoms

of terminal differentiation (endoreduplication of the bacterial genome together with enlargement of the cell, loss of cell division capacity, increased membrane permeability) indicating that these peptides govern the differentiation process (Van de Velde et al., 2010). Further evidence came by blocking the transport of NCR peptides to the bacteroids in the *M. truncatula* signal peptidase complex mutant, which resulted in the complete absence of bacteroid differentiation (Van de Velde et al., 2010; Wang et al., 2010). The mode of action of the more than 600 NCR peptides in *M. truncatula* remains elusive except for a few cases and their activity might be quite different based on the low sequence similarity of the individual members. The high diversity in amino acid sequence and composition of the mature peptides provide large variations in their physicochemical properties, which is reflected also by the wide spectrum of isoelectric points (p*I*) ranging from 3.5 to 10.5. Roughly one-third of the NCRs are cationic while the rest are anionic or neutral (Montiel et al., 2017).

Most studies have been focused on the cationic NCR peptides as *in vitro* they possess strong antimicrobial activity against a wide range of Gram-negative and Gram-positive bacteria as well as unicellular and filamentous fungi (Tiricz et al., 2013; Ördögh et al., 2014). This bactericidal and mycocidal activity is mediated *via* the disruption of the integrity of microbial membranes (Ördögh et al., 2014; Mikuláss et al., 2016). However, these experiments were performed at high concentrations of synthetic NCR peptides, which do not reflect and are most likely incomparable with the peptide concentrations in the nodule cells where, in addition, many other NCRs, cationic and non-cationic ones are present that might act together.

Terminal bacteroid differentiation is accompanied by endoreduplication when the genome of the bacteria duplicated without cell division. A crucial step of the cell division is the formation of the *Z*-ring, assembled by polymerization and localization of the FtsZ protein at the future site of the septum, required for separating the mother and daughter cells (Lutkenhaus and Addinall, 1997). The cationic NCR247 peptide does not provoke membrane damage at sub-lethal concentrations but enters the bacterial cytosol and drastically alters the physiology of the bacterium manifested by disappearance of proliferating cells and appearance of septum-less elongated cells. The NCR247 peptide binds to FtsZ and this interaction abolishes polymerization of FtsZ and thereby septum formation (Farkas et al., 2014). Interestingly, the same symbiotic cells produce another peptide, NCR035, which in growing bacterial cultures localizes to the septum and this localization can be abolished by treating rhizobia with the NCR247 peptide (Farkas et al., 2014). Thus, it seems that more than one peptide might affect single biological pathways and processes, particularly those with key importance in symbiosis, such as stopping bacterial proliferation in the host cell. Moreover, the NCR247 peptide attenuates the expression of critical cell cycle regulator genes *ctrA*, *gcrA*, *dnaA* as well as cell division genes, including genes required for *Z*-ring function, among others (Tiricz et al., 2013; Penterman et al., 2014), another way to regulate the cell cycle of the developing bacteroids. In addition, this peptide inhibits translation not only by downregulating the expression of ribosomal genes

but also *via* binding to several ribosomal proteins. NCR247 might thus contribute to the altered proteome and physiology of the bacteroids. These effects could be amplified by binding of NCR247 to the GroEL chaperon modifying presumably interaction of GroEL with other proteins (Farkas et al., 2014).

Other Roles of NCR Peptides

At present, it is unknown what the role and the extent of the NCRs' antimicrobial activity are during symbiosis. In addition to the expected much lower NCR concentrations in the nodules than the ones used in *in vitro* experiments, non-cationic peptides present in the NCR cocktails produced by a given symbiotic cell might counteract the killing effect of the cationic peptides. An intriguing possibility is that the cationic peptides facilitate somehow the uptake of the acidic and neutral ones or act in complexes in the membranes and/or in the cytoplasm. Indeed, NCR247 was shown to interact with two anionic NCRs (Farkas et al., 2014), but the promotion of uptake of these peptides by NCR247 has not been demonstrated. The antibacterial activity of NCRs likely keeps rhizobia on the verge of destruction, manifested in the immediate and NCR-dependent death of *bacA* mutant bacteria in the nodule. The BacA protein is a peptide transporter that provides tolerance toward the antimicrobial activity of the NCR peptides (Haag et al., 2011).

The role of neutral and anionic NCRs, on the other hand, is a great enigma. They were shown to accumulate in the bacteroids and they might be major players providing a plethora of novel activities. However, at present, it is unknown how they enter the bacteria and what they do there. Unlike the cationic NCRs, none of the tested neutral or anionic NCRs showed antimicrobial activity (Tiricz et al., 2013; Ördögh et al., 2014). The NCR211 required for the development of effective nodules (see below) is so far the only anionic peptide, which has a mild antibacterial activity (Kim et al., 2015).

Despite the high amino acid sequence variation, the large number of NCR peptides suggested redundant functions making their genetic analysis difficult. However, map-based cloning from plant mutants unable to establish nitrogen-fixing symbiosis led to the identification of single peptides (NCR169 and NCR211) that are required for the development of the effective interaction. Their absence resulted in the arrest of bacteroid differentiation and/or in the loss of bacteroid persistence (Horváth et al., 2015; Kim et al., 2015) *via* a mechanism unknown at present. An interesting and kind of opposing activity of the NCR peptides is their involvement in the selection of the bacterial partner: Incompatibility between *M. truncatula* ecotype Jemalong and *Sinorhizobium meliloti* strains Rm41 and A145, which form effective symbiosis with other *Medicago* partners, results in the elimination of the bacterial partner from the nodule region where nitrogen fixation should take place. This elimination is mediated by allelic variants of two NCR peptides, called NFS1 and NFS2, but the process cannot be explained by a stronger antimicrobial activity of the incompatible variants because the sensitivity of compatible and incompatible bacteria is quite similar (Yang et al., 2017; Wang et al., 2017, 2018). The phenotypes of mutants in *NCR169* and *NCR211* and the *NFS1* and *NFS2* alleles clearly suggest that antimicrobial activity

is not the only mode of action of members of the NCR family.

The Evolution of NCRs in the IRLC

The IRLC is a clade of legumes, which is mostly constituted by temperate herbaceous tribes such as the *Galegeae*, *Carmichaelieae*, *Cicereae*, *Hedysareae*, *Trifolieae*, *Vicieae*, as well as the tropical tribe Millettieae (*Callerya*, *Wisteria*, and related genera) (Wojciechowski et al., 2000). Based on similarity searches, NCRs were first recognized in nodule EST sequencing data of several IRLC species from distinct genera and subclades (Frühling et al., 2000; Györgyey et al., 2000; Jimenez-Zurdo et al., 2000; Crockard et al., 2002; Fedorova et al., 2002; Kaijalainen et al., 2002; Kato et al., 2002; Mergaert et al., 2003; Chou et al., 2006). Nodule transcriptome sequencing from species representing the main subclades (Hedysaroid, Astragalean, and Vicioid) of the IRLC as well as analysis of RNA-Seq data from other IRLC species shed light on the evolution of NCRs in the IRLC (Montiel et al., 2017). It was shown that the numbers of NCR genes are highly variable (from 7 to >700) and expanded independently in different lineages of IRLC legumes. In nodules of *Glycyrrhiza uralensis* (the most basal IRLC legume) infected with *Mesorhizobium tianshanense*, only seven NCR genes have been identified, none of them encoding peptides with positive charge (Montiel et al., 2017). *M. tianshanense* bacteroids display symptoms of terminal differentiation, however, the swelling of the bacteroids represents a mild morphological response, compared to the drastic enlargement of Y-shaped bacteroids in several Vicioid legumes (Montiel et al., 2016). The small cocktail of NCRs produced by *G. uralensis* and the absence of cationic peptides seems to be insufficient to induce irreversible differentiation in *Sinorhizobium fredii*, a rhizobial strain abnormally resistant to the antimicrobial action of NCR247 and NCR335 (Crespo-Rivas et al., 2016). The GuNCRs identified until now, are likely the ancestor symbiotic peptides in the IRLC, since each of them have at least one putative ortholog in another IRLC legume from different genera (Montiel et al., 2017). The presence of recognizable orthologs between different genera is, however, rather rare. For example, only a few orthologs can be predicted among the closely related species *M. truncatula* and *M. sativa*. The Vicioid legume *Cicer arietinum* represents a peculiar case, where 20 CaNCRs have putative orthologs in the non-Vicioid species *G. uralensis*, *O. lamberti*, *A. canadensis*, and *O. viciifolia*, but surprisingly none in any Vicioid legume. In addition, the 63 NCRs found in the nodule transcriptome of *C. arietinum* represent a considerably low number, compared to the large gene families in *Galega orientalis*, *Ononis spinosa*, *Pisum sativum*, and *Medicago* spp., all of them part of the Vicioid subclade (Wojciechowski et al., 2004; Montiel et al., 2017). Additionally, the swollen and spherical bacteroids of *C. arietinum* nodules contrast with the elongated-branched bacteroids in other legumes located in the same subclade such as *Medicago* sp., *G. orientalis*, *P. sativum*, *V. faba*, and *Trifolium repens* (Mergaert et al., 2003, 2006; Montiel et al., 2016). In general, a positive correlation was found between the degree of bacteroid elongation and the number of the expressed *NCR*s (Montiel et al., 2017). Legumes with elongated-branched bacteroids express hundreds

of NCRs, characterized by a large proportion of cationic peptides with a well-defined isoelectric point (Montiel et al., 2017). Spherical bacteroids can be found both in *C. arietinum* and *O. spinosa* nodules; however, these species share no evident NCR pattern (Lee and Copeland, 1994; Montiel et al., 2016; Montiel et al., 2017). These clues indicate that *NCR* gene families took different evolutionary trajectories, showing variable duplication rates (Alunni et al., 2007; Montiel et al., 2017) that were likely favored by transposable elements located in flanking regions of *NCR* genes as shown in *M. truncatula* (Satgé et al., 2016). Clearly, the enrichment of cationic NCRs with particular isoelectric points had great impact on the morphology of the hosted endosymbionts.

NCR Genes in Dalbergoid Legumes

Terminal bacteroid differentiation is not restricted to the IRLC legumes (Oono et al., 2010). The Dalbergoid clade is one of the other legume groups in which bacteroids, similar to the IRLC legumes, differentiate into polyploid and strongly enlarged bacteria (Czernic et al., 2015). Depending on the host species, these bacteroids have either an elongated morphology similar as in *Medicago* (e.g., *Aeschynomene afraspera*, *Aeschynomene nilotica*) or they can be almost perfect, large spheres as in *C. arietinum* and *O. spinosa* (e.g., in *Aeschynomene indica*, *Aeschynomene evenia*, and *Arachis hypogaea*). Transcriptome analysis in different *Aeschynomene* species identified a family of peptide genes with similar features as the IRLC NCRs. They are secretory peptides, characterized by conserved cysteine motifs in the mature domain. However, the *Aeschynomene* NCR peptides have no sequence similarity to the IRLC NCRs [for example, the spacing and number (six or eight) of cysteines is different]. They form thus a separate family of peptides. The genes encoding the *Aeschynomene* NCRs are only expressed in nodules and they are activated just before the onset of bacteroid differentiation. They are expressed only in the symbiotic nodule cells and a proteome analysis of purified bacteroids demonstrated that the peptides are targeted to them. Moreover, blocking the secretory pathway by RNAi targeting one of the subunits of the signal peptidase complex inhibits bacteroid differentiation (Czernic et al., 2015) as was described before in *M. truncatula* mutated in the orthologous gene (Van de Velde et al., 2010). Another parallel with the bacteroid differentiation in *Medicago* is the requirement of a BacA-like peptide transporter, named BclA, in *Bradyrhizobium* symbionts for the interaction with *Aeschynomene* (Guefrachi et al., 2015). In the absence of this transporter in the *bclA* mutant, the bacteroids do not differentiate into their polyploid and elongated forms and die, exactly as the phenotype of the *S. meliloti bacA* mutant in *Medicago* nodules (Haag et al., 2011).

Together, these similitudes between the *Aeschynomene* and the IRLC suggest that bacteroid differentiation in the Dalbergioid clade, which evolved independently from the bacteroid differentiation in the IRLC clade (Oono et al., 2010), is based on very similar mechanisms used by IRLC legumes. Nevertheless, some unresolved questions remain. One of them is that a recent transcriptome analysis in *A. hypogaea* failed to detect homologs of the *Aeschynomene* NCR genes (Karmakar et al., 2018). However, this study identified another, small, family

of cysteine-containing peptides, related to the antimicrobial PR-1 family and the authors suggested the involvement of these peptides in bacteroid differentiation. Alternatively, the *A. hypogaea* and *Aeschynomene* NCR peptides diverged too much to be identified by homology and a specific bioinformatics search for small, secreted, and cysteine-rich peptides would be required to identify them. A second open question that requires further investigation is the absence of detectable *in vitro* activity of the *Aeschynomene* peptides (Czernic et al., 2015). As described above, many of the *Medicago* peptides show a strong action against bacteria, including arrest of division and membrane permeabilization and complete cell lysis (Tiricz et al., 2013; Ördögh et al., 2014; Mikuláss et al., 2016). None of the thus far tested *Aeschynomene* peptides displayed such an activity on the *Brabyrhizobium* symbionts of *Aeschynomene* nor on any other tested bacterium including ones that show a strong response to the *Medicago* peptides. All identified peptides in the *Aeschynomene* transcriptome are either neutral or anionic while the most active antimicrobial NCR peptides of *Medicago* are positively charged. In addition, the *Bradyrhizobium* symbionts seem to be very robust bacteria that are highly resistant to antimicrobial peptides including the most active *Medicago* NCR peptides with a broad spectrum of activity. This robustness of bradyrhizobia is due to their very tough cell envelope. Contrary to most other rhizobia, bradyrhizobial envelopes contain hopanoids, a class of bacterial lipids, similar to eukaryotic steroids (cholesterol). They are known to render bacterial membranes more rigid and resistant to membrane stresses, including ones caused by antimicrobial peptides (Belin et al., 2018). And indeed, hopanoid mutants of *Bradyrhizobium* become more sensitive to *Medicago* NCR peptides and other antimicrobial peptides (Kulkarni et al., 2015). Altogether, this leaves open the question how the *Bradyrhizobium* symbionts in the symbiotic cells of *Aeschynomene* nodules are manipulated by the host to respond to these host signals.

The formation of spherical bacteroids in some hosts like the Dalbergoid *A. indica* or *A. evenia* as well as in the IRLC *C. arietinum* and *O. spinosa* is an additional unsettled point. In *A. indica*, the spherical bacteroids are formed through an intermediate stage of elongation similar to the bacteroids in other *Aeschynomene* or in *Medicago* (Czernic et al., 2015). It is unknown if the transition from elongated to spherical morphotypes is the result of the action of specific NCR peptides or of another host factor.

Nodule-Specific Glycine-Rich Proteins (GRPs)

Glycine-rich protein-encoding genes are a second group of nodule-specific transcripts that seems to be restricted to the IRLC. They were originally identified in *V. faba* and *Medicago* spp. (Küster et al., 1995; Schröder et al., 1997; Györgyey et al., 2000; Jimenez-Zurdo et al., 2000; Kevei et al., 2002; Alunni et al., 2007). The *GRP* gene family is much smaller than the NCR one with less than 30 members in *M. truncatula* (Alunni et al., 2007).

Glycine-rich proteins have been described in a wide variety of plant species performing variable roles including activity in biotic and abiotic interactions of the plants with their environment (Sachetto-Martins et al., 2000). Semi-repetitive glycine regions

characterize GRP sequences that can be classified according to the presence of different binding motifs or a signal peptide (Mangeon et al., 2010). Usually, GRPs have around 80% glycine content arranged in specific motifs, but the nodule expressed secreted GRPs are shorter polypeptides than the usual GRPs and possess only 20–30% glycine residues without any recognizable motif. Interestingly, the signal peptide sequence of the nodule-specific GRPs found in IRLC legumes is also a distinctive feature not shared with the signal peptides of GRPs from other plant species (Kevei et al., 2002; Alunni et al., 2007).

A recent search for GRPs revealed that these peptides are also expressed in the nodules of representative species from the Astragalean and Hedysaroid subclades along with *G. uralensis* (J. Montiel, unpublished). However, the size of GRP families is considerably lower in the non-Vicioid species compared to the Vicioid legumes *G. orientalis*, *C. arietinum*, *O. spinosa*, and *Medicago* spp. Unlike the *NCR* gene families, the enrichment and diversification of the *GRP* families show no correlation with the morphotype of the hosted bacteroids, and rather seems to be specific for members of the Vicioid subclade (J. Montiel, unpublished). The expression profile of the *GRP*s in the different nodule zones is another relevant difference to *NCR*s. In *M. truncatula*, 39% of *GRP* transcripts are present in the infection zone, the nodule tissue where bacteroid differentiation takes places (J. Montiel, unpublished), while this region contributes only to 18% of *NCR* transcripts (Roux et al., 2014; Montiel et al., 2017). This observation indicates that several GRPs are potentially involved in bacteroid differentiation (Kevei et al., 2002; Kondorosi et al., 2013). Gene characterization of different *GRP*s through reverse genetics could help to understand the role(s) played by these proteins in nodulation and their high diversification within Vicioid legumes.

SNARPs or LEED..PEEDs

The SNARPs or LEED..PEEDs form a small family, 10–13 members, of small secreted and nodule specific peptides in *Medicago* (Laporte et al., 2010; Trujillo et al., 2014). These peptides are not longer than 70 amino acids and are characterized by one or two conserved domains of acidic amino acid residues, the LEED or PEED domains, hence, one of their names (Trujillo et al., 2014). They were also characterized as RNA-binding peptides, from there, the other name, small nodulin acidic RNA-binding protein, or SNARP in short (Laporte et al., 2010). Intriguingly, this peptide family is specific to the *Medicago* lineage (*M. truncatula* and *M. sativa*) because homologous sequences are absent in all other genomes of legumes or other plant species, showing that the family arose within this clade during the past 25 million years (Trujillo et al., 2014). Their expression, similarly to the above described NCR and GRP peptides, is absolutely nodule specific with transcripts only found in the distal and proximal infection zones, the interzone and the nitrogen fixation zone of nodules while they are absent in all other plant tissues. This expression pattern suggests a specific role of these peptides in the later stages of nodule development, potentially in symbiotic cell differentiation or bacteroid formation. As described above, two members of the family, SNARP1 and SNARP2 were identified in a yeast three-hybrid screen for proteins that interact with the

MtENOD40 mRNA (Campalans et al., 2004). *In vitro* biochemical studies further demonstrated that the SNARP2 protein has non-specific binding activity to single-stranded RNA (Laporte et al., 2010). On the other hand, the LEED..PEED/SNARP proteins are secretory proteins indicating that they should be localized in the endomembrane system, the symbiosomes, or in the extracellular space (Laporte et al., 2010). How these putative localizations, which—except for the bacteroids in the symbiosomes—supposedly do not contain RNA, can be reconciled with the RNA-binding activity of these peptides needs further investigation. However, even if their molecular role is still unclear, the importance of SNARP peptides in nodule development is strongly supported by RNAi inactivation of the *MtSNARP2* gene, which led to the formation of abnormal nodules (Laporte et al., 2010). In these nodules, infection of symbiotic cells and bacteroid formation seemed to proceed normally but the symbiotic cells and their bacteroids were not stably maintained and degenerated prematurely. Thus, even if these reverse genetic experiments conclusively demonstrate the importance of the SNARP peptides for normal symbiotic cell formation, they raise at the same time new questions. Why are these peptides essential in *Medicago* nodules while these peptides are absent in closely related legumes such as pea, clover, or chickpea (Trujillo et al., 2014), which form nodules very similar in structure and function to the *Medicago* nodules?

CONCLUDING REMARKS

The large number of different peptides and peptide families that we have described here are those for which at least a minimal amount of evidence demonstrates a specific role in symbiosis. But they might just as well be only the tip of the proverbial "peptide-iceberg." Small proteins have traditionally escaped gene prediction efforts in plant genomes because algorithms were biased against them by a concern to avoid wrongful annotations. However, in recent years, predicting peptide genes in plant genomes by dedicated bioinformatics tools have provided the insight that functional genes encoding small peptides are massively hidden in plant genomes (Silverstein et al., 2005, 2007; Lease and Walker, 2006; Hanada et al., 2013; Pan et al., 2013; Ghorbani et al., 2015;

de Bang et al., 2017). Combined with transcriptomics, peptide predictions have been the impetus for the functional characterization of many of the above described peptides. In a recent large scale effort, the Mt4.0 and Mt3.5v5 releases of the *M. truncatula* genome were re-annotated using a suite of bioinformatics programs with the specific aim to search for ORFs encoding small secreted peptides (SSPs) (de Bang et al., 2017). This approach yielded a comprehensive catalog of almost 2,000 genes from 46 previously defined SSP families, including all the above described families. In addition, another catalog of almost 2,500 genes encoding putative novel SSPs was established. Focusing on SSP genes, known or suspected to function *via* receptor-mediated signaling, a transcriptome analysis by RNA-seq was performed during a time course of nodule formation and in response to Nod factors. This analysis revealed 365 differentially expressed known signaling SSPs plus an additional several hundred genes encoding putative novel SSPs. The very large majority of these differentially expressed genes were up- or downregulated in developing or mature nodules. Their differential regulation during the early stages of nodulation and Nod factor signaling has not yet been tested thoroughly. But in any case, these results suggest an unanticipated complexity and importance of peptide-mediated signaling in the orchestration of the symbiosis.

AUTHOR CONTRIBUTIONS

AK, PM, JM, GE, and ÉK wrote the manuscript and read and approved the final version of the manuscript.

FUNDING

REFERENCES

Alunni, B., Kevei, Z., Redondo-Nieto, M., Kondorosi, A., Mergaert, P., and Kondorosi, E. (2007). Genomic organization and evolutionary insights on *GRP* and *NCR* genes, two large nodule-specific gene families in *Medicago truncatula*. *Mol. Plant Microbe Interact.* 20, 1138–1148. doi: 10.1094/MPMI-20-9-1138

Andrews, S. J., and Rothnagel, J. A. (2014). Emerging evidence for functional peptides encoded by short open reading frames. *Nat. Rev. Genet.* 15, 193–204. doi: 10.1038/nrg3520

Antolín-Llovera, M., Ried, M. K., and Parniske, M. (2014). Cleavage of the symbiosis receptor-like kinase ectodomain promotes complex formation with nod factor receptor 5. *Curr. Biol.* 24, 422–427. doi: 10.1016/j.cub.2013.12.053

Bardou, F., Merchan, F., Ariel, F., and Crespi, M. (2011). Dual RNAs in plants. *Biochimie* 93, 1950–1954. doi: 10.1016/j.biochi.2011.07.028

Beijerinck, M. W. (1888). Die bacterien der papilionaceenknöllchen. *Botanische Zeitung* 46, 725–804.

Belin, B. J., Busset, N., Giraud, E., Molinaro, A., Silipo, A., and Newman, D. K. (2018). Hopanoid lipids: from membranes to plant–bacteria interactions. *Nat. Rev. Microbiol.* 16, 304–315. doi: 10.1038/nrmicro.2017.173

Butenko, M. A., Patterson, S. E., Grini, P. E., Stenvik, G., Amundsen, S. S., Mandal, A., et al. (2003). *INFLORESCENCE DEFICIENT IN ABSCISSION* controls floral organ abscission in *Arabidopsis* and identifies a novel family of putative ligands in plants. *Plant Cell* 15, 2296–2307. doi: 10.1105/tpc.014 365.1

Caetano-Anolles, G., and Gresshoff, P. M. (1990). Early induction of feedback regulatory responses governing nodulation in soybean. *Plant Sci.* 71, 69–81. doi: 10.1016/0168-9452(90)90069-Z

Caetano-Anolles, G., and Gresshoff, P. M. (1991). Plant genetic control of nodulation. *Annu. Rev. Microbiol.* 45, 345–382. doi: 10.1146/annurev.mi.45. 100191.002021

Campalans, A., Kondorosi, A., and Crespi, M. (2004). *Enod40*, a short open reading frame–containing mRNA, induces cytoplasmic localization of a nuclear RNA

binding protein in *Medicago truncatula*. *Plant Cell* 16, 1047–1059. doi: 10.1105/tpc.019406

Carroll, B. J., McNeil, D. L., and Gresshoff, P. M. (1985). Isolation and properties of soybean (*Glycine max* (L)merr.) mutants that nodulate in the presence of high nitrate concentrations. *Proc. Natl. Acad. Sci. U.S.A.* 82, 4164–4166. doi: 10.1073/pnas.82.12.4162

Charon, C., Johansson, C., Kondorosi, E., Kondorosi, A., and Crespi, M. (1997). *Enod40* induces dedifferentiation and division of root cortical cells in legumes. *Proc. Natl. Acad. Sci. U.S.A.* 94, 8901–8906. doi: 10.1073/pnas.94.16.8901

Charon, C., Sousa, C., Crespi, M., and Kondorosi, A. (1999). Alteration of enod40 expression modifies *Medicago truncatula* root nodule development induced by *Sinorhizobium meliloti*. *Plant Cell* 11, 1953–1966. doi: 10.1105/tpc.11.10.1953

Chou, M. X., Wei, X. Y., Chen, D. S., and Zhou, J. C. (2006). Thirteen nodule-specific or nodule-enhanced genes encoding products homologous to cysteine cluster proteins or plant lipid transfer proteins are identified in *Astragalus sinicus* L. by suppressive subtractive hybridization. *J. Exp. Bot.* 57, 2673–2685. doi: 10.1093/jxb/erl030

Clark, S. E., Williams, R. W., and Meyerowitz, E. M. (1997). The *CLAVATA1* gene encodes a putative receptor kinase that controls shoot and floral meristem size in *Arabidopsis*. *Cell* 89, 575–585. doi: 10.1016/S0092-8674(00)80239-31

Cock, J. M., and McCormick, S. (2001). A large family of genes that share homology with CLAVATA3. *Plant Physiol.* 126, 939–942. doi: 10.1104/pp.126.3.939

Combier, J.-P., De Billy, F., Gamas, P., Niebel, A., and Rivas, S. (2008a). Trans-regulation of the expression of the transcription factor *MtHAP2-1* by a uORF controls root nodule development. *Genes Dev.* 22, 1549–1559. doi: 10.1101/gad.461808

Combier, J.-P., Frugier, F., De Billy, F., Boualem, A., El-Yahyaoui, F., Moreau, S., et al. (2006). *MtHAP2-1* is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in *Medicago truncatula*. *Genes Dev.* 20, 3084–3088. doi: 10.1101/gad.402806

Combier, J.-P., Küster, H., Journet, E.-P., Hohnjec, N., Gamas, P., and Niebel, A. (2008b). Evidence for the involvement in nodulation of the two small putative regulatory peptide-encoding genes *MtRALFL1* and *MtDVL1*. *Mol. Plant Microbe Interact.* 21, 1118–1127. doi: 10.1094/MPMI-21-8-1118

Constabel, C. P., Bergey, D. R., and Ryan, C. A. (1995). Systemin activates synthesis of wound-inducible tomato leaf polyphenol oxidase via the octadecanoid defense signaling pathway. *Proc. Natl. Acad. Sci. U.S.A.* 92, 407–411. doi: 10.1073/pnas.92.2.407

Costa, L. M., Marshall, E., Tesfaye, M., Silverstein, K. A. T., Mori, M., Umetsu, Y., et al. (2014). Central cell–derived peptides regulate early embryo patterning in flowering plants. *Science* 344, 168–173. doi: 10.1126/science.1243005

Couzigou, J.-M., and Combier, J.-P. (2016). Plant microRNAs: key regulators of root architecture and biotic interactions. *New Phytol.* 212, 22–35. doi: 10.1111/nph.14058

Couzigou, J.-M., Olivier André, O., Guillotin, B., Alexandre, M., and Combier, J.-P. (2016). Use of microRNA-encoded peptide miPEP172c to stimulate nodulation in soybean. *New Phytol.* 211, 379–381. doi: 10.1111/nph.13991

Crespo-Rivas, J. C., Guefrachi, I., Mok, K. C., Villaécija-Aguilar, J. A., Acosta-Jurado, S., Pierre, O., et al. (2016). *Sinorhizobium fredii* HH103 bacteroids are not terminally differentiated and show altered O-antigen in nodules of the inverted repeat-lacking clade legume *Glycyrrhiza uralensis*. *Environ. Microbiol.* 18, 2392–2404. doi: 10.1111/1462-2920.13101

Crockard, A., Bjourson, J., Dazzo, B., and Cooper, J. E. (2002). A white clover nodulin gene, *dd23b*, encoding a cysteine cluster protein, is expressed in roots during the very early stages of interaction with *Rhizobium leguminosarum* biovar *trifolii* and after treatment with chitolipooligosaccharide Nod factors. *J. Plant Res.* 115, 439–447. doi: 10.1007/s10265-002-0053-57

Czernic, P., Gully, D., Cartieaux, F., Moulin, L., Guefrachi, I., Patrel, D., et al. (2015). Convergent evolution of endosymbiont differentiation in dalbergioid and inverted repeat-lacking clade legumes mediated by nodule-specific cysteine-rich peptides. *Plant Physiol.* 169, 1254–1265. doi: 10.1104/pp.15.00584

de Bang, T. C., Lundquist, P. K., Dai, X., Boschiero, C., Zhuang, Z., Pant, P., et al. (2017). Genome-wide identification of *Medicago* peptides involved in macronutrient responses and nodulation. *Plant Physiol.* 175, 1669–1689. doi: 10.1104/pp.17.01096

Delay, C., Imin, N., and Djordjevic, M. A. (2013). *CEP* genes regulate root and shoot development in response to environmental cues and are specific to seed plants. *J. Exp. Bot.* 64, 5383–5394. doi: 10.1093/jxb/ert332

Delves, A. C., Mathews, A., Day, D. A., Carter, A. S., and Gresshoff, P. M. (1986). Regulation of the soybean-*Rhizobium* nodule symbiosis by shoot and root factors. *Plant Physiol.* 82, 588–590. doi: 10.1104/pp.82.2.588

Djordjevic, M. A., Mohd-Radzman, N. A., and Imin, N. (2015). Small-peptide signals that control nodule number, development, and symbiosis. *J. Exp. Bot.* 66, 5171–5181. doi: 10.1093/jxb/erv357

Doblas, V. G., Smakowska-luzan, E., Fujita, S., Alassimone, J., Barberon, M., Madalinski, M., et al. (2017). Root diffusion barrier control by a vasculature-derived peptide binding to the SGN3 receptor. *Science* 355, 280–284. doi: 10.1126/science.aaj1562

Durgo, H., Klement, E., Hunyadi-Gulyas, E., Szucs, A., Kereszt, A., Medzihradszky, K. F., et al. (2015). Identification of nodule-specific cysteine-rich plant peptides in endosymbiotic bacteria. *Proteomics* 15, 2291–2295. doi: 10.1002/pmic.201400385

Endre, G., Kereszt, A., Kevei, Z., Mihacea, S., Kaló, P., and Kiss, G. B. (2002). A receptor kinase gene regulating symbiotic nodule development. *Nature* 417, 962–966. doi: 10.1038/nature00841.1

Farkas, A., Maróti, G., Dürgő, H., Györgypál, Z., Lima, R. M., Medzihradszky, K. F., et al. (2014). *Medicago truncatula* symbiotic peptide NCR247 contributes to bacteroid differentiation through multiple mechanisms. *Proc. Natl. Acad. Sci. U.S.A.* 111, 5183–5188. doi: 10.1073/pnas.1404169111

Fedorova, M., Van De Mortel, J., Matsumoto, P. A., Cho, J., Town, C. D., Vandenbosch, K. A., et al. (2002). Genome-wide identification of nodule-specific transcripts in the model legume *Medicago truncatula*. *Plant Physiol.* 130, 519–537. doi: 10.1104/pp.006833

Fletcher, J. C., Brand, U., Running, M. P., Simon, R., and Meyerowitz, E. M. (1999). Signaling of cell fate decisions by *CLAVATA3* in *Arabidopsis* shoot meristems. *Science* 283, 1911–1914. doi: 10.1126/science.283.5409.1911

Frugier, F., Kosuta, S., Murray, J. D., Crespi, M., and Szczyglowski, K. (2008). Cytokinin: secret agent of symbiosis. *Trends Plant Sci.* 13, 115–120. doi: 10.1016/j.tplants.2008.01.003

Frühling, M., Albus, U., Hohnjec, N., Geise, G., Pühler, A., and Perlick, A. M. (2000). A small gene family of broad bean codes for late nodulins containing conserved cysteine clusters. *Plant Sci.* 152, 67–77. doi: 10.1016/S0168-9452(99)00219-218.

Ge, Z., Bergonci, T., Zhao, Y., Zou, Y., Du, S., Liu, M.-C., et al. (2017). *Arabidopsis* pollen tube integrity and sperm release are regulated by RALF-mediated signaling. *Science* 358, 1596–1600. doi: 10.1126/science.aao3642

Ghorbani, S., Lin, Y. C., Parizot, B., Fernandez, A., Njo, M. F., Van de Peer, Y., et al. (2015). Expanding the repertoire of secretory peptides controlling root development with comparative genome analysis and functional assays. *J. Exp. Bot.* 66, 5257–5269. doi: 10.1093/jxb/erv346

Guefrachi, I., Nagymihaly, M., Pislariu, C. I., Van de Velde, W., Ratet, P., Mars, M., et al. (2014). Extreme specificity of NCR gene expression in *Medicago truncatula*. *BMC Genomics* 15:712. doi: 10.1186/1471-2164-15-712

Guefrachi, I., Pierre, O., Timchenko, T., Alunni, B., Barriere, Q., Czernic, P., et al. (2015). *Bradyrhizobium* BclA is a peptide transporter required for bacterial differentiation in symbiosis with *Aeschynomene* legumes. *Mol. Plant Microbe Interact.* 28, 1155–1166. doi: 10.1094/MPMI-04-15-0094-R

Guo, P., Yoshimura, A., Ishikawa, N., Yamaguchi, T., Guo, Y., and Tsukaya, H. (2015). Comparative analysis of the RTFL peptide family on the control of plant organogenesis. *J. Plant Res.* 128, 497–510. doi: 10.1007/s10265-015-0703-701

Györgyey, J., Vaubert, D., Jimenez-Zurdo, J. I., Charon, C., Troussard, L., Kondorosi, A., et al. (2000). Analysis of *Medicago truncatula* nodule expressed sequence tags. *Mol. Plant Microbe Interact.* 13, 62–71. doi: 10.1094/MPMI.2000.13.1.62

Haag, A. F., Baloban, M., Sani, M., Kerscher, B., Pierre, O., Farkas, A., et al. (2011). Protection of *Sinorhizobium* against host cysteine-rich antimicrobial peptides is critical for symbiosis. *PLoS Biol.* 9:e1001169. doi: 10.1371/journal.pbio.1001169

Hanada, K., Higuchi-takeuchi, M., Okamoto, M., Yoshizumi, T., Shimizu, M., Nakaminami, K., et al. (2013). Small open reading frames associated with morphogenesis are hidden in plant genomes. *Proc. Natl. Acad. Sci. U.S.A.* 110, 2395–2400. doi: 10.1073/pnas.1213958110

Haruta, M., Sabat, G., Stecker, K., Minkoff, B. B., and Sussman, M. R. (2014). A peptide hormone and its receptor protein kinase regulate plant cell expansion. *Science* 343, 408–412. doi: 10.1126/science.1244454

Hellens, R. P., Brown, C. M., Chisnall, M. A. W., Waterhouse, P. M., and Macknight, R. C. (2016). The emerging world of small ORFs. *Trends Plant Sci.* 21, 317–328. doi: 10.1016/j.tplants.2015.11.005

Higashiyama, T. (2010). Peptide signaling in pollen–pistil interactions. *Plant Cell Physiol.* 51, 177–189. doi: 10.1093/pcp/pcq008

Horváth, B., Domonkos, Á., Kereszt, A., Szűcs, A., Ábrahám, E., Ayaydin, F., et al. (2015). Loss of the nodule-specific cysteine rich peptide, NCR169, abolishes symbiotic nitrogen fixation in the *Medicago truncatula dnf7* mutant. *Proc. Natl. Acad. Sci. U.S.A.* 112, 15232–15237. doi: 10.1073/pnas.1500777112

Hou, S., Wang, X., Chen, D., Yang, X., Wang, M., Turrà, D., et al. (2014). The secreted peptide PIP1 amplifies immunity through receptor-like kinase. *PLoS Pathog.* 10:e1004331. doi: 10.1371/journal.ppat.1004331

Huault, E., Laffont, C., Wen, J., Mysore, K. S., Ratet, P., and Frugier, F. (2014). Local and systemic regulation of plant root system architecture and symbiotic nodulation by a receptor- like kinase. *PLoS Genet.* 10:e1004891. doi: 10.1371/journal.pgen.1004891

Huffaker, A., Pearce, G., and Ryan, C. A. (2006). An endogenous peptide signal in *Arabidopsis* activates components of the innate immune response. *Proc. Natl. Acad. Sci. U.S.A.* 103, 10098–10103. doi: 10.1073/pnas.0603727103

Hunt, L., Gray, J. E., Court, F., Bank, W., and Sheffield, S. (2009). Report the signaling peptide EPF2 controls asymmetric cell divisions during stomatal development. *Curr. Biol.* 19, 864–869. doi: 10.1016/j.cub.2009.03.069

Igarashi, D., Tsuda, K., and Katagiri, F. (2012). The peptide growth factor, phytosulfokine, attenuates pattern-triggered immunity. *Plant J.* 71, 194–204. doi: 10.1111/j.1365-313X.2012.04950.x

Ikeuchi, M., Yamaguchi, T., Kazama, T., Ito, T., Horiguchi, G., and Tsukaya, H. (2011). ROTUNDIFOLIA4 regulates cell proliferation along the body axis in Arabidopsis shoot. *Plant Cell Physiol.* 52, 59–69. doi: 10.1093/pcp/pcq138

Imin, N., Mohd-radzman, N. A., Ogilvie, H. A., and Djordjevic, M. A. (2013). The peptide-encoding *CEP1* gene modulates lateral root and nodule numbers in *Medicago truncatula*. *J. Exp. Bot.* 64, 5395–5409. doi: 10.1093/jxb/ert369

Imin, N., Patel, N., Corcilius, L., and Payne, R. J. (2018). CLE peptide tri-arabinosylation and peptide domain sequence composition are essential for SUNN-dependent autoregulation of nodulation in *Medicago truncatula*. *New Phytol.* 218, 73–80. doi: 10.1111/nph.15019

Ishihara, H., Koriyama, H., Osawa, A., Zehirov, G., Yamaura, M., Kucho, K., et al. (2011). Characteristics of bacteroids in indeterminate nodules of the leguminous tree *Leucaena glauca*. *Microbes Environ.* 26, 156–159. doi: 10.1264/jsme2.ME11104

Jimenez-Zurdo, J. I., Frugier, F., Crespi, M. D., and Kondorosi, A. (2000). Expression profiles of 22 novel molecular markers for organogenetic pathways acting in alfalfa nodule development. *Mol. Plant Microbe Interact.* 13, 96–106. doi: 10.1094/MPMI.2000.13.1.96

Kaijalainen, S., Schroda, M., and Lindström, K. (2002). Cloning of nodule-specific cDNAs of *Galega orientalis*. *Physiol. Plant.* 114, 588–593. doi: 10.1034/j.1399-3054.2002.1140412.x

Karmakar, K., Kundu, A., Rizvi, A. Z., Dubois, E., Severac, D., Czernic, P., et al. (2018).Transcriptomic analyses with the progress of symbiosis in 'crack-entry' legume Arachis hypogaea highlight its contrast with 'Infection thread' adapted legumes. *bioRxiv* doi: 10.1101/251181

Kassaw, T., Nowak, S., Schnabel, E., and Frugoli, J. (2017). ROOT DETERMINED NODULATION1 is required for *M. truncatula* CLE12, but not CLE13, peptide signaling through the SUNN receptor kinase. *Plant Physiol.* 174, 2445–2456. doi: 10.1104/pp.17.00278

Kato, T., Kawashima, K., Miwa, M., Mimura, Y., Tamaoki, M., Kouchi, H., et al. (2002). Expression of genes encoding late nodulins characterized by a putative signal peptide and conserved cysteine residues is reduced in ineffective pea nodules. *Mol. Plant Microbe Interact.* 15, 129–137. doi: 10.1094/MPMI.2002.15.2.129

Kevei, Z., Vinardell, J. M., Kiss, G. B., Kondorosi, A., and Kondorosi, E. (2002). Glycine-rich proteins encoded by a nodule-specific gene family are implicated in different stages of symbiotic nodule development in *Medicago* spp. *Mol. Plant Microbe Interact.* 15, 922–931. doi: 10.1094/MPMI.2002.15.9.922

Kim, M., Chen, Y., Xi, J., Waters, C., Chen, R., and Wang, D. (2015). An antimicrobial peptide essential for bacterial survival in the nitrogen-fixing symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* 112, 15238–15243. doi: 10.1073/pnas.1500123112

Komori, R., Amano, Y., Ogawa-Ohnishi, M., and Matsubayashi, Y. (2009). Identification of tyrosylprotein sulfotransferase in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 106, 15067–15072. doi: 10.1073/pnas.0902801106

Kondorosi, E., Mergaert, P., and Kereszt, A. (2013). A paradigm for endosymbiotic life: cell differentiation of *Rhizobium* bacteria provoked by host plant factors. *Annu. Rev. Microbiol.* 67, 611–628. doi: 10.1146/annurev-micro-092412-155630

Kosslak, R. M., and Bohlool, B. B. (1984). Suppression of nodule development of one side of a root system of soybeans caused by prior inoculation of the other side. *Plant Physiol.* 75, 125–130. doi: 10.1104/pp.75.1.125

Krusell, L., Sato, N., Fukuhara, I., Koch, B. E., Grossmann, C., Okamoto, S., et al. (2011). The *Clavata2* genes of pea and *Lotus japonicus* affect autoregulation of nodulation. *Plant J.* 65, 861–871. doi: 10.1111/j.1365-313X.2010.04474.x

Kulkarni, G., Busset, N., Molinaro, A., Gargani, D., Chaintreuil, C., Silipo, A., et al. (2015). Specific hopanoid classes differentially affect free-living and symbiotic states of *Bradyrhizobium diazoefficiens*. *mBio* 6:e01251-15. doi: 10.1128/mBio.01251-1215

Kumagai, H., Kinoshita, E., Ridge, R. W., and Kouchi, H. (2006). RNAi knock-down of *ENOD40s* leads to significant suppression of nodule formation in *Lotus japonicus*. *Plant Cell Physiol.* 47, 1102–1111. doi: 10.1093/pcp/pcj081

Küster, H., Schröder, G., Frühling, M., Pich, U., Rieping, M., Schubert, I., et al. (1995). The nodule-specific *VfENOD-GRP3* gene encoding a glycine-rich early nodulin is located on chromosome I of *Vicia faba* L. and is predominantly expressed in the interzone II-III of root nodules. *Plant Mol. Biol.* 28, 405–421. doi: 10.1007/BF00020390

Laporte, P., Satiat-Jeunemaître, B., Velasco, I., Csorba, T., Van de Velde, W., Campalans, A., et al. (2010). A novel RNA-binding peptide regulates the establishment of the *Medicago truncatula–Sinorhizobium meliloti* nitrogen-fixing symbiosis. *Plant J.* 62, 24–38. doi: 10.1111/j.1365-313X.2009.04121.x

Lauressergues, D., Couzigou, J.-M., Clemente, H. S., Martinez, Y., Dunard, C., Bécard, G., et al. (2015). Primary transcripts of microRNAs encode regulatory peptides. *Nature* 520, 90–93. doi: 10.1038/nature14346

Lease, K. A., and Walker, J. C. (2006). The Arabidopsis unannotated secreted peptide database, a resource for plant peptidomics. *Plant Physiol.* 142, 831–838. doi: 10.1104/pp.106.086041

Lee, H., and Copeland, L. (1994). Ultrastructure of chickpea nodules. *Protoplasma* 185, 32–38. doi: 10.1007/BF01403686

Lee, J. S., Hnilova, M., Maes, M., Lin, Y. L., Putarjunan, A., Han, S., et al. (2015). Competitive binding of antagonistic peptides fine-tunes stomatal patterning. *Nature* 522, 439–443. doi: 10.1038/nature14561

Li, C., Wu, H., and Cheung, A. Y. (2016). FERONIA and her pals: functions and mechanisms. *Plant Physiol.* 171, 2379–2392. doi: 10.1104/pp.16.00667

Li, D., Kinkema, M., and Gresshoff, P. M. (2009). Autoregulation of nodulation (AON) in *Pisum sativum* (pea) involves signalling events associated with both nodule primordia development and nitrogen fixation. *J. Plant Physiol.* 166, 955–967. doi: 10.1016/j.jplph.2009.03.004

Lin, Y. H., Ferguson, B. J., Kereszt, A., and Gresshoff, P. M. (2010). Suppression of hypernodulation in soybean by a leaf-extracted, NARK- and Nod factor-dependent, low molecular mass fraction. *New Phytol.* 185, 1074–1086. doi: 10.1111/j.1469-8137.2009.03163.x

Lutkenhaus, J., and Addinall, S. G. (1997). Bacterial cell division and the Z ring. *Annu. Rev. Biochem.* 66, 93–116. doi: 10.1146/annurev.biochem.66.1.93

Makarewich, C. A., and Olson, E. N. (2017). Mining for micropeptides. *Trends Cell Biol.* 27, 685–696. doi: 10.1016/j.tcb.2017.04.006 685.

Mangeon, A., Junqueira, R. M., and Sachetto-Martins, G. (2010). Functional diversity of the plant glycine-rich proteins superfamily. *Plant Signal. Behav.* 5, 99–104. doi: 10.4161/psb.5.2.10336

Maróti, G., and Kondorosi, É. (2014). Nitrogen-fixing *Rhizobium*-legume symbiosis: are polyploidy and host peptide-governed symbiont differentiation general principles of endosymbiosis? *Front. Microbiol.* 5:326. doi: 10.3389/fmicb.2014.00326

Maróti, G., Kereszt, A., Kondorosi, É., and Mergaert, P. (2011). Natural roles of antimicrobial peptides in microbes, plants and animals. *Res. Microbiol.* 162, 363–374. doi: 10.1016/j.resmic.2011.02.005

Marx, H., Minogue, C. E., Jayaraman, D., Richards, A. L., Kwiecien, N. W., Siahpirani, A. F., et al. (2016). A proteomic atlas of the legume *Medicago truncatula* and its nitrogen-fixing endosymbiont *Sinorhizobium meliloti*. *Nat. Biotechnol.* 34, 1198–1205. doi: 10.1038/nbt.3681

Mathews, A., Kosslak, R. M., Sengupta-Gopalan, C., Appelbaum, E. R., Carroll, B. J., and Gresshoff, P. M. (1989). Biological characterization of root exudates and extracts from non-nodulating and supernodulating soybean mutants. *Mol. Plant Microbe Interact.* 2, 283–290. doi: 10.1094/MPMI-2-283

Matsubayashi, Y., Ogawa, M., and Morita, A. (2002). An LRR receptor kinase involved in perception of a peptide plant hormone, phytosulfokine. *Science* 296, 1470–1473. doi: 10.1126/science.1069607

Matsuzaki, Y., Ogawa-ohnishi, M., Mori, A., and Matsubayashi, Y. (2010). Secreted peptide signals required for maintenance of root stem cell niche in *Arabidopsis*. *Science* 329, 1065–1068. doi: 10.1126/science.1191132

Mecchia, M. A., Santos-fernandez, G., Duss, N. N., Somoza, S. C., Boisson-dernier, A., Gagliardini, V., et al. (2017). RALF4/19 peptides interact with LRX proteins to control pollen tube growth in *Arabidopsis*. *Science* 358, 1600–1603. doi: 10.1126/science.aao5467

Mergaert, P., Nikovics, K., Kelemen, Z., Maunoury, N., Vaubert, D., Kondorosi, A., et al. (2003). A novel family in *Medicago truncatula* consisting of more than 300 nodule-specific genes coding for small, secreted polypeptides with conserved cysteine motifs. *Plant Physiol.* 132, 161–173. doi: 10.1104/pp.102.018192

Mergaert, P., Uchiumi, T., Alunni, B., Evanno, G., Cheron, A., Catrice, O., et al. (2006). Eukaryotic control on bacterial cell cycle and differentiation in the *Rhizobium*-legume symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* 103, 5230–5235. doi: 10.1073/pnas.0600912103

Mikulász, K. R., Nagy, K., Bogos, B., Szegletes, Z., Kovács, E., Farkas, A., et al. (2016). Antimicrobial nodule-specific cysteine-rich peptides disturb the integrity of bacterial outer and inner membranes and cause loss of membrane potential. *Ann. Clin. Microbiol. Antimicrob.* 15:43. doi: 10.1186/s12941-016-0159-8

Miyazawa, H., Oka-Kira, E., Sato, N., Takahashi, H., Wu, G. J., Sato, S., et al. (2010). The receptor-like kinase KLAVIER mediates systemic regulation of nodulation and non-symbiotic shoot development in *Lotus japonicus*. *Development.* 137, 4317–4325. doi: 10.1242/dev.058891

Mohd-Radzman, N. A., Binos, S., Truong, T. T., Imin, N., and Mariani, M. (2015). Novel MtCEP1 peptides produced *in vivo* differentially regulate root development in *Medicago truncatula*. *J. Exp. Bot.* 66, 5289–5300. doi: 10.1093/jxb/erv008

Mohd-Radzman, N. A., Laffont, C., Ivanovici, A., Patel, N., Reid, D., Stougaard, J., et al. (2016). Different pathways act downstream of the CEP peptide receptor CRA2 to regulate lateral root and nodule development. *Plant Physiol.* 171, 2536–2548. doi: 10.1104/pp.16.00113

Montiel, J., Downie, J. A., Farkas, A., Bihari, P., Herczeg, R., Balint, B., et al. (2017). Morphotype of bacteroids in different legumes correlates with the number and type of symbiotic NCR peptides. *Proc. Natl. Acad. Sci. U.S.A.* 114, 5041–5046. doi: 10.1073/pnas.1704217114

Montiel, J., Szucs, A., Boboescu, I. Z., Gherman, V. D., Kondorosi, E., and Kereszt, A. (2016). Terminal bacteroid differentiation is associated with variable morphological changes in legume species Belonging to the inverted repeat-lacking clade. *Mol. Plant Microbe Interact.* 29, 210–219. doi: 10.1094/MPMI-09-15-0213-R

Mortier, V., De Wever, E., Vuylsteke, M., Holsters, M., and Goormachtig, S. (2012). Nodule numbers are governed by interaction between CLE peptides and cytokinin signaling. *Plant J.* 70, 367–376. doi: 10.1111/j.1365-313X.2011.04881.x

Mortier, V., Den Herder, G., Whitford, R., Van de Velde, W., Rombauts, S., D'haeseleer, K., et al. (2010). CLE peptides control *Medicago truncatula* nodulation. *Plant Physiol.* 153, 222–237. doi: 10.1104/pp.110.153718

Murphy, E., and De Smet, I. (2014). Understanding the RALF family: a tale of many species. *Trends Plant Sci.* 19, 664–671. doi: 10.1016/j.tplants.2014.06.005

Nakayama, T., Shinohara, H., Tanaka, M., Baba, K., Ogawa-Ohnishi, M., and Matsubayashi, Y. (2017). A peptide hormone required for Casparian strip diffusion barrier formation in *Arabidopsis* roots. *Science* 355, 284–286. doi: 10.1126/science.aai9057

Narita, N. N., Moore, S., Horiguchi, G., Kubo, M., Demura, T., Fukuda, H., et al. (2004). Overexpression of a novel small peptide ROTUNDIFOLIA4 decreases cell proliferation and alters leaf shape in *Arabidopsis thaliana*. *Plant J.* 38, 699–713. doi: 10.1111/j.1365-313X.2004.02078.x

Nishida, H., Tanaka, S., Handa, Y., Ito, M., Sakamoto, Y., Matsunaga, S., et al. (2018). A NIN-LIKE PROTEIN mediates nitrate-induced control of root nodule symbiosis in *Lotus japonicus*. *Nat. Commun.* 9:499. doi: 10.1038/s41467-018-02831-x

Nishimura, R., Hayashi, M., Wu, G. J., Kouchi, H., Imaizumi-Anraku, H., Murakami, Y., et al. (2002). HAR1 mediates systemic regulation of symbiotic organ development. *Nature* 420, 426–429. doi: 10.1038/nature01231

Nutman, P. S. (1952). Studies on the physiology of nodule formation: III. experiments on the excision of root-tips and nodules. *Ann. Bot.* 16, 79–103. doi: 10.1093/oxfordjournals.aob.a083304

Ogilvie, H. A., Imin, N., and Djordjevic, M. A. (2014). Diversification of the C-TERMINALLY ENCODED PEPTIDE (CEP) gene family in angiosperms, and evolution of plant-family specific CEP genes. *BMC Genomics* 15:870. doi: 10.1186/1471-2164-15-870

Ohkubo, Y., Tanaka, M., Tabata, R., Ogawa-Ohnishi, M., and Matsubayashi, Y. (2017). Shoot-to-root mobile polypeptides involved in systemic regulation of nitrogen acquisition. *Nat. Plants.* 3:17029. doi: 10.1038/nplants.2017.29

Okamoto, S., Ohnishi, E., Sato, S., Takahashi, H., Nakazono, M., Tabata, S., et al. (2009). Nod factor/nitrate-induced *CLE* genes that drive HAR1-mediated systemic regulation of nodulation. *Plant Cell Physiol.* 50, 67–77. doi: 10.1093/pcp/pcn194

Okamoto, S., Shinohara, H., Mori, T., Matsubayashi, Y., and Kawaguchi, M. (2013). Root-derived CLE glycopeptides control nodulation by direct binding to HAR1 receptor kinase. *Nat. Commun.* 4:2191. doi: 10.1038/ncomms3191

Okuda, S., Tsutsui, H., Shiina, K., Sprunck, S., Takeuchi, H., Yui, R., et al. (2009). Defensin-like polypeptide LUREs are pollen tube attractants secreted from synergid cells. *Nature* 458, 357–361. doi: 10.1038/nature07882

Oldroyd, G. E., and Downie, J. A. (2008). Coordinating nodule morphogenesis with rhizobial infection in legumes. *Annu. Rev. Plant Biol.* 59, 519–546. doi: 10.1146/annurev.arplant.59.032607.092839

Olsson, J. E., Nakao, P., Benbohlool, B., and Gresshoff, P. M. (1989). Lack of systemic suppression of nodulation in split root systems of supernodulating soybean (*Glycine max* [L] Merr) mutants. *Plant Physiol.* 90, 1347–1352. doi: 10.1104/pp.90.4.1347

Oono, R., and Denison, R. F. (2010). Comparing symbiotic efficiency between swollen versus nonswollen rhizobial bacteroids. *Plant Physiol.* 154, 1541–1548. doi: 10.1104/pp.110.163436

Oono, R., Schmitt, I., Sprent, J. I., and Denison, R. F. (2010). Multiple evolutionary origins of legume traits leading to extreme rhizobial differentiation. *New Phytol.* 187, 508–520. doi: 10.1111/j.1469-8137.2010.03261.x

Ördögh, L., Vörös, A., Nagy, I., Kondorosi, E., and Kereszt, A. (2014). Symbiotic plant peptides eliminate *Candida albicans* both *in vitro* and in an epithelial infection model and inhibit the proliferation of immortalized human cells. *BioMed Res. Int.* 2014:320796. doi: 10.1155/2014/320796

Pan, B. H., Sheng, J., Sun, W. N., Zhao, Y. H., Hao, P., and Li, X. (2013). OrysPSSP: a comparative platform for small secreted proteins from rice and other plants. *Nucleic Acids Res.* 41, D1192-D1198. doi: 10.1093/nar/gks1090

Park, S. J., and Buttery, B. R. (1989). Inheritance of nitrate-tolerant supernodulation in ems-induced mutants of common bean (*Phaseolus-vulgaris* L.). *J. Hered.* 80, 486–488. doi: 10.1093/oxfordjournals.jhered.a110907

Patel, N., Mohd-Radzman, N. A., Corcilius, L., Crossett, B., Connolly, A., Cordwell, S. J., et al. (2018). Diverse peptide hormones affecting root growth identified in the *Medicago truncatula* secreted peptidome. *Mol. Cell. Proteomics* 17, 160–174. doi: 10.1074/mcp.RA117.000168

Pearce, G., Moura, D. S., Stratmann, J., and Ryan, C. A. (2001). RALF, a 5-kDa ubiquitous polypeptide in plants, arrests root growth and development. *Proc. Natl. Acad. Sci. U.S.A.* 98, 12843–12847. doi: 10.1073/pnas.201416998

Penmetsa, R. V., and Cook, D. R. (1997). A legume ethylene-insensitive mutant hyperinfected by its rhizobial symbiont. *Science* 275, 527–530. doi: 10.1126/science.275.5299.527

Penmetsa, R. V., Frugoli, J. A., Smith, L. S., Long, S. R., and Cook, D. R. (2003). Dual genetic pathways controlling nodule number in *Medicago truncatula*. *Plant Physiol.* 131, 998–1008. doi: 10.1104/pp.015677

Penterman, J., Abo, R. P., De Nisco, N. J., Arnold, M. F. F., Longhi, R., Zanda, M., et al. (2014). Host plant peptides elicit a transcriptional response to control the *Sinorhizobium meliloti* cell cycle during symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* 111, 3561–3566. doi: 10.1073/pnas.1400450111

Pierce, M., and Bauer, W. D. (1983). A rapid regulatory response governing nodulation in soybean. *Plant Physiol.* 73, 286–290. doi: 10.1104/pp.73.2.286

Reid, D. E., Ferguson, B. J., Hayashi, S., Lin, Y. H., and Gresshoff, P. M. (2011). Molecular mechanisms controlling legume autoregulation of nodulation. *Ann. Bot.* 108, 789–795. doi: 10.1093/aob/mcr205

Röhrig, H., John, M., and Schmidt, J. (2004). Modification of soybean sucrose synthase by S-thiolation with ENOD40 peptide A. *Biochem. Biophys. Res. Commun.* 325, 864–870. doi: 10.1016/j.bbrc.2004.10.100

Röhrig, H., Schmidt, J., Miklashevichs, E., Schell, J., and John, M. (2002). Soybean *ENOD40* encodes two peptides that bind to sucrose synthase. *Proc. Natl. Acad. Sci. U.S.A.* 99, 1915–1920. doi: 10.1073/pnas.022664799

Roux, B., Rodde, N., Jardinaud, M. F., Timmers, T., Sauviac, L., Cottret, L., et al. (2014). An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. *Plant J.* 77, 817–837. doi: 10.1111/tpj.12442

Sachetto-Martins, G., Franco, L. O., and De Oliveira, D. E. (2000). Plant glycine-rich proteins: a family or just proteins with a common motif? *Biochim. Biophys. Acta* 1492, 1–14. doi: 10.1016/S0167-4781(00)00064-66

Santiago, J., Brandt, B., Wildhagen, M., Hohmann, U., Hothorn, L. A., Butenko, M. A., et al. (2016). Mechanistic insight into a peptide hormone signaling complex mediating floral. *eLife* 5:e15075. doi: /10.7554/eLife.15075

Sasaki, T., Suzaki, T., Soyano, T., Kojima, M., Sakakibara, H., and Kawaguchi, M. (2014). Shoot-derived cytokinins systemically regulate root nodulation. *Nat. Commun.* 5:4983. doi: 10.1038/ncomms5983

Satgé, C., Moreau, S., Sallet, E., Lefort, G., Auriac, M. C., Rembliere, C., et al. (2016). Reprogramming of DNA methylation is critical for nodule development in Medicago truncatula. *Nat. Plants.* 2, 16166. doi: 10.1038/nplants.2016.166.

Saur, I. M., Oakes, M., Djordjevic, M. A., and Imin, N. (2011). Crosstalk between the nodulation signaling pathway and the autoregulation of nodulation in *Medicago truncatula*. *New Phytol.* 190, 865–874. doi: 10.1111/j.1469-8137.2011.03738.x

Schnabel, E. L., Kassaw, T. K., Smith, L. S., Marsh, J. F., Oldroyd, G. E., Long, S. R., et al. (2011). The *ROOT DETERMINED NODULATION1* gene regulates nodule number in roots of Medicago truncatula and defines a highly conserved, uncharacterized plant gene family. *Plant Physiol.* 157, 328–340. doi: 10.1104/pp.111.178756

Schnabel, E., Journet, E. P., De Carvalho-Niebel, F., Duc, G., and Frugoli, J. (2005). The *Medicago truncatula SUNN* gene encodes a *CLV1*-like leucine-rich repeat receptor kinase that regulates nodule number and root length. *Plant Mol. Biol.* 58, 809–822. doi: 10.1007/s11103-005-8102-y

Schröder, G., Fruhling, M., Puhler, A., and Perlick, A. M. (1997). The temporal and spatial transcription pattern in root nodules of Vicia faba nodulin genes encoding glycine-rich proteins. *Plant Mol. Biol.* 33, 113–123. doi: 10.1023/A:1005779116272

Searle, I. R., Men, A. E., Laniya, T. S., Buzas, D. M., Iturbe-Ormaetxe, I., Carroll, B. J., et al. (2003). Long-distance signaling in nodulation directed by a CLAVATA1-like receptor kinase. *Science* 299, 109–112. doi: 10.1126/science.1077937

Sen, P., and Weaver, R. W. (1984). A basis for different rates of N2-fixation by the same strains of *Rhizobium* in peanut and cowpea root nodules. *Plant Sci. Lett.* 34, 239–246.

Silverstein, K. A. T., Graham, M. A., Paape, T. D., and Vandenbosch, K. A. (2005). Genome organization of more than 300 defensin-like genes in Arabidopsis. *Plant Physiol.* 138, 600–610. doi: 10.1104/pp.105.060079

Silverstein, K. A. T., Moskal, W. A., Wu, H. C., Underwood, B. A., Graham, M. A., Town, C. D., et al. (2007). Small cysteine-rich peptides resembling antimicrobial peptides have been under-predicted in plants. *Plant J.* 51, 262–280. doi: 10.1111/j.1365-313X.2007.03136.x

Sousa, C., Johansson, C., Charon, C., Manyani, H., Sautter, C., Kondorosi, A., et al. (2001). Translational and structural requirements of the early nodulin gene enod40, a short-open reading frame-containing RNA, for elicitation of a cell-specific growth response in the alfalfa root cortex. *Mol. Cell. Biol.* 21, 354–366. doi: 10.1128/Mcb.21.1.354-366.2001

Soyano, T., Hirakawa, H., Sato, S., Hayashi, M., and Kawaguchi, M. (2014). NODULE INCEPTION creates a long-distance negative feedback loop involved in homeostatic regulation of nodule organ production. *Proc. Natl. Acad. Sci. U.S.A.* 111, 14607–14612. doi: 10.1073/pnas.1412716111

Srivastava, R., Liu, J.-X., and Howell, S. H. (2008). Proteolytic processing of a precursor protein for a growth-promoting peptide by a subtilisin serine protease in Arabidopsis. *Plant J.* 56, 219–227. doi: 10.1111/j.1365-313X.2008.03598.x

Stegmann, M., Monaghan, J., Smakowska-Luzan, E., Rovenich, H., Lehner, A., Holton, N., et al. (2017). The receptor kinase FER is a RALF-regulated scaffold controlling plant immune signaling. *Science* 355, 287–289. doi: 10.1126/science.aal2541

Sugano, S. S., Shimada, T., Imai, Y., Okawa, K., Tamai, A., Mori, M., et al. (2010). Stomagen positively regulates stomatal density in *Arabidopsis*. *Nature* 463, 241–244. doi: 10.1038/nature08682

Tabata, R., Sumida, K., Yoshii, T., Ohyama, K., Shinohara, H., and Matsubayashi, Y. (2014). Perception of root-derived peptides by shoot LRR-RKs mediates systemic N-demand signaling. *Science* 346, 343–346. doi: 10.1126/science.1257800

Takahara, M., Magori, S., Soyano, T., Okamoto, S., Yoshida, C., Yano, K., et al. (2013). TOO MUCH LOVE, a novel kelch repeat-containing F-box protein, functions in the long-distance regulation of the legume-*Rhizobium* symbiosis. *Plant Cell Physiol.* 54, 433–447. doi: 10.1093/pcp/pct022

Takahashi, F., Suzuki, T., Osakabe, Y., Betsuyaku, S., Kondo, Y., Dohmae, N., et al. (2018). A small peptide modulates stomatal control via abscisic acid in long-distance signalling. *Nature* 556, 235–238. doi: 10.1038/s41586-018-0009-2

Taleski, M., Imin, N., and Djordjevic, M. A. (2018). CEP peptide hormones: key players in orchestrating nitrogen-demand signalling, root nodulation, and lateral root development. *J. Exp. Bot.* 69, 1829–1836. doi: 10.1093/jxb/ery037

Tavormina, P., De Coninck, B., Nikonorova, N., De Smet, I., and Cammue, B. P. A. (2015). The plant peptidome: an expanding repertoire of structural features and biological functions. *Plant Cell* 27, 2095–2118. doi: 10.1105/tpc.15.00440

Tiricz, H., Szucs, A., Farkas, A., Pap, B., Lima, R. M., Maroti, G., et al. (2013). Antimicrobial nodule-specific cysteine-rich peptides induce membrane depolarization-associated changes in the transcriptome of *Sinorhizobium meliloti*. *Appl. Environ. Microbiol.* 79, 6737–6746. doi: 10.1128/AEM.01791-1713

Trujillo, D. I., Silverstein, K. A. T., and Young, N. D. (2014). Genomic characterization of the LEED..PEEDs, a gene family unique to the *Medicago* lineage. *G3* 4, 2003–2012. doi: 10.1534/g3.114.011874

Valdivia, E. R., Chevalier, D., Sampedro, J., Taylor, I., Niederhuth, C. E., and Walker, J. C. (2012). *DVL* genes play a role in the coordination of socket cell recruitment and differentiation. *J. Exp. Bot.* 63, 1405–1412. doi: 10.1093/jxb/err378

van Brussel, A. A. N., Tak, T., Boot, K. J. M., and Kijne, J. W. (2002). Autoregulation of root nodule formation: signals of both symbiotic partners studied in a split-root system of *Vicia sativa* subsp. *nigra*. *Mol. Plant Microbe Interact.* 15, 341–349. doi: 10.1094/Mpmi.2002.15.4.341

Van de Velde, W., Zehirov, G., Szatmari, A., Debreczeny, M., Ishihara, H., Kevei, Z., et al. (2010). Plant peptides govern terminal differentiation of bacteria in symbiosis. *Science* 327, 1122–1126. doi: 10.1126/science.1184057

Wan, X., Hontelez, J., Lillo, A., Guarnerio, C., Van De Peut, D., Fedorova, E., et al. (2007). *Medicago truncatula ENOD40-1* and *ENOD40-2* are both involved in nodule initiation and bacteroid development. *J. Exp. Bot.* 58, 2033–2041. doi: 10.1093/jxb/erm072

Wang, C., Yu, H. X., Zhang, Z. M., Yu, L. L., Xu, X. S., Hong, Z. L., et al. (2015). Phytosulfokine is involved in positive regulation of *Lotus japonicus* nodulation. *Mol. Plant Microbe Interact.* 28, 847–855. doi: 10.1094/Mpmi-02-15-0032-R

Wang, D., Griffitts, J., Starker, C., Fedorova, E., Limpens, E., Ivanov, S., et al. (2010). A nodule-specific protein secretory pathway required for nitrogen-fixing symbiosis. *Science* 327, 1126–1129. doi: 10.1126/science.1184096

Wang, Q., Liu, J., Li, H., Yang, S., Kormoczi, P., Kereszt, A., et al. (2018). Nodule-specific cysteine-rich peptides negatively regulate nitrogen-fixing symbiosis in a strain-specific manner in *Medicago truncatula*. *Mol. Plant Microbe Interact.* 31, 240–248. doi: 10.1094/MPMI-08-17-0207-R

Wang, Q., Yang, S., Liu, J., Terecskei, K., Abraham, E., Gombar, A., et al. (2017). Host-secreted antimicrobial peptide enforces symbiotic selectivity in *Medicago truncatula*. *Proc. Natl. Acad. Sci. U.S.A.* 114, 6854–6859. doi: 10.1073/pnas.1700715114

Wang, Y. N., Wang, L. X., Zou, Y. M., Chen, L., Cai, Z. M., Zhang, S. L., et al. (2014). Soybean miR172c targets the repressive AP2 transcription factor NNC1 to activate *ENOD40* expression and regulate nodule initiation. *Plant Cell* 26, 4782–4801. doi: 10.1105/tpc.114.131607

Wen, J. Q., Lease, K. A., and Walker, J. C. (2004). DVL, a novel class of small polypeptides: overexpression alters *Arabidopsis* development. *Plant J.* 37, 668–677. doi: 10.1111/j.1365-313X.2003.01994.x

Whitford, R., Fernandez, A., Tejos, R., Perez, A. C., Kleine-Vehn, J., Vanneste, S., et al. (2012). GOLVEN secretory peptides regulate auxin carrier turnover during plant gravitropic responses. *Dev. Cell* 22, 678–685. doi: 10.1016/j.devcel.2012.02.002

Wojciechowski, M. F., Lavin, M., and Sanderson, M. J. (2004). A phylogeny of legumes (Leguminosae) based on analysis of the plastid *matK* gene resolves many well-supported subclades within the family. *Am. J. Bot.* 91, 1846–1862. doi: 10.3732/ajb.91.11.1846

Wojciechowski, M. F., Sanderson, M. J., Steele, K. P., and Liston, A. (2000). "Molecular phylogeny of the 'Temperate Herbaceous Tribes' of papilionoid legumes: a supertree approach," in *Advances in Legumes Systematics*, ed. P. S. Herendeen and A. Bruneau (Richmond: Royal Botanic Gardens, Kew), 277–298.

Wopereis, J., Pajuelo, E., Dazzo, F. B., Jiang, Q. Y., Gresshoff, P. M., De Bruijn, F. J., et al. (2000). Short root mutant of *Lotus japonicus* with a dramatically altered symbiotic phenotype. *Plant J.* 23, 97–114. doi: 10.1046/j.1365-313x.2000.00799.x

Yang, S. M., Wang, Q., Fedorova, E., Liu, J. G., Qin, Q. L., Zheng, Q. L., et al. (2017). Microsymbiont discrimination mediated by a host-secreted peptide in *Medicago truncatula. Proc. Natl. Acad. Sci. U.S.A.* 114, 6848–6853. doi: 10.1073/pnas.1700460114

Zhang, H., Hu, Z., Lei, C., Zheng, C., Wang, J., Shao, S., et al. (2018). A plant phytosulfokine peptide initiates auxin-dependent immunity through cytosolic Ca2 + signaling in tomato. *Plant Cell* 30, 652–667. doi: 10.1105/tpc.17.00537

Comparative Transcriptomic Analysis of Two Actinorhizal Plants and the Legume *Medicago truncatula* Supports the Homology of Root Nodule Symbioses and Is Congruent With a Two-Step Process of Evolution in the Nitrogen-Fixing Clade of Angiosperms

Kai Battenberg[1]*, Daniel Potter[1], Christine A. Tabuloc[2], Joanna C. Chiu[2] and Alison M. Berry[1]

[1] Department of Plant Sciences, University of California, Davis, Davis, CA, United States, [2] Department of Entomology and Nematology, University of California, Davis, Davis, CA, United States

Root nodule symbiosis (RNS) is a symbiotic interaction established between angiosperm hosts and nitrogen-fixing soil bacteria in specialized organs called root nodules. The host plants provide photosynthate and the microsymbionts supply fixed nitrogen. The origin of RNS represents a major evolutionary event in the angiosperms, and understanding the genetic underpinnings of this event is of major economic and agricultural importance. Plants that engage in RNS are restricted to a single angiosperm clade known as the nitrogen-fixing clade (NFC), yet occur in multiple lineages scattered within the NFC. It has been postulated that RNS evolved in two steps: a gain-of-predisposition event occurring at the base of the NFC, followed by a gain-of-function event in each host plant lineage. Here, we first explore the premise that RNS has evolved from a single common background, and then we explore whether a two-step process better explains the evolutionary origin of RNS than either a single-step process, or multiple origins. We assembled the transcriptomes of root and nodule of two actinorhizal plants, *Ceanothus thyrsiflorus* and *Datisca glomerata*. Together with the corresponding published transcriptomes of the model legume *Medicago truncatula*, the gene expression patterns in roots and nodules were compared across the three lineages. We found that orthologs of many genes essential for RNS in the model legumes are expressed in all three lineages, and that the overall nodule gene expression patterns were more similar to each other than expected by random chance, a finding that supports a common evolutionary background for RNS shared by the three lineages. Moreover, phylogenetic analyses suggested that a substantial portion of the genes experiencing selection pressure changes at the base of the NFC also experienced additional changes at the base of each host plant lineage. Our results (1) support the

occurrence of an event that led to RNS at the base of the NFC, and (2) suggest a subsequent change in each lineage, most consistent with a two-step origin of RNS. Among several conserved functions identified, strigolactone-related genes were down-regulated in nodules of all three species, suggesting a shared function similar to that shown for arbuscular mycorrhizal symbioses.

INTRODUCTION

Root nodule symbiosis is a symbiotic interaction established between certain groups of angiosperm hosts and nitrogen-fixing soil bacteria that are housed in specialized organs called root nodules. The host plants provide photosynthate to their microsymbionts, and in turn the microsymbionts provide fixed nitrogen to their host plants.

This symbiotic relationship enables host species to thrive in nutrient-poor soils, and thus these RNS hosts play a major role in terrestrial ecosystems as pioneer plants (Chapin et al., 1994). Moreover, legumes play key roles in agriculture, where plant-based biological nitrogen fixation accounts for as much as 10% of the total nitrogen fixed in the world (Herridge et al., 2008; Fowler et al., 2013). Thus, understanding the genetic underpinnings of the origin of RNS not only provides insight into a major biological event in the evolution of angiosperms, but is also of major economic and agricultural importance.

RNS occurs in ten families of angiosperms within four orders: Fabales, Rosales, Cucurbitales, and Fagales. Molecular phylogenetic studies have revealed that these four orders, which were previously considered to be distantly related within the angiosperms (Cronquist, 1988), together form a clade known as the nitrogen-fixing clade (NFC) (Soltis et al., 1995). Within each of the four orders, RNS occurs in a subset of the families, which are phylogenetically scattered within each order (Swensen, 1996), and within each family, RNS is restricted to a subset of the genera.

There are several possible hypotheses regarding the evolutionary origin of RNS that can explain this restricted (found only among orders of the NFC) yet scattered (found only in some families and genera of the NFC) distribution of RNS hosts (Doyle, 2011). The single-origin hypothesis proposes that the capability of forming nitrogen-fixing root nodules evolved once in the MRCA of the NFC and was subsequently lost multiple times independently in the currently non-fixing lineages. The multiple-origin hypothesis proposes that the evolution of RNS occurred independently at least six and as many as ten times (Doyle, 2011). The two-step hypothesis postulates that a predisposition for, i.e., propensity to subsequently gain, RNS was first gained at the base of the NFC, which was then followed by a gain of function that occurred independently in the aforementioned six to ten different lineages (Soltis et al., 1995; Swensen, 1996; Werner et al., 2014).

The two-step hypothesis has been supported by phylogenetic analysis based on the distribution of RNS hosts within the NFC (Werner et al., 2014). The hypothesis can parsimoniously explain the restricted yet scattered phylogenetic distribution of RNS hosts, but raises the question, what was the genetic basis of the "predisposition" to RNS?

In addition to the phylogenetic evidence, several shared cellular, molecular, and genetic characteristics of RNS hosts in different lineages are consistent with a common evolutionary predisposition to RNS: (1) all RNSs result in a stable accommodation of the microsymbiont within the host cells (Pawlowski and Demchenko, 2012); (2) homologs of many essential genes required for RNS in the model legumes *Medicago truncatula* and *Lotus japonicus* have been shown to be expressed in the nodules of non-legume RNS hosts (Hocher et al., 2011; Demina et al., 2013); (3) calcium oscillation, an early host physiological response, is induced during initiation of RNS in both legume and non-legume hosts (Granqvist et al., 2015).

Multiple lineages sharing these aforementioned series of characters support the common descent of RNS across multiple lineages (i.e., supports single-origin and two-step hypotheses over multiple-origin hypothesis), but some morphological, cellular, and molecular characteristics are clearly distinct in different lineages of RNS hosts (Pawlowski and Demchenko, 2012), which could favor a multiple-origin hypothesis. Most notably, RNS hosts in two families (Fabaceae and Cannabaceae) associate with rhizobia as their microsymbiont, while the hosts in the remaining eight families associate with members of the actinobacterial genus *Frankia*. These eight families are collectively called the actinorhizal plants. Most *Frankia* genomes lack the genes coding for Nod factor, the signaling molecule responsible for the initiation of RNS in the model legumes (Oldroyd, 2013), but homologs for the *nodABC* genes have been identified in some groups of *Frankia* (Persson et al., 2015; Nguyen et al., 2016). Similarly, some legume hosts can be nodulated by rhizobia without the Nod factor signaling pathway (Okazaki et al., 2015). Therefore, a range of different mechanisms for initiating RNS must exist among the legumes and the actinorhizal plants.

Transcriptomes generated via RNA-seq represent a powerful source of data that can provide a comprehensive set of characters. In the present study, we have used evidence from comparative transcriptomics and molecular evolutionary analyses to test the competing hypotheses for the evolutionary origin(s) of RNS. To this end, we assembled the root nodule and root transcriptomes of two actinorhizal plant species, *Ceanothus thyrsiflorus* (Rhamnaceae, Rosales) and *Datisca glomerata* (Datiscaceae, Cucurbitales) and compared them to

---

Abbreviations: ABA, abscisic acid; AM, arbuscular mycorrhizal; CDS, coding sequence; *dN/dS*, ratio of non-synonymous to synonymous mutations; IAA, indole-3-acetic acid; MRCA, most recent common ancestor; NFC, nitrogen-fixing clade; NO, nitric oxide; RNS, root nodule symbiosis.

published transcriptomes of *M. truncatula* (Fabaceae, Fabales) (Roux et al., 2014). We conducted differential gene expression analysis between nodules and roots to determine a set of genes that are root- or nodule-enhanced for each species. Then, to allow interspecific comparisons, phylogeny-based orthology prediction was conducted across the three species and other taxa that are either members or close outgroups of the NFC.

We first explored if RNS has evolved from a single common evolutionary event that occurred at the base of the NFC, regardless of whether this event gave rise to the function or the predisposition of RNS. To test the homology of RNS in the three species, we first focused on the presence/absence of orthologs in *C. thyrsiflorus* and *D. glomerata* for 19 genes required for the initiation and development of root nodules in the model legumes *M. truncatula* or *L. japonicus*, because both the single-origin hypothesis and the two-step hypothesis would require that at least some of the orthologs of genes required for nodulation would be shared among all RNS hosts. Determining orthology is an improvement with respect to previous studies that have identified homologs of genes required for RNS in legumes in several actinorhizal plants (Hocher et al., 2011; Demina et al., 2013), as orthologs are a subset of homologs that are most likely to be functionally equivalent according to ortholog conjecture (Koonin, 2005; Altenhoff et al., 2012). A recent study has taken a similar approach, comparing nodule gene expression in *Parasponia* spp. the only non-legume host that can establish RNS with rhizobia, to *M. truncatula* (van Velzen et al., 2018).

Second, we compared the overall expression patterns of orthologous genes across the three species between roots and nodules. Our assumption was that a significant degree of similarity across the three species, which are known to belong to three different lineages of RNS hosts (Doyle, 2011), is not expected if they had completely independent origins of RNS. Thus, sharing a significant degree of similarity would refute the multiple-origin hypothesis, and indicate a single common ancestor for RNS. Gene expression analysis and orthology predictions allowed us to identify a set of genes that showed a consistent pattern of differential expression across the three species, which we designated as the core set of genes for RNS.

Then, we have further explored which of the three competing hypotheses best explains the origin of RNS, particularly whether a single-origin or a two-step process better explains the evolutionary basis of RNS. For this, we employed a model-based phylogenetic test. We focused on the fact that the three hypotheses each assume different timing(s) for the gain-of-predisposition or gain-of-function event leading to RNS in the evolutionary history of the NFC. We assumed that gain of a new function would result in a change in selection pressure on that gene, which should be reflected in the average ratio of non-synonymous to synonymous mutations (*dN*/*dS*) (Hurst, 2002) in the coding regions of the gene. Thus, for each set of orthologs, we tested when (if ever) each member gained a new function by calculating *dN*/*dS* on key branches of their respective phylogenetic trees.

RESULTS

Transcriptome Assembly and Completeness

Illumina HiSeq platform generated a total of 376,256,549 and 303,175,236 paired-end reads (150 bp) for *C. thyrsiflorus* and *D. glomerata,* respectively. After cleaning, >97% (342,814,393 reads) and >91% (296,864,053 reads) of these pairs were kept for the transcriptome assembly (**Supplementary Table S1**, NCBI BioProject ID: PRJNA422680). Trinity (Grabherr et al., 2011) generated root + nodule transcriptomes for *C. thyrsiflorus* and *D. glomerata* with 675,696 transcripts (480,254 genes) and 444,766 transcripts (309,847 genes), respectively. After curating (three screenings and collapsing of alleles), the cleaned transcriptome of *C. thyrsiflorus* and *D. glomerata* consisted of 15,245 and 15,448 genes, respectively (**Supplementary Table S2**, NCBI BioProject ID: PRJNA422680). In both transcriptomes, N50, the length of the shortest transcript that covers 50% of the transcriptome, was >2.1 kb, average was >1.8 kb, and median was >1.6 kb. Over 76% and 81% of the cleaned reads mapped onto the cleaned transcriptomes of *C. thyrsiflorus* and *D. glomerata,* respectively (**Supplementary Table S2**).

Within each of the three root + nodule transcriptomes (*C. thyrsiflorus*, *D. glomerata*, and *M. truncatula*), BUSCO found 62.4%, 82.2%, and 93.6% of the plant-universal orthologs (**Supplementary Table S3**). In the *M. truncatula* genome, BUSCO found 95.5% of the plant-universal genes (**Supplementary Table S3**). The number of ECs found for each biosynthetic pathway based on KEGG-KAAS server searches overall showed similar results across the three transcriptomes and the genome: of the 410 separate pathways listed in KEGG, only 15 of them differed in their enzyme counts by more than 2 between any transcriptome or genome (**Table 1** and **Supplementary Table S4**). Transcripts coding for 11 out of the 14 enzymes in the sequiterpenoid/triterpenoid pathway were expressed in *M. truncatula* while only five and four enzymes were found in *C. thyrsiflorus* and *D. glomerata* transcriptomes, respectively; for the isoflavonoid biosynthesis pathway, 9 of the 14 enzymes were expressed in *M. truncatula*, while only one each was expressed in *C. thyrsiflorus* and *D. glomerata*.

Annotation, Differential Gene Expression

The root + nodule transcriptomes of *C. thyrsiflorus* and *D. glomerata* and the genome of *M. truncatula* were annotated (by KO, EC, and/or GO) for 88.1%, 85.0%, and 60.9% of the genes, respectively (**Supplementary Table S5**). No abnormalities in the mean variance trend of gene expressions were found by limma (Ritchie et al., 2015), and there was a clear distinction in gene expression patterns between the roots and the nodules for all three species (**Supplementary Image S1**). Differential gene expression analysis found 19.2% (2,932 root-enhanced and 745 nodule-enhanced), 34.2% (2,726 root-enhanced and 2,550 nodule-enhanced), and 34.6% (4,819 root-enhanced and 4,965 nodule-enhanced) of genes significantly differentially

TABLE 1 | Enzymes in selected KEGG biosynthetic pathways represented in transcriptomes and *M. truncatula* genome, showing more than two enzyme differences per pathway among the three hosts.

KEGG pathway ID	KEGG pathway name	Total enzymes in KEGG pathway	*C. thyrsiflorus* (T)	*D. glomerata* (T)	*M. truncatula* (T)	*M. truncatula* (G)
130	Ubiquinone and other terpenoid-quinone biosynthesis	47	19	23	24	24
230	Purine metabolism	106	57	61	62	62
240	Pyrimidine metabolism	75	36	41	40	41
540	Lipopolysaccharide biosynthesis	33	8	12	13	13
670	One carbon pool by folate	24	9	13	13	13
710	Carbon fixation in photosynthetic organisms	27	19	21	23	23
730	Thiamine metabolism	29	14	16	18	18
790	Folate biosynthesis	30	11	14	15	15
860	Porphyrin and chlorophyll metabolism	87	23	30	32	32
906	Carotenoid biosynthesis	40	22	25	25	25
909	Sesquiterpenoid and triterpenoid biosynthesis	77	5	4	11	11
941	Flavonoid biosynthesis	19	16	13	13	15
943	Isoflavonoid biosynthesis	14	1	1	9	9
1200	Carbon metabolism	221	92	93	96	97
1210	2-Oxocarboxylic acid metabolism	58	30	29	32	32

T, transcriptome; G, genome.

expressed in the *C. thyrsiflorus*, *D. glomerata*, and *M. truncatula* transcriptomes, respectively.

In the GO enrichment analysis, although particular annotated functions (GO terms) were enriched in roots or nodules of an individual species, when the results were compared across the three species, no common or similar patterns of enrichment were found (**Supplementary Table S6**).

Validation of RNA-seq With qPCR of cDNA Library

A subset of the gene expression patterns calculated based on RNA-seq were validated by qPCR. Of the 14 genes tested (seven genes in each of the two actinorhizal species), 13 showed similar log-fold changes in expression as the RNA-seq results. Of these 13 genes, eight of which were <2 log2-fold change apart from the RNA-seq results. *NRT1.8* in *C. thyrsiflorus* showed an expression pattern in conflict with the RNA-seq results (**Supplementary Table S7**).

Orthology Predictions

Orthologs for each gene in the *C. thyrsiflorus* and *D. glomerata* transcriptomes and the *M. truncatula* genome were predicted independently using OrthoReD (Battenberg et al., 2017), a tree-based orthology prediction tool. OrthoReD predicted orthologs for all genes within the *C. thyrsiflorus* transcriptome, *D. glomerata* transcriptome, and the *M. truncatula* genome (total of 81,587 genes). Because OrthoReD predicts a set of orthologs for each gene independently, the predicted sets of orthologs for different genes are not always mutually exclusive. Thus, in order to avoid analyzing the same gene multiple times for the same analysis, orthologous sets were merged into 27,367 mutually exclusive groups (MergedOrthoGroups).

We note here that these orthologs are predicted based on the specific dataset that was analyzed. Including a more complete dataset for example by replacing the *C. thyrsiflorus* and *D. glomerata* transcriptomes with genomes, or by including more species may result in different predictions.

Presence and Expression Patterns of RNS Pathway Genes

We selected 19 genes that are required for the development of nodules in *M. truncatula* and *L. japonicus* to specifically search for orthologs in *C. thyrsiflorus* and *D. glomerata* transcriptomes. Of these, OrthoReD identified orthologs for 17 and 18 genes in *C. thyrsiflorus* and *D. glomerata* transcriptomes, respectively (**Table 2**). These orthologs spanned from the most upstream step (e.g., *NFR1*, *SYMRK*) to *NIN*, one of the key transcription factors for RNS (**Figure 1**). OrthoReD found transcripts with high sequence similarity (putative homologs) at its intermediate step in orthology prediction for all 19 genes including *FNR5* and *HMGR* in *C. thyrsiflorus* and *HMGR* in *D. glomerata* (**Table 2**).

Three of the 19 RNS pathway genes were found to be universally nodule-enhanced: *SYMREM* (Lefebvre et al., 2010), a remorin that is known to interact with *NFR1*, in the most upstream portion of the RNS pathway; *NIN* (Schauser et al., 1999), a key transcription factor that is one of the more downstream RNS pathway genes required for nodule organogenesis; and *RPG* (Arrighi et al., 2008), a gene required for the proper growth and regulation of the infection thread that is key to the establishment of RNS in *M. truncatula* (**Table 2**). Fourteen of the RNS pathways genes were found to have an ortholog in all three taxa, but there was not a significant universal enhancement of gene expression in this group of genes (i.e., passing both criteria for fold change and *p*-value) in a single tissue (root vs. nodule). At the same time, none of these genes showed a significant conflicting pattern

TABLE 2 | Fold changes (nodule over root) and adjusted p-values of symbiotic pathway genes.

Gene Name	MergedOrtho Group	*C. thyrsiflorus* ortholog Gene ID	Log2FC	P-value	*D. glomerata* ortholog Gene ID	Log2FC	P-value	*M. truncatula/L. japonicus* ortholog Gene ID	Log2FC	P-value	GenBank ID
NFR1	–	Ct_CtTrNR03429	–0.37	0.151	Dg_DgTrNR03578	2.75	0.000	Mt_Medtr5g 086130	–2.10	0.000	AY372406.1
					Dg_DgTrNR03579	–1.39	0.000				
					Dg_DgTrNR03580	2.54	0.000				
NFR5	–	Ct_CtTrNR01660, Ct_CtTrNR02016, Ct_CtTrNR02556, Ct_CtTrNR03429, Ct_CtTrNR03658, Ct_CtTrNR10088, Ct_CtTrNR13226, Ct_CtTrNR13284, Ct_CtTrNR14238, Ct_CtTrNR14293	Dg_DgTrNR09631	–1.52	0.000	Mt_Medtr5g 019040	–2.43	0.000	DQ496250.1			
SYMRK	MergedOrthoGroup005636	Ct_CtTrNR07134	–0.19	0.413	Dg_DgTrNR12393	0.46	0.000	Mt_Medtr5g030920	0.60	0.002	AF491998.1
SYMREM	MergedOrthoGroup006520	Ct_CtTrNR08408	2.10	0.003	Dg_DgTrNR05672	5.65	0.000	Mt_Medtr5g010590	–0.22	0.775	–
								Mt_Medtr8g097320	9.88	0.000	JQ061257.1
HMGR	–	Ct_CtTrNR06569, Ct_CtTrNR06801, Ct_CtTrNR08716, Ct_CtTrNR12229	Dg_DgTrNR07804, Dg_DgTrNR14611, Dg_DgTrNR14612	Mt_Medtr5g 087550	–1.35	0.000	XM_003617018.2				
NUP85	MergedOrthoGroup003212	Ct_CtTrNR03893	–0.87	0.020	Dg_DgTrNR02299	0.12	0.132	Mt_Medtr1g006690	–0.19	0.665	XM_0035888338.2
NUP133	MergedOrthoGroup009404	Ct_CtTrNR12969	–0.20	0.263	Dg_DgTrNR15290	0.58	0.000	Lj_Lj2g3v3337540	–	–	AJ890251.1
								Lj_Lj0g3v0050449	–	–	–
								Mt_Medtr5g097260	–0.48	0.064	
CASTOR	MergedOrthoGroup008762	Ct_CtTrNR11902	–0.39	0.071	Dg_DgTrNR14656	0.37	0.001	Mt_Medtr7g117580	–0.23	0.431	FJ974130.1
POLLUX	MergedOrthoGroup009011	Ct_CtTrNR12292	–0.43	0.090	Dg_DgTrNR14917	–0.83	0.000	Mt_Medtr2g005870	1.61	0.000	AY497771.1
CCAMK	MergedOrthoGroup006145Sub001	Ct_CtTrNR07875	–0.49	0.064	Dg_DgTrNR14486	0.02	0.857	Mt_Medtr8g043970	0.59	0.004	AY496049.1
CYCLOPS	MergedOrthoGroup005459Sub002	Ct_CtTrNR09813	0.25	0.294	Dg_DgTrNR00459	1.60	0.000	Mt_Medtr5g026850	3.18	0.000	EF117279.1

(Continued)

TABLE 2 | Continued

Gene Name	MergedOrtho Group	C. thyrsiflorus ortholog			D. glomerata ortholog			M. truncatula/L. japonicus ortholog			
		Gene ID	Log₂FC	P-value	Gene ID	Log₂FC	P-value	Gene ID	Log₂FC	P-value	GenBank ID
NSP1	MergedOrtho Group009317	Ct_CtTrNR12805	−0.10	0.595	Dg_DgTrNR11911	0.96	0.000	Mt_Medtr3g085310	0.00	1.000	–
NSP2	MergedOrtho Group000090	Ct_CtTrNR00101	0.50	0.232	Dg_DgTrNR00363	−1.38	0.000	Mt_Medtr8g020840	1.62	0.000	AJ972478.1
								Mt_Medtr3g072710	−1.34	0.000	AJ832138.1
ERN1	MergedOrtho Group010414	Ct_CtTrNR15219	−2.56	0.003	Dg_DgTrNR01580	0.99	0.001	Mt_Medtr5g058860	−2.59	0.000	–
					Dg_DgTrNR02845	−0.66	0.006	Mt_Medtr6g029180	−2.53	0.000	–
NIN	MergedOrtho Group008966	Ct_CtTrNR12221	5.48	0.003	Dg_DgTrNR10008	4.95	0.000	Mt_Medtr7g085810	−0.39	0.373	EU038802.2
		Ct_CtTrNR12222	6.47	0.004	Dg_DgTrNR00885	8.30	0.000	Mt_Medtr5g099060	7.42	0.000	FJ719774.1
					Dg_DgTrNR00886	7.56	0.000				
LHK1	MergedOrtho Group007561	Ct_CtTrNR09972	1.16	0.156	Dg_DgTrNR03717	−0.92	0.000	Lj_Lj0g3v0108729	–	–	DQ848999.1
								Mt_Medtr2g067240	0.00	1.000	–
								Mt_Medtr5g044100	0.00	1.000	–
								Mt_Medtr8g106150	−0.22	0.524	–
PIR	–	Ct_CtTrNR13403	0.13	0.688	Dg_DgTrNR01709	0.17	0.048	Lj_Lj1g3v5020900	–	–	AM946364.1
					Dg_DgTrNR07598	−0.51	0.000				
RPG	MergedOrtho Group006667 Sub001	Ct_CtTrNR08628	−0.78	0.022	Dg_DgTrNR05600	8.90	0.000	Mt_Medtr1g090807	6.45	0.000	DQ854741.1
		Ct_CtTrNR14741	5.14	0.003	Dg_DgTrNR07056	−0.25	0.102				
CERBERUS	MergedOrtho Group002490 Sub002	Ct_CtTrNR11435	0.24	0.265	Dg_DgTrNR05473	0.87	0.000	Mt_Medtr1g090320	1.25	0.000	EU926661.1

For each symbiotic pathway gene, the corresponding MergedOrthoGroup and the ortholog(s) are shown. Each ortholog also has its log2-fold change and adjusted p-values are indicated. The yellow cells are not orthologs, but are putative homologs found by OrthoReD. All genes included in this table were selected from Kouchi et al. (2010) and/or Oldroyd (2013).
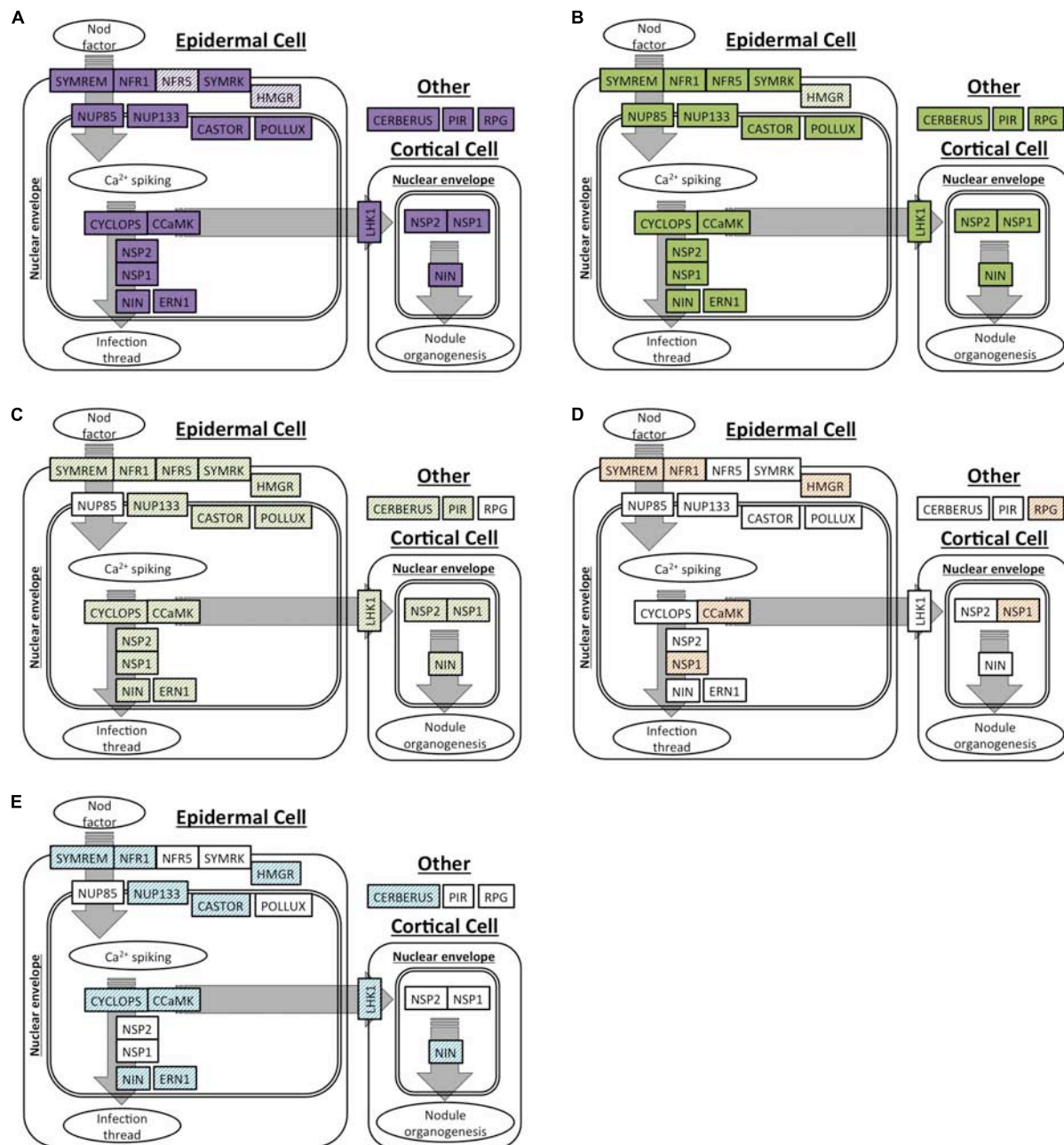
FIGURE 1 | Orthologs or homologs of genes involved in the initiation of root nodule symbiosis identified in four actinorhizal hosts. A spatial and sequential model of genes (indicated as boxes) involved in the initiation of root nodule symbiosis based on *M. truncatula* and/or *L. japonicus*. Genes are diagrammed at locations where the model expects the gene product to be localized in the root tissue. Orthologs are shown as solid colors; homologs are shaded. **(A,B)** Genes with orthologs detected in this study: **(A)** *C. thyrsiflorus*, and **(B)** *D. glomerata*. **(C–E)** Respectively: homologs detected in *D. glomerata* according to Demina et al. (2013); homologs detected in *A. glutinosa* according to Hocher et al. (2011); homologs detected in *C. glauca* according to Hocher et al. (2011).

of gene expression (nodule-enhanced in one ortholog while root-enhanced in another) across different species (**Table 2**).

Expression Similarity Analysis

We used two approaches to compare differential (root vs. nodule) gene expression patterns among the three species, on a pairwise basis. We parsed out 3,894 whole and 3,033 subsets

of MergedOrthoGroups (total of 6,927 MergedOrthoGroups) as the representative MergedOrthoGroups for these analyses (**Supplementary Table S8**). Of the 19 RNS pathway genes, 15 were represented within these MergedOrthoGroups (**Table 2**). The representative MergedOrthoGroups included 49.2%, 51.0%, and 38.7% of the root + nodule transcriptomes of *C. thyrsiflorus*, *D. glomerata*, and *M. truncatula*, respectively. Among the

6,927 representative MergedOrthoGroups, 3,365 (48.6%) of them were enhanced in at least one tissue in at least one of the three species (**Figure 2**). Among these, there were 103 core MergedOrthoGroups that showed universal enhancement either in the nodule (*n* = 51) or in the root (*n* = 52) across the three species (**Figure 2**). The nodule-enhanced core MergedOrthoGroup included sets of orthologs for *SYMREM*, *NIN*, and *RPG* as described above in section "*Presence of RNS Pathway Genes and Their Gene Expression Patterns.*"

First, based on the Pearson correlation coefficient, we found that the gene expression fold changes between roots and nodules were weakly to moderately correlated (*r* = 0.18–0.38 for the three pairwise species comparisons) when all 6,927 representative MergedOrthoGroups were analyzed. The correlations were stronger for MergedOrthoGroups whose expression was enhanced in roots compared to nodules, of both species being compared (*r* = 0.25–0.47), and strongest for MergedOrthoGroups whose expression was enhanced in the nodules of both species being compared (*r* = 0.43–0.60) (**Figure 3**). The correlations were statistically significant in all cases (*p* < 0.01) (**Figure 3**).

Second, the degree and significance of overall similarity in gene expression between each pair of species was assessed using a scale, which we call the dissonance score (see section "*Expression Similarity Analysis*" in Materials and Methods for detail and **Supplementary Image S2** for a summary). Dissonance scores across all MergedOrthoGroups for all pairwise species comparison showed that the expression patterns of nodules compared to roots in the three species were more similar to each other than by random chance in all pairwise comparisons across the three species. In fact the dissonance score from
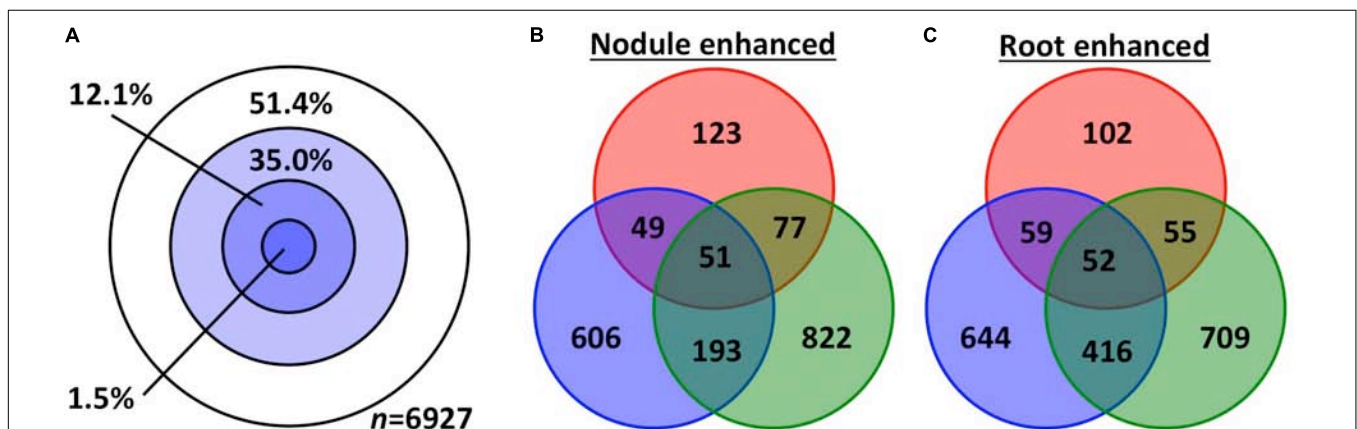
FIGURE 2 | Gene expression overlap among the representative MergedOrthoGroups. **(A)** Concentric circles each represent MergedOrthoGroups that showed significant enhancement in different number of species: The innermost circle represents the MergedOrthoGroups showing enhancement in all three species; the next circle represents significant enhancement in two species then one, and the outermost while circle is for MergedOrthoGroups showing no significant enhancement in any species. **(B,C)** The number of MergedOrthoGroups that showed significant enhancement in nodules **(B)** or in roots **(C)**. The three colors red, blue, and green correspond to *C. thyrsiflorus*, *D. glomerata*, and *M. truncatula*, respectively.
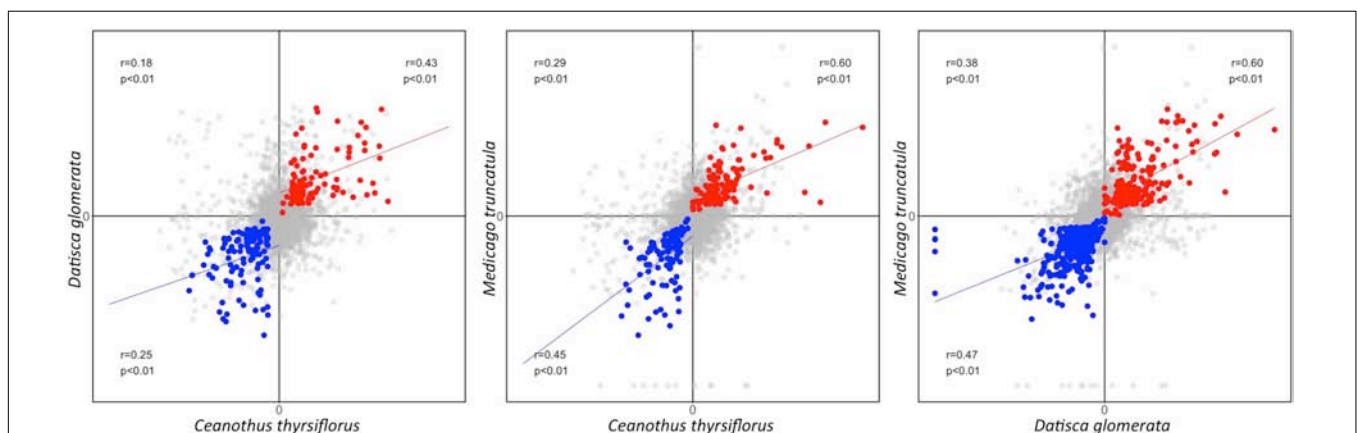
FIGURE 3 | Pearson correlation of gene expression fold change between two species. Gene expression fold change (nodule over root) is plotted as a comparison between each pair of species. The Pearson correlation coefficient calculated from all representative MergedOrthoGroup (in gray) is displayed in the second quadrant along with the corresponding *p*-value. The Pearson correlation coefficient of root co-enhanced MergedOrthoGroups (in blue) and nodule co-enhanced MergedOrthoGroups (in red) are displayed in the third and the first quadrant, respectively. Blue and the red lines each show the regression lines of root and nodule co-enhanced MergedOrthoGroups, respectively.

any of the 10,000 random permutations conducted for each pairwise comparison was never lower than the observed value ($p < 0.0001$).

Gene Ontology (GO) Enrichment Analysis of the Core MergedOrthoGroups

Gene Ontology terms (GO terms) provide gene annotations that permit grouping into integrated biological processes (Ashburner et al., 2000; Gene Ontology Consortium, 2017). A few GO terms were universally enriched in the comparison of core MergedOrthoGroups across the three host species: nitrate transport (GO:0015706), and metabolic processes of zeatin and trans-zeatin (GO:0033397, GO:0033398, GO:0033400, GO:0033466) were universally enriched in nodules; while strigolactone biosynthesis (GO:1901600, GO:1901601), secondary shoot formation (GO:0010223, GO:0010346), and cellular response to NO (GO:0034614, GO:0071731, GO:0071732, and GO:1902170) were universally enriched in roots (**Table 3** and **Supplementary Image S3**).

Enrichment of GO terms related to a nitrate transporter in the nodule was due to universal up-regulation of a set of orthologs (MergedOrthoGroup008002) corresponding to *NRT1.8* in *Arabidopsis*, a low-affinity nitrate transporter hypothesized to export nitrate from xylem to xylem parenchyma cell (Li et al., 2010).

The set of orthologs responsible for the enrichment of GO terms related to cytokinin (MergedOrthoGroup005893) was *CYP735A*, a cytokinin hydroxylase that catalyzes the biosynthesis of trans-zeatin in *Arabidopsis* (Takei et al., 2004). We also found another set of orthologs (MergedOrthoGroup008876) coding for IPT (isopentenyl transferase) (Azarakhsh et al., 2015) up-regulated in the nodules of *C. thyrsiflorus* and *D. glomerata*, but down-regulated in the *M. truncatula* nodule.

Enrichment of GO terms related to strigolactone biosynthesis in the roots was due to universal up-regulation of two sets of orthologs: *MAX4* (MergedOrthoGroup000413), a likely carotenoid dioxygenase (Sorefan et al., 2003), *D27* (MergedOrthoGroup006746), a carotenoid isomerase (van Zeijl et al., 2015a). A third set of universally up-regulated orthologs, *DLK2* (MergedOrthoGroup000413), has been classified under the strigolactone GO term, but was recently shown not to be involved in strigolactone signaling (Waters et al., 2012; Bennett et al., 2016; Vegh et al., 2017). These three sets of orthologs were also annotated to be related to secondary shoot formation.

Enrichment of the GO term related to NO in the roots was due to universal up-regulation of two sets of orthologs: MergedOrthoGroup006609, a gene coding for a histidine kinase (*AHK5*) (Iwama et al., 2007), and MergedOrthoGroup001248Sub001 which did not have a well-documented member.

dN/dS Analysis

We used a molecular phylogenetic approach to model changes in selection pressure along key branches in the phylogenies of selected MergedOrthoGroups. These consisted of 3,894 MergedOrthoGroups that (1) contained at least one member

from each of *C. thyrsiflorus*, *D. glomerata*, and *M. truncatula*, and (2) were identical to a set of orthologs predicted from a single gene with respect to all 15 species, were used in this analysis. These included 58 core genes (showing a consistent pattern of differential expression among the three target nitrogen-fixing species) as well as 9 of the 19 known RNS pathway genes: *SYMRK*, *NUP85*, *NUP133*, *CASTOR*, *POLLUX*, *NSP2*, *ERN1*, *NIN*, and *LHK1*.

For each MergedOrthoGroup, the likelihoods of up to three different hypothesis-based scenarios (SINGLE, MULTI, and TWOSTEP), each of which predicts a different timing for a change in dN/dS, were compared to a NULL scenario, which assumes a constant dN/dS throughout the gene phylogeny (**Figure 4**).

For each of the 3,894 MergedOrthoGroups, we tested whether any of the three hypothesis-based scenario (MULTI, SINGLE, or TWOSTEP) was significantly better than the NULL scenario (see section "dN/dS Analysis" in Materials and Methods for detail and **Figure 4** for a summary). Among the 3,894 MergedOrthoGroups, 2,668 did not reject the NULL scenario and the remaining 1,226 did not reject one or more of the alternative scenarios (MULTI, SINGLE, or TWOSTEP).

Focusing on the branch at the base of the NFC (branch *a* in **Figure 4**), 1,166 MergedOrthoGroups did not reject scenarios that predict a change in selection pressure here (SINGLE and/or TWOSTEP). Of these, 364 MergedOrthoGroups (31.2%) rejected neither SINGLE nor TWOSTEP, 6 (0.5%) rejected TWOSTEP, and 796 (68.3%) rejected SINGLE. For branches leading to *C. thyrsiflorus*, *D. glomerata*, and the legumes (branches *b*, *c*, *d* in **Figure 4**, respectively), 1,220 MergedOrthoGroups supported scenarios that predict changes in selection pressure along those branches (MULTI and/or TWOSTEP). Of these, 1,150 MergedOrthoGroups (94.3%) rejected neither MULTI nor TWOSTEP, 60 (4.9%) rejected TWOSTEP, and 10 (0.8%) rejected MULTI (**Table 4**).

The 1,226 MergedOrthoGroups included a subset of core MergedOrthoGroups (9 universally nodule-enhanced, 10 universally root-enhanced). The pattern of scenarios not rejected in the analysis of the core MergedOrthoGroups was similar to that of the overall results: Among the core MergedOrthoGroups that did not reject a change of selection pressure at branch *a* (i.e., either SINGLE or TWOSTEP), a majority (76.5%) rejected the SINGLE scenario. In addition, among the core MergedOrthoGroups that did not reject changes in selection pressure at branches *b*, *c*, and *d,* most (89.5%) rejected neither the MULTI nor the TWOSTEP scenario (**Table 4**). Among the RNS pathway genes, only *ERN1* rejected the NULL scenario and supported MULTI and TWOSTEP scenarios.

DISCUSSION

Newly Assembled Transcriptomes Are of High Quality

Both *C. thyrsiflorus* and *D. glomerata* root + nodule transcriptomes scored well in multiple measures of quality (e.g., proper insert size, long N50, and high % fragment mapped)

TABLE 3 | Enriched Gene Ontology (GO) terms among the root- and the nodule-enhanced core MergedOrthoGroups that are shared across the three species.

GO ID	GO term	P-value (*C. thyrsiflorus*)	P-value (*D. glomerata*)	P-value (*M. truncatula*)
(A) Nodule-enhanced GO terms shared across the three species.				
GO:0009690	Cytokinin metabolic process	4.9.E-05	2.5.E-04	4.2.E-04
GO:0015706	Nitrate transport	5.8.E-05	2.3.E-03	3.0.E-03
GO:0034754	Cellular hormone metabolic process	2.0.E-04	4.4.E-04	1.1.E-03
GO:0015698	Inorganic anion transport	4.0.E-04	2.9.E-03	5.4.E-03
GO:0006820	Anion transport	4.2.E-04	5.9.E-04	2.1.E-04
GO:0042445	Hormone metabolic process	8.1.E-04	3.8.E-04	1.1.E-03
GO:0010817	Regulation of hormone levels	3.4.E-03	2.8.E-03	4.5.E-03
GO:0010028	Xanthophyll cycle	3.8.E-03	1.8.E-05	1.6.E-05
GO:0015882	L-Ascorbic acid transport	3.8.E-03	1.8.E-05	1.6.E-05
GO:0033397	Zeatin metabolic process	3.8.E-03	1.8.E-05	1.6.E-05
GO:0033398	Zeatin biosynthetic process	3.8.E-03	1.8.E-05	1.6.E-05
GO:0033400	*Trans*-zeatin metabolic process	3.8.E-03	1.8.E-05	1.6.E-05
GO:0033466	*Trans*-zeatin biosynthetic process	3.8.E-03	1.8.E-05	1.6.E-05
GO:0010101	Post-embryonic root morphogenesis	6.7.E-03	9.2.E-04	1.3.E-03
GO:0010102	Lateral root morphogenesis	0.00665	0.00092	0.00132
GO:0051180	Vitamin transport	0.00767	0.000054	0.00019
(B) Root-enhanced GO terms shared across the three species.				
GO:1901334	Lactone metabolic process	1.4.E-06	4.3.E-07	5.7.E-11
GO:1901336	Lactone biosynthetic process	1.4.E-06	4.3.E-07	5.7.E-11
GO:1901600	Strigolactone metabolic process	1.4.E-06	4.3.E-07	5.7.E-11
GO:1901601	Strigolactone biosynthetic process	1.4.E-06	4.3.E-07	5.7.E-11
GO:0034756	Regulation of iron ion transport	2.8.E-06	3.4.E-04	1.4.E-04
GO:0071732	Cellular response to nitric oxide	2.8.E-06	5.9.E-06	1.4.E-09
GO:1902170	Cellular response to reactive nitrogen species	2.8.E-06	1.3.E-05	5.8.E-09
GO:0071281	Cellular response to iron ion	3.9.E-06	4.5.E-03	2.5.E-04
GO:0071731	Response to nitric oxide	4.8.E-06	8.8.E-06	5.8.E-09
GO:0010039	Response to iron ion	9.6.E-06	6.4.E-03	4.3.E-04
GO:0010223	Secondary shoot formation	6.1.E-05	3.7.E-05	2.6.E-08
GO:0010346	Shoot axis formation	6.1.E-05	3.7.E-05	2.6.E-08
GO:0071241	Cellular response to inorganic substance	6.1.E-05	1.2.E-03	4.9.E-07
GO:0001763	Morphogenesis of a branching structure	1.1.E-04	6.9.E-05	3.7.E-08
GO:0010959	Regulation of metal ion transport	1.5.E-04	2.3.E-03	2.8.E-03
GO:0042445	Hormone metabolic process	2.7.E-04	4.2.E-05	3.1.E-03
GO:0034614	Cellular response to reactive oxygen species	9.5.E-04	4.4.E-04	3.6.E-07
GO:0016106	Sesquiterpenoid biosynthetic process	1.9.E-03	3.7.E-05	2.1.E-08
GO:0006714	Sesquiterpenoid metabolic process	0.00242	0.000069	0.00000012
GO:1905393	Plant organ formation	0.00271	0.00484	0.000056
GO:0006812	Cation transport	0.00378	0.00605	0.00175
GO:0055114	Oxidation–reduction process	0.00409	0.00986	0.00451
GO:0071369	Cellular response to ethylene stimulus	0.00433	0.00404	0.00015
GO:0016114	Terpenoid biosynthetic process	0.00457	0.00056	0.0000089
GO:0034599	Cellular response to oxidative stress	0.0057	0.0026	0.0000052
GO:0055085	Transmembrane transport	0.00588	0.0012	0.00216
GO:0006811	Ion transport	0.00659	0.00782	0.00047
GO:0006721	Terpenoid metabolic process	0.00659	0.00099	0.000017

throughout the assembly process. After the assembly, KEGG annotation found the two transcriptomes to have similar numbers of ECs to the *M. truncatula* transcriptome (and even to the *M. truncatula* genome) for most of the pathways (**Table 1** and **Supplementary Table S4**). Both transcriptomes were annotated (by KO, EC, and/or GO) for >85% of the transcripts. Moreover,

BUSCO found 62.4%, 82.2% of the plant-universal orthologs for *C. thyrsiflorus* and *D. glomerata,* respectively. Furthermore, all the qPCR validated the RNA-seq results except for one gene (**Supplementary Table S7**). This all together indicate that the root + nodule transcriptomes of *C. thyrsiflorus* and *D. glomerata* are of high quality.

FIGURE 4 | Different scenarios tested based on competing hypotheses of the origin of RNS. A phylogenetic tree of the NFC and its outgroup. The branches *a* through *d* designate the base of the NFC (*a*), *C. thyrsiflorus* (*b*), *D. glomerata* (*c*), and the legumes (*d*), respectively. The table below indicates the three hypothesis-based scenarios and where each scenario expects a change in *dN/dS* in the phylogeny. Species without a family label each belong to a separate family: *A. thaliana* (Brassicaceae), *M. esculenta* (Euphorbiaceae), *P. trichocarpa* (Salicaceae), *C. thyrsiflorus* (Rhamnaceae), *C. sativus* (Cucurbitaceae), and *D. glomerata* (Datiscaceae). Mal., Malpighiales. The topology of the tree is based on Angiosperm Phylogeny Group III (APG, 2009).

Nodule Gene Expression Patterns in *C. thyrsiflorus*, *D. glomerata*, and *M. truncatula* Are More Similar to Each Other Than Would be Expected by Random Chance

Our analysis supports the homology (shared by common descent) of RNS among the three plant species based on multiple lines of evidence. First, orthologs for most of the 19 RNS pathway genes required for the proper nodulation in *M. truncatula* were present and expressed in the nodules of *C. thyrsiflorus* (17 out of 19 found) and *D. glomerata* (18 out of 19 found) (**Figure 1**). Transcripts with high sequence similarity were found,

even for the few genes that did not have an ortholog predicted (**Figure 1** and **Table 2**). Results of this orthology-based analysis strengthens previous homology-based reports (inferred based on high scores in BLAST searches) throughout the RNS pathway in *D. glomerata*, *Alnus glutinosa* (Betulaceae, Fagales) and *Casuarina glauca* (Casuarinaceae, Fagales) (Hocher et al., 2011; Demina et al., 2013); and are consistent with reports demonstrating functional equivalents (presumed orthologs) of a specific member of the RNS pathway, such as *SYMRK* (Markmann et al., 2008), *CCaMK* (Svistoonoff et al., 2013), and *NIN* (Clavijo et al., 2015), across multiple lineages. This is the first time, to our knowledge, that the entire pathway has been detected comprehensively in the context of a phylogenetically based orthology framework. It is also important to emphasize that the orthologs identified in our analysis included *NFR1*, *NSP1*, *ERN1*, and *NIN*, genes not shared with the more ancient AM symbiosis (Oldroyd, 2013). The presence and the expression of orthologs across these three species indicates that their existence predates the NFC, which is required for the RNS to share a common function and common origin across them.

In all pairwise comparisons of gene expression patterns between two species, we found a moderate to strong positive correlation for genes that were significantly enhanced in the nodules ($r = 0.43$–0.60); and there was a weak to moderate positive correlation for genes that were significantly enhanced in the roots ($r = 0.25$–0.47). The correlation was much weaker when all genes (including genes that are not necessarily presumed to be involved in RNS) were compared at once across two transcriptomes ($r = 0.18$–0.38) (**Figure 2**). While it is common to describe Pearson correlation coefficients of $r = 0.20$–0.39, $r = 0.40$–0.59, $r = 0.6$–0.79, as indicating weak, moderate, and strong correlation, respectively, meaningful interpretations of particular values depend on the context in which they were obtained, in this case, comparisons of transcriptomes across a considerable phylogenetic distance. For example, the Pearson correlation coefficient between the gene expression patterns between two sets of 20 plants in one species, *Arabidopsis thaliana,* was 0.81 (Kempema et al., 2007). By contrast, the comparisons made in our study are between pairs of different plant species that have been diverging for nearly 100 million years (Bell et al., 2010); thus, we consider that the values of $r = 0.43$–0.60 found for nodule-enhanced genes indicate a high degree of similarity and conservation, compared with 0.18–0.38 for all genes.

Finally, permutation tests based on the dissonance scores indicated that the overall gene expression patterns of the nodules in the three RNS hosts tested are more similar to each other than expected by random chance ($p < 0.0001$).

These results strongly support the homology of RNS in all three lineages, i.e., that their similarity is due to common descent. It is possible that other factors, such as similarity in the age of these tissues across the three transcriptomes, may have contributed to the similarity of the gene expression patterns. An increased spatiotemporal resolution for *C. thyrsiflorus* and *D. glomerata*, as obtained for *M. truncatula* through a time course transcriptome (Larrainzar et al., 2015) or tissue-specific transcriptome (Roux et al., 2014) would provide further clarity.

TABLE 4 | Number and fractions of MergedOrthoGroups that show a change in dN/dS at branches *a*, *b*, *c*, and *d* in **Figure 4**.

All MergedOrthoGroups			Reject TWOSTEP	Reject neither	Reject SINGLE	Total
	a	Count	6	364	796	1,166
		Percentage	0.5%	31.2%	68.3%	
			Reject TWOSTEP	Reject neither	Reject MULTI	Total
	b, c, d	Count	60	1150	10	1,220
		Percentage	4.9%	94.3%	0.8%	
Core MergedOrthoGroups			Reject TWOSTEP	Reject neither	Reject SINGLE	Total
	a	Count	0	4	13	17
		Percentage	0.0%	23.5%	76.5%	
			Reject TWOSTEP	Reject neither	Reject MULTI	Total
	b, c, d	Count	2	17	0	19
		Percentage	10.5%	89.5%	0.0%	

Features of RNS Conserved Among the Three Lineages

Orthologs of 17 or 18 of the 19 RNS pathway genes were expressed in the root + nodule transcriptomes of *C. thyrsiflorus* and *D. glomerata,* respectively, and three of them (*SYMREM, NIN,* and *RPG*) were universally nodule-enhanced (**Table 2**). The up-regulation of *SYMREM, NIN,* and RPG in the nodules was also found in *A. glutinosa,* and *C. glauca* (Hocher et al., 2011), and is consistent with what was found for *NIN* in *D. glomerata* (Demina et al., 2013). With the inclusion of *C. thyrsiflorus,* we now show that an up-regulation of *SYMREM, NIN,* and *RPG* in the nodules of RNS hosts is found in all four orders within the NFC.

Because the initiation and establishment of RNS consists of multiple developmental stages (Pawlowski and Demchenko, 2012; Roux et al., 2014), a high degree of spatiotemporal resolution is crucial to accurately trace the expression pattern of a gene (Roux et al., 2014; Larrainzar et al., 2015): Genes up-regulated only at a specific stage of nodule development may not be up-regulated within a transcriptome that is inclusive of the entire nodule. Thus, genes that are up-regulated (relative to root) in the whole nodule are expected either (1) to be so strongly nodule-enhanced in a given stage that they are detected as up-regulated even after averaging expression values over the whole nodule, or (2) to be nodule-enhanced throughout the process of nodulation. The latter expression pattern has been well documented in the case of *NIN* (Schauser et al., 1999). For the remaining genes, whose orthologs were not universally up-regulated in either nodules or roots (some were up-regulated in two hosts and non-significant in the third), a higher resolution of space and time would be helpful for accurate comparisons across species.

Among the processes found to be universally enriched in either nodules or in roots among the core genes, we focused on the following four processes based on the potential relevance to RNS.

Nitrate Transporter

Orthologs of *NRT1.8*, a low-affinity nitrate transporter, were significantly up-regulated in the nodules of all three species of RNS hosts. In *A. thaliana*, *NRT1.8* is up-regulated by nitrate, and is hypothesized to export nitrate from xylem conducting cells to xylem parenchyma (Li et al., 2010). In *M. truncatula*, 50% of the expression of *NRT1.8* ortholog was located in Zone I (Roux et al., 2014), which corresponds to the nodule meristem. However, there are no functional conducting xylem elements in root meristems of higher plants, which would be the tissue equivalent of Zone I of the root nodule. Moreover, nitrate is known to suppress nodulation in *Ceanothus* (Thomas and Berry, 1989) and in *L. japonicus* (Soyano et al., 2014). In *L. japonicus*, nitrate and *NIN* work antagonistically against each other (Soyano et al., 2014): *NIN* expression suppresses the activation of nitrate-induced genes while nitrate suppresses the activation of *NF-YA1* and *NF-YB1*, the known targets of *NIN*. Taken together, these lines of evidence suggest that the function of *NRT1.8* in the root nodules is not related to nitrate. While the understanding of *NRT1.8* is still limited, the *NRT1* family proteins are also capable of transporting signaling compounds and phytohormones: In *A. thaliana*, *NRT1.1*, *NRT1.2*, and *NRT1.10* are involved in the transport of IAA, ABA, and glucosinolates, respectively (Chiba et al., 2015).

Strigolactone Biosynthesis

The genetic dissection of RNS has revealed that the pathway for its establishment shares many of its genes with the more ancient pathway to form AM symbiosis (Oldroyd, 2013). It is thus of substantial importance that *D27* and *MAX4*, genes coding for the enzymes of the first and the third steps of strigolactone biosynthesis (Ruyter-Spira et al., 2013) are apparently suppressed in the nodules of all three nitrogen-fixing hosts (**Table 3B**). Strigolactone has a number of functions in plants (Besserer et al., 2006; Gomez-Roldan et al., 2008), particularly as a key regulator in root development, with possible regulatory feedback interactions with auxin transport and metabolism (Kapulnik and Koltai, 2014). Strigolactone is also a signaling molecule that initiates AM symbiosis by stimulating the branching and growth of hyphae of the AM fungi (Besserer et al., 2006). Within the mature AM symbiotic tissue, however, strigolactone is down-regulated (Kapulnik and Koltai, 2014).

A similar pattern may be inferred for RNS. Strigolactone has been shown to play a role in promoting nodulation in *Pisum*

sativum, in that a strigolactone mutant formed fewer nodules than wild-type (Foo and Davies, 2011), although, strigolactone is not required for nodulation, since nodulation does occur in the mutant. The influence of strigolactone on nodulation in *P. sativum* seems to be limited to the early stage of infection thread formation (McAdam et al., 2017).

Thus, the down-regulation of strigolactone biosynthesis in the nodule tissue of three phylogenetically distinct hosts that was observed in this study could be a function derived from AM symbiosis, to inhibit a portion of the root system from forming further infection sites, or to limit a stage of root development from further growth, allowing for nodule development or maintenance.

The enrichment of the GO term related to secondary shoot formation in the root transcriptome is likely related to the pattern of strigolactone gene expression observed, since *MAX4* plays a role in regulating branching in the shoot (Sorefan et al., 2003).

Cytokinin Biosynthesis

Up-regulation of genes associated with cytokinin biosynthesis and/or metabolism during the establishment of RNS is well documented in *L. japonicus* and *M. truncatula* (Tirichine et al., 2007; van Zeijl et al., 2015b; Gamas et al., 2017). Cytokinin in legume RNS is a major signal to the cortical cells to initiate nodule organogenesis by induction of *NIN* through the activation of a cytokinin receptor *LHK1* (Tirichine et al., 2007). Since the up-regulation of *NIN* has been found in the nodules of *A. glutinosa*, *C. glauca*, and *D. glomerata* (Hocher et al., 2011; Demina et al., 2013), *NIN* is considered to play a major role in RNS in actinorhizal plants as well as in the legumes. We found a universal up-regulation of *CYP735A*, a gene coding for the biosynthesis of trans-zeatin. *IPT* was also up-regulated in the nodules in *C. thyrsiflorus* and *D. glomerata*. Although *IPT* was down-regulated overall in *M. truncatula*, this is likely explained by the fact that >98% of *M. truncatula IPT* ortholog (Medtr2g022140) expression was restricted to the pre-infected cells (Roux et al., 2014).

Nitric Oxide Response

The cellular response to NO was universally down-regulated in the nodules of the three host plants in this study. Two sets of orthologs responsible for this pattern were detected, but only one had an assigned name: *AHK5*, a histidine kinase originally identified in *A. thaliana* as a regulator of stress response in guard cells (Desikan et al., 2008). In *M. truncatula*, NO was found to be a regulator of nodule senescence: increased and decreased levels of NO led to quickening and delay of nodule senescence, respectively (Cam et al., 2012). Since the nodules used in this study were all in relatively early stages of development, NO production in the nodules would not be expected to be high. Moreover, NO binds to and reduces activity of glutamine synthetase (Melo et al., 2011), the key enzyme in primary N assimilation. Thus, the down-regulation of *AHK5* would be important to maintain low concentrations of NO in the nodules. Alternatively, in *Arabidopsis*, *AHK5* is known to confer resistance to pathogens such as *Pseudomonas syringae* and *Botrytis cinerea* (Pham et al., 2012). Moreover, *AHK5* was most highly expressed

in the roots of *A. thaliana* (Desikan et al., 2008). An inverse relationship exists between the host plant immune response and symbiotic processes established in the root nodules (Toth and Stacey, 2015). Down-regulation of *AHK5* could be part of the mechanism that enables the harboring of bacterial cells within the plant cells.

dN/dS Analysis Disfavored the Single-Origin Hypothesis

Of the 3,894 MergedOrthoGroups, 31.5% (1,226) rejected the NULL scenario, which assumes a single rate of *dN/dS* throughout the tree. Of these 1,226 MergedOrthoGroups, 95.1% (1,166) supported a change in selection pressure at the base of the NFC. Among these, only 0.5% (6) rejected the TWOSTEP scenario while 68.3% (796) rejected the SINGLE scenario. The 58 core genes (MergedOrthoGroups), which should be strong candidates for playing key roles in the evolutionary origin of RNS, showed the same general pattern: 32.8% (19) rejected the NULL scenario, of which 89.5% (17) supported a change in selection pressure at the base of the NFC. Among these, 76.5% (13) rejected the SINGLE scenario, but none rejected the TWOSTEP scenario. We did not determine how many, if any, of the MergedOrthoGroups (i.e., sets of orthologous genes) are in fact the gene(s) that gave rise to RNS; thus our findings do not reject the single-origin hypothesis. However, the results of analyzing nearly 4,000 genes clearly disfavor the single-origin hypothesis.

Even the MergedOrthoGroups that rejected the TWOSTEP scenario are not in conflict with the two-step hypothesis, because the two-step hypothesis does not require the same gene to be responsible for both the gain-of-predisposition and the subsequent gain-of-function.

CONCLUSION

The evolution of RNS represents a major event in the biology of plant-microbe interactions (Doyle, 2016), and different explanations of the evolutionary origins have been proposed. We have demonstrated the genetic homology of RNS in the three lineages based on the presence of RNS pathway orthologs and the high similarity of gene expression patterns across the three species, thus demonstrating that RNS shares a common evolutionary event at the base of the NFC. At the same time, we show that most genes (regardless of whether the gene is involved in the process of RNS or not) that experience change in selection pressure at the base of the NFC also experienced subsequent changes in selection pressure at the base of each RNS host lineage. Taken together, our results are most consistent with the two-step hypothesis of the origin of RNS. The work of Werner et al. (2014) supported the two-step hypothesis, but was based on a single trait (capability to establish RNS) and had been criticized for being based on a flawed phylogenetic tree (Doyle, 2016; LPWG, 2017). Our findings provide additional support for the two-step hypothesis.

On the other hand, two recent papers suggest a more ancient origin of functional RNS within the NFC followed by multiple losses (Griesmann et al., 2018; van Velzen et al., 2018). In

Cannabaceae (Rosales), *Parasponia* retains the capability for RNS, whereas closely related *Trema* has lost it (van Velzen et al., 2018). Within a single genus, *Dryas octopetala* (Rosaceae) apparently does not form root nodules, whereas other *Dryas* species retain this trait (Becking, 1984). In a larger scale, two recent studies found that *NIN* and *RPG* have been lost among plants within the NFC that are not RNS hosts multiple times (Griesmann et al., 2018; van Velzen et al., 2018). Since the known functions of these genes are specific to RNS, multiple losses are difficult to explain under two-step hypothesis where these genes would be maintained for millions of years after the gain-of-predisposition event until the gain-of-function event.

Based on the known phylogenetic distribution of RNS hosts, the gain-of-predisposition at the MRCA of the NFC followed by a gain-of-function has been postulated as a parsimonious hypothesis since the discovery of the NFC (Soltis et al., 1995). What is the genetic nature of the predisposition assumed in the two-step hypothesis? Natural selection can only operate on a "predisposition" if the predisposition has a function of its own. Otherwise, the propensity for a gain-of-function could not have been conserved for tens of millions of years in multiple lineages (Doyle, 2011; Werner et al., 2014; Li et al., 2015). Likewise, a single-origin hypothesis needs an explanation for its apparently unparsimonious distribution of RNS hosts within the NFC. The high cost of RNS might be an explanation (Griesmann et al., 2018), but no direct evidence is available yet. In either case, the key to answer this question depends on an understanding of the genetic underpinnings that led to RNS, which still remains incomplete.

MATERIALS AND METHODS

Plant Growth Conditions

Ceanothus thyrsiflorus var. *thyrsiflorus* plants as rooted cuttings were purchased from Cornflower Farms (Elk Grove, CA, United States). Plants were grown in a greenhouse in the Plant Sciences Department, University of California, Davis, Davis, CA, United States, under natural daylight (14–15 h light/9–10 h dark), and irrigated daily with deionized water. Shortly after arrival, original media was removed and the plants were transplanted into Stuewe D40H pots (6.4 cm × 25.4 cm) filled with media consisting of perlite:sand:fir bark:peat moss, 3:1:1:1. The roots were inoculated at the time of transplant. No preexisting root nodules were observed in any plants. Soil collected from the rhizosphere of *Ceanothus velutinus* shrubs growing in Sagehen Experimental Forest (Truckee, CA, United States) was used as inoculum. The soil inoculum was directly applied to the exposed root ball.

Datisca glomerata seeds were collected from wild *D. glomerata* plants growing in Gates Canyon, Vacaville, CA, United States. Surface-sterilized seeds were germinated on moistened autoclaved vermiculite in a Magenta$^{\text{TM}}$ GA-7 Plant Culture Box, in a controlled environment (25°C, 16 h light/8 h dark). Once seedlings reached approximately 2 cm, they were transplanted into 5 cm × 5 cm × 7 cm pots filled with the same media used for the *C. thyrsiflorus* cuttings, and moved to the same

greenhouse as the *C. thyrsiflorus* cutting. Seedlings were irrigated daily with deionized water.

One-half-strength Hoagland's solution without nitrogen (Hoagland and Arnon, 1938) was applied twice a week for 9 weeks. The roots were then cut back and repotted with fresh media. The re-grown roots were collected 6 days later, and frozen immediately in liquid nitrogen (see section "*Tissue Sampling*").

Three days after the roots had been collected, seedlings were inoculated. Crushed nodules were collected from a different set of *D. glomerata* seedlings and used as inoculum. These previous seedlings had been inoculated with crushed nodules that originated from the same inoculum source as the *C. thyrsiflorus* cuttings described above.

Tissue Sampling

Roots tips and nodules of both *C. thyrsiflorus* and *D. glomerata* were collected as pairs from the same individual plants: for *C. thyrsiflorus*, pairs of root and nodule were collected from three individual plants; for *D. glomerata* pairs of root and nodule were collected from 6 individual plants. The *C. thyrsiflorus* nodules were collected 75 days or 96 days (11 or 14 weeks) after inoculation, and *D. glomerata* nodules were collected 24 days after inoculation. All plants were first placed under a halogen lamp at 250 μEm^{-2}s^{-1} for over 1 h to stabilize photosynthesis. For roots, root tips (approximately 2.5 cm) were collected, and for nodules, whole nodules (ranging from single- to multi-lobed) were collected (**Supplementary Image S4**). To remove media particles, roots or nodules were rinsed in deionized water immediately before flash freezing in liquid nitrogen. The sampling process was kept under 5 min per plant. Frozen tissues were stored at −80°C until RNA extraction.

RNA Extraction, Sequencing, and Transcriptome Assembly

Total RNA was extracted from each sample (see **Supplementary Table S9** for the concentrations and integrity scores for each RNA extracts). Barcode-indexed RNA-seq libraries were prepared in the DNA Technologies and Expression Analysis Cores at the University of California, Davis Genome Center (see **Supplementary Table S10** for the adapter and barcode sequences, see **Supplementary Image S5** for fragment sizes). The libraries were pooled in equimolar ratios for sequencing.

High-throughput sequencing was carried out in the DNA Technologies and Expression Analysis Cores at the UC Davis Genome Center. For *C. thyrsiflorus*, the pooled libraries were sequenced on a single lane of Illumina HiSeq 4000 (Illumina, San Diego, CA, United States) platform with paired-end 150 bp reads (PE150). For *D. glomerata* RNA-seq, the pooled libraries were sequenced on two lanes of Illumina HiSeq 2500 (Illumina) platform (PE150).

Raw sequence reads were trimmed based on read quality and adapter contamination using Scythe v0.991 (Buffalo, 2014) and Sickle v1.22 (Joshi and Fass, 2011) and the quality of cleaned reads were assessed using FastQC v0.11.2 (Andrews, 2010). Insert sizes were verified using Bowtie2 v2.2.6 (Langmead and Salzberg,

2012), Samtools v1.2 (Li, 2011), and Picard (Nazaire, 2017) (**Supplementary Images S5, S6**).

Cleaned high-quality paired-end reads were used to assemble a single root + nodule transcriptome using Trinity v2.20 (Grabherr et al., 2011) for *C. thyrsiflorus* and v2.06 for *D. glomerata*. The *C. thyrsiflorus* transcriptome was assembled on DIAG (Data Intensive Academic Grid) (White et al., 2010), and the *D. glomerata* transcriptome was assembled on the UC Davis Bioinformatics Core high-performance computing cluster. The newly assembled raw root + nodule transcriptomes were curated by passing through three independent screenings. In short, transcripts with low (<5) transcripts per million (TPM) in both tissues were removed in Screen-1 using RSEM v1.2.31 (Li and Dewey, 2011), transcripts with none or only short (<298 bp) CDSs were removed in Screen-2 using TransDecoder v3.0.1 (Haas et al., 2013) and ORFfinder (Wheeler et al., 2013), and transcripts with no similar sequence found in GenBank (Benson et al., 2005) non-redundant (nr) database were removed in Screen-3 using BLASTP v2.5.0+ (Altschul et al., 1990) (i.e., sequences with no hits better than 1e-20 were removed). Because neither the *C. thyrsiflorus* cuttings nor *D. glomerata* seedlings were clonal, alleles were collapsed using Allelepipe v1.0.28 (Dlugosch et al., 2013) (see **Supplementary Data Sheet S1** for full description).

Validation of RNA-seq With Real-Time Quantitative Reverse Transcription PCR

The assembled sequences and their respective fold changes determined via RNA-seq were validated using real-time quantitative reverse transcription PCR (RT-qRT-PCR). A total of nine genes were selected for RT-qRT-PCR validation in each plant (**Supplementary Table S7**) using orthologs in *D. glomerata* and *C. thyrsiflorus* determined by OrthoReD (**Supplementary Table S5**): five genes that are universally nodule-enhanced (*NIN, SYMREM, RPG, NRT1.8, CYP735A*), two genes that are universally root-enhanced (*MAX4, AHK5*) according to the RNA-seq results, and two additional reference genes (Ubiquitin ligase, Glyceraldehyde 3-P Dehydrogenase) used as controls. The controls were selected based on the reference genes tested in *Medicago sativa* (Castonguay et al., 2015).

Primer-BLAST (Ye et al., 2012) was used to design all PCR primers (**Supplementary Table S7**). Default settings were used except the size of the product was limited to 90–160 bps. PCR primers were chosen to amplify a segment near the middle of the transcript. All PCR primer pairs had a self-complementarity score and self-3′-complementarity score below 8.

Roots and root nodules were sampled from three different individuals of *C. thyrsiflorus* and three different individuals of *D. glomerata* with the same methods used to sample material for RNA-seq. Here, the nodules of *C. thyrsiflorus* and *D. glomerata* nodules were 84 days (12 weeks) and 48 days after inoculation, respectively, and the inoculation was about a month earlier in the season compared to the samples used for RNA-seq.

RNA extraction was performed as described in RNA-seq experiments. This was followed by DNAse treatment using Turbo DNA-free kit (Thermo Fisher, CA, San Diego, United States). Efficacy of DNAse treatment was confirmed by negative PCR amplification targeting plastidic *trnL* gene (data not shown). cDNA libraries were constructed using SuperScript™ III First-Strand Synthesis System Kit (Thermo Fisher, CA, United States) following manufacturer's protocol. RT qRT-PCR was performed using VeriQuest Fast SYBR Green qPCR Master Mix (Thermo Fisher, CA, United States) following manufacturer's protocol on a 7500 Fast Real-Time PCR system (Thermo Fisher, CA, United States).

Once the threshold cycle (C_T) was determined for each reaction, the average C_T for each gene was calculated for each tissue for each species. Then the C_T values for the seven target genes (*NIN, SYMREM, RPG, NRT1.8, CYP735A, MAX4, AHK5*) were normalized using the C_T of the reference. The nodule/root log2-fold change was calculated as the difference between the normalized C_T values of nodules and roots. This was compared to the nodule/root log2-fold fold change estimated based on RNA-seq.

M. truncatula Transcriptome and Genome

Published root and nodule (15 days after inoculation) transcriptome data were obtained from a previous study (Roux et al., 2014). The genome sequence of *M. truncatula*, i.e., the full length CDSs of 50,894 genes, was obtained from Phytozome v11.0 (Goodstein et al., 2012) (**Supplementary Table S3**). For genes that were not expressed in either of the tissues, the expression levels, the log2-fold changes, and the associated *p*-values were considered 0, 0, and 1, respectively. For genes that were tissue specific, the log2-fold changes were set as 15 (+15 for nodule-specific, −15 for root-specific genes), and the associated *p*-values were set as 1 whenever the values were not already provided. Henceforth, *M. truncatula* genome will refer to all the genes within the genome while *M. truncatula* transcriptome will refer to the subset (28,260 genes) that is expressed in the roots and/or in the nodules.

Transcriptome and Genome Annotation

The *C. thyrsiflorus* and *D. glomerata* transcriptomes and the *M. truncatula* genome were annotated using InterProScan v5.21 (Jones et al., 2014) (options: -goterms -pa) and Trinotate v3.0.1 (Haas et al., 2013) (options: -E 1e-5 –pfam_cutoff DNC) in parallel, to provide transcript annotations according to KEGG orthology (KO), Enzyme commission (EC) number, InterPro ID, Pfam ID, EggNOG ID, GO ID. The *M. truncatula* genome was previously annotated (Roux et al., 2014), but was reannotated to have consistency across all three species.

Transcriptome and Genome Completeness

Two analyses were conducted in parallel to assess and compare the completeness of the three root + nodule transcriptomes and *M. truncatula* genome. BUSCO v2.0 with Embryophyte reference (Simao et al., 2015) (options: -m proteins -sp arabidopsis) was used to estimate the completeness based on the

fraction of plant-universal single-copy orthologs found in each transcriptome or genome. Each transcriptome or genome was also passed through the KEGG-KAAS server (Moriya et al., 2007) as amino acids and used as a query in searches against all plant genomes via BLAST and annotated based on single-directional best hits (SBHs), to determine how many enzymes in terms of EC number are present within each biosynthetic pathway listed in KEGG.

Differential Gene Expression Analysis

Analyses of differential gene expression between the two tissues (nodule and root) for *C. thyrsiflorus* (n = 3) and *D. glomerata* (n = 6) were conducted using R v3.3.1 (R Core Team, 2013) with limma v3.3.3 Bioconductor package (Ritchie et al., 2015). Expected reads for each gene were estimated by using RSEM v1.2.31 (Li and Dewey, 2011) based on the mapping results of cleaned reads on to the root and the nodule transcriptomes and were used as the basis for the differential gene expression analysis. For *M. truncatula* (n = 3), the gene expression levels in the root and in the nodule, nodule/root log2-fold change, and associated adjusted *p*-values were obtained from a previous study (Roux et al., 2014). Genes with the log2-fold change over ± 1 and *p*-values (adjusted for repeated measures) of <0.005 for *C. thyrsiflorus* or <0.001 for *D. glomerata* and *M. truncatula* were considered to represent significantly differential expression. The *p*-value threshold for *C. thyrsiflorus* was set higher than the other two plants because limma did not generate any *p*-values < 0.001 due to less statistical power as a result of the fewer biological replicates for each tissue compared to *D. glomerata*.

Gene Ontology enrichment analysis (Gene Ontology Consortium, 2017) was conducted on both the root- and nodule-enhanced genes against the root + nodule transcriptome to detect GO terms and corresponding functions enriched in the root and the nodule. GO enrichment analysis was conducted using R v3.3.1 with topGO v2.28.0 Bioconductor package (Alexa and Rahnenfuhrer, 2016) based on gene counts of each GO term using runTest() (options: algorithm = "classic", statistic = "fisher"). After calculating the raw *p*-values, to account for multiple testing, the significance of each GO term was tested through Benjamini–Hochberg procedure. The *p*-value thresholds was set at α = 0.01, and the false discovery rate was set at Q = 0.20.

Database for Orthology Comparison

To enable the comparison of orthologous sets of genes across and beyond the NFC, a nucleotide sequence database was constructed with 12 genomes (total CDSs) from species within the NFC and close outgroups in combination with the *C. thyrsiflorus* and *D. glomerata* transcriptomes and the *M. truncatula* genome (see **Supplementary Table S11** for the complete list). These sequences were collected from three different databases: Phytozome v11 (Goodstein et al., 2012), Genome Explorer (Carleton College, 2010), and miyakogusa.jp 3.0 (Sato et al., 2008). A small fraction (<0.1%) of the CDSs originally collected could not be translated reliably due to ambiguous reading frames or premature stop codons. These sequences were removed from the analyses.

Orthology Predictions

Orthologs for each gene in the *C. thyrsiflorus* and *D. glomerata* transcriptomes and the *M. truncatula* genome were predicted independently using OrthoReD (Battenberg et al., 2017) (options: –blast_type NCBI –loci_threshold 10 –sander_schneider YES –overlap_threshold 0) with *Vitis vinifera* (Vitaceae, Vitales) set as the outgroup. In order to capture orthologs beyond the in-paralogs (Sonnhammer and Koonin, 2002) expected from the genome duplications in Fabaceae (Cannon et al., 2010) and Rosaceae (Potter et al., 2007), OrthoReD was modified to ignore gene duplication(s) within a single order (Customized version of OrthoReD[1]). For genes with multiple isoforms or alleles, orthologs were predicted independently for each member, and the union of the predictions of all isoforms was considered the set of predicted orthologs for the gene. The predicted orthologs were then merged into 27,367 non-overlapping MergedOrthoGroups using a custom Perl script (**Supplementary Table S5**).

In addition, orthologs of 19 genes from *M. truncatula* or *L. japonicus* whose functions in the establishment of RNS have been well characterized (Kouchi et al., 2010; Oldroyd, 2013) were also predicted (see **Table 1** for the gene list). Reference sequences for these 19 RNS pathway genes were collected from GenBank, and verified in the database assembled in this study. OrthoReD with the same parameter settings as other sequences were used to predict orthology for *L. japonicus* genes.

Presence of RNS Pathway Genes

For each of the 19 RNS pathway genes, a set of genes was collected comprising every *C. thyrsiflorus* or *D. glomerata* gene that was (1) predicted as a putative ortholog starting with the reference gene as a query, and/or (2) predicted the reference gene as its putative ortholog. The presence/absence of orthologs for each gene identified as a part of the RNS pathway established in *M. truncatula* and/or *L. japonicus* was scored for each newly assembled transcriptome and compared to previous studies published on *Alnus glutinosa* and *Casuarina glauca* (Hocher et al., 2011), and *D. glomerata* (Demina et al., 2013).

Expression Similarity Analysis

From the 27,367 MergedOrthoGroups, we first parsed out whole MergedOrthoGroups, each of which: (1) contained at least one member from each of *C. thyrsiflorus*, *D. glomerata*, and *M. truncatula*; and (2) was identical to a set of orthologs predicted from a single gene. We also parsed subsets within the remaining MergedOrthoGroups that met the aforementioned two criteria only with respect to the three species *C. thyrsiflorus*, *D. glomerata*, and *M. truncatula*.

Then the gene expression pattern was scored as root-enhanced, nodule-enhanced, or not significantly different between the two tissues based on the differential gene expression analysis for each species for each MergedOrthoGroup. For MergedOrthoGroups where a single species was represented by more than one paralog, the gene expression pattern for this species was considered root- or nodule-enhanced if there was

---

[1]https://github.com/kbattenb/OrthoReD_vNFC

at least one significantly differentially expressed paralog, unless one or more of them were enhanced in one tissue and other paralog(s) were enhanced in the other tissue, in which case the MergedOrthoGroup was considered to be both root- and nodule-enhanced for that species. Among the representative MergedOrthoGroups, those that were either nodule-enhanced or root-enhanced in all three species were collected and designated as the core MergedOrthoGroups.

In order to determine the functions of RNS establishment that are conserved across the three species, root- and nodule-enhanced genes within the core MergedOrthoGroup were collected separately for each species, and GO enrichment analysis was conducted on these genes against their respective root + nodule transcriptome using the same conditions as described above.

Next, we tested how strongly the differences in gene expression between roots and nodules in each species are correlated with those of other species using the approach previously applied for measuring the similarities between biological replicates (Kempema et al., 2007). For each pair of species, the gene expression level fold changes between the two tissues were plotted for all MergedOrthoGroups. The Pearson correlation coefficient was then calculated separately for all MergedOrthoGroups, for MergedOrthoGroups that were nodule-enhanced in both species, and for those that were root-enhanced in both species using cor.test() available in R v3.3.1, stats package v3.3.1. For MergedOrthoGroups where at least one of the species was represented by more than one paralog, the most similar pair of fold changes was used as the representative.

Finally, the degree and significance of overall similarity in gene expression between each pair of species was assessed using dissonance scores (summarized in **Supplementary Image S2**). For each pair of species for each MergedOrthoGroup, the degree of dissonance between the species was calculated as the average pairwise difference in fold change between all transcripts for the two species belonging to that group (for some groups, one or more species was represented by more than one transcript). For example, the dissonance score (D) between two species A and B for a given MergedOrthoGroup each with m and n representative transcripts, was calculated as $D = \left[\sum_{i=1}^{m} \left(\sum_{j=1}^{n} |A_i - B_j| \right) \right] / (m * n)$. Next, the overall dissonance score between species A and B was calculated as the total sum of the dissonance scores of all representative MergedOrthoGroups. Finally, in order to determine the statistical significance of this measure, a permutation test (random resampling of fold change values with no replacement) was conducted to calculate the overall dissonance score for 10,000 replicates, and the p-value was calculated as the fraction of permutations that yielded an overall dissonance score equal to or smaller than the observed data.

The normal distributions of the total dissonance scores among the permutations were verified using R v3.3.1 with stats package v3.3.1 based on quantile-quantile plots and histograms generated using qqnorm() and hist() (**Supplementary Image S7**).

dN/dS Analysis

The aforementioned 3,894 MergedOrthoGroups, i.e., sets of orthologs that (1) contained at least one member from each of *C. thyrsiflorus*, *D. glomerata*, and *M. truncatula*, and (2) were identical to a set of orthologs predicted from a single gene, were used in this analysis.

The three competing hypotheses (single-origin, multiple-origin, and two-step hypothesis) differ in the expected timing(s) of gaining predisposition or function of RNS, the events that would have caused changes in selection pressures on the genes involved. A change in selection pressure can be detected as a change in the average ratio of non-synonymous to synonymous mutations (*dN/dS*) (Hurst, 2002).

For each MergedOrthoGroup, the likelihoods of up to three different hypothesis-based scenarios (SINGLE, MULTI, and TWOSTEP), each of which predicts a different timing for a change in *dN/dS*, were compared to a NULL scenario, which assumes a constant *dN/dS* throughout the gene phylogeny (**Figure 4**). The SINGLE scenario, based on the hypothesis of a single origin of RNS, assumes a different *dN/dS* along the branch leading to the base of the NFC (branch *a* in **Figure 4**) compared to the rest of the tree. The MULTI scenario, based on the hypothesis of multiple independent origins of RNS, assumes different values of *dN/dS* along the branches leading to *C. thyrsiflorus*, *D. glomerata*, and the legumes, (branches *b*, *c*, and *d*, respectively, in **Figure 4**) compared to the rest of the tree. The TWOSTEP scenario, based on the hypothesis of a gain of predisposition to RNS at the base of the NFC followed by independent gains of function in each lineage exhibiting RNS, assumes different values of *dN/dS* for all of the aforementioned branches (*a* through *d* in **Figure 4**) compared to the rest of the tree.

To serve as the basis for comparisons between these scenarios, a multiple sequence alignment (MSA) and a most-likely gene phylogeny were generated for each MergedOrthoGroup. Each MSA was generated using MAFFT v7.272 (Katoh and Standley, 2013) (options: −amino −localpair −retree 2 −maxiterate 1000). Two phylogenetic trees were constructed using RAxML v8.2.8 (Stamatakis, 2006) (options: -f d -m GTRGAMMAIX). The topology of one tree was constrained to the family level according to the established phylogeny (APG, 2009) (see **Figure 4** as a reference), while the topology of the second tree was not constrained. For each tree, four parallel searches were carried out to prevent lodging on to local optima. To select one of the two trees for further analysis, the two trees were compared using a K-H test (Kishino and Hasegawa, 1989) through PAML v4.8 (Yang, 2007) (options: codeml, clock = 0, aaDist = 0, model = 0, NSsites = 0, fix_blength = −1). Unless the unconstrained tree was significantly better than the constrained tree ($p < 0.05$) the constrained tree was used for further analyses.

For each MergedOrthoGroup, the MSA was fitted onto the most-likely tree under the four scenarios to calculate the likelihood scores for each using PAML v4.8 (options: codeml, clock = 0, aaDist = 0, model = 2, NSsites = 0, fix_blength = −1). An S-H test (Shimodaira and Hasegawa, 1999) was carried out simultaneously in PAML. Scenarios significantly worse than the best ($p < 0.05$) were rejected, and MergedOrthoGroups that

did not reject the NULL scenario were assumed to conform to the NULL scenario. In some MergedOrthoGroups for which the unconstrained topology was used, some scenarios were not tested because the clade required for the scenario was missing from the tree.

Hypothesis Testing

Based on the results generated by the methods described above, the three competing hypotheses (single-origin, multiple-origin, and two-step) were tested against each other. The multiple-origin hypothesis differs from the other two hypotheses in that it assumes completely independent gains of RNS in different lineages. To test this hypothesis, we used the presence/absence of RNS pathway genes in the two actinorhizal plants, and the gene expression patterns from the three different lineages for each orthologous set of genes, to see if the three lineages were significantly more similar to one another than expected by random chance (see section "*Expression Similarity Analysis*" under Material and Methods). A significant result would indicate the presence of a common evolutionary background among the three species compared (which are phylogenetically distantly related species distributed throughout the NFC), supporting single-origin or two-step hypothesis, and rejecting the multiple-origin hypothesis.

The dN/dS analysis directly compares the three hypotheses. Among MergedOrthoGroups that rejected the NULL scenario, MergedOrthoGroups that fail to reject only the MULTI scenario would be consistent with both the multiple-origin hypothesis and the two-step hypothesis, because the two-step hypothesis does not require the same gene to be the cause of both the gain-of-predisposition and the gain-of-function. However, such genes would be in conflict with the single-origin hypothesis because additional changes for each RNS host lineage are not assumed in this hypothesis. Likewise, genes that fail to reject only the SINGLE scenario would be consistent with either the single-origin hypothesis or the two-step hypothesis, but would be in conflict with the multiple-origin hypothesis. Finally, genes that fail to reject only the TWOSTEP scenario would be in conflict with both the multiple-origin and single-origin hypotheses, and would only be consistent with the two-step hypothesis.

AUTHOR CONTRIBUTIONS

KB conducted the experiments, analyzed the data, and substantially wrote the manuscript. CT conducted the RT-qPCR reactions together with KB. JC, DP, and AB provided suggestions on the experimental design, interpretations for comparative transcriptomics, phylogenetics, and the biology of RNS, and contributed to editing the manuscript. AB provided an initial project framework.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

REFERENCES

Alexa, A., and Rahnenfuhrer, J. (2016). *Topgo: Enrichment Analysis for Gene Ontology. R Package Version* 2.26.0. Available at: https://bioconductor.org/packages/release/bioc/html/topGO.html

Altenhoff, A. M., Studer, R. A., Robinson-Rechavi, M., and Dessimoz, C. (2012). Resolving the ortholog conjecture: orthologs tend to be weakly, but significantly, more similar in function than paralogs. *PLoS Comput. Biol.* 8:e1002514. doi: 10.1371/journal.pcbi.1002514

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data.* Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

APG (2009). An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* 161, 105–121. doi: 10.1016/j.jep.2015.05.035

Arrighi, J., Godfroy, O., Billy, F., Saurat, O., Jauneau, A., and Gough, C. (2008). The RPG gene of *Medicago truncatula* controls Rhizobium-directed polar growth during infection. *Proc. Natl. Acad. Sci. U.S.A.* 105, 9817–9822. doi: 10.1073/pnas.0710273105

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556

Azarakhsh, M., Kirienko, A. N., Zhukov, V. A., Lebedeva, M. A., Dolgikh, E. A., and Lutova, L. A. (2015). Knotted1-like homeobox 3: a new regulator of symbiotic nodule development. *J. Exp. Bot.* 66, 7181–7195. doi: 10.1093/jxb/erv414

Battenberg, K., Lee, E. K., Chiu, J. C., Berry, A. M., and Potter, D. (2017). OrthoReD: a rapid and accurate orthology prediction tool with low computational requirement. *BMC Bioinformatics* 18:310. doi: 10.1186/s12859-017-1726-5

Becking, J. H. (1984). "Identification of the endophyte of Dryas and *Rubus* (Rosaceae)," in *Frankia Symbioses*, eds A. D. L. Akkermans, D. Baker, K. Huss-Danell, and J. D. Tjepkema (Dordrecht: Springer), 105–128. doi: 10.1007/978-94-009-6158-6_11

Bell, C. D., Soltis, D. E., and Soltis, P. S. (2010). The age and diversification of the angiosperm re-revisited. *Am. J. Bot.* 97, 1296–1303. doi: 10.3732/ajb.0900346

Bennett, T., Liang, Y., Seale, M., Ward, S., Muller, D., and Leyser, O. (2016). Strigolactone regulates shoot development through a core signalling pathway. *Biol. Open* 5, 1806–1820. doi: 10.1242/bio.021402

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Wheeler, D. L. (2005). GenBank. *Nucleic Acids Res.* 33, D34–D38.

Besserer, A., Puech-Pages, V., Kiefer, P., Gomez-Roldan, V., Jauneau, A., Roy, S., et al. (2006). Strigolactones stimulate arbuscular mycorrhizal fungi by activating mitochondria. *PLoS Biol.* 4:e226. doi: 10.1371/journal.pbio.0040226

Buffalo, V. (2014). *Scythe: A Bayesian Adapter Trimmer.* Available at: https://github.com/vsbuffalo/scythe

Cam, Y., Pierre, O., Boncompagni, E., Herouart, D., Meilhoc, E., and Bruand, C. (2012). Nitric oxide (NO): a key player in the senescence of *Medicago truncatula* root nodules. *New Phytol.* 196, 548–560. doi: 10.1111/j.1469-8137.2012.04282.x

Cannon, S. B., Ilut, D., Farmer, A. D., Maki, S. L., May, G. D., Singer, S. R., et al. (2010). Polyploidy did not predate the evolution of nodulation in all legumes. *PLoS One* 5:e11630. doi: 10.1371/journal.pone.0011630

Carleton College (2010). *The Science Education Resource Center. Genome Explorer.* Available at: https://serc.carleton.edu/exploring_genomics/index.html [accessed Dec 13, 2017].

Castonguay, Y., Michaud, J., and Dube, M. (2015). Reference genes for RT-qPCR analysis of environmentally and developmentally regulated gene expression in Alfalfa. *Am. J. Plant Sci.* 6, 132–143. doi: 10.4236/ajps.2015.61015

Chapin, F. S., Walker, L. R., Fastie, C. L., and Sharman, L. C. (1994). Mechanisms of primary succession following deglaciation at Glacier Bay. Alaska. *Ecol. Monogr.* 64, 149–175. doi: 10.1007/BF00396734

Chiba, Y., Shimizu, T., Miyakawa, S., Kanno, Y., Koshiba, T., Kamiya, Y., et al. (2015). Identification of *Arabidopsis thaliana* NRT1/PTR Family (NPF) proteins capable of transporting plant hormones. *J. Plant Res.* 128, 679–686. doi: 10.1007/s10265-015-0710-2

Clavijo, F., Diedhiou, I., Vaissayre, V., Brottier, L., Acolatse, J., Moukouanga, D., et al. (2015). The *Casuarina* NIN gene is transcriptionally activated throughout *Frankia* root infection as well as in response to bacterial diffusible signals. *New Phytol.* 208, 887–903. doi: 10.1111/nph.13506

Core Team, R. (2013). *R: A Language and Environment for Statistical Computing.* Veinna: R Foundation for Statistical Computing.

Cronquist, A. (1988). *The Evolution and Classification of Flowering Plants,* 2nd Edn. New York Botanical Garden, NY: The University of California.

Demina, I. V., Persson, T., Santos, P., Plaszczyca, M., and Pawlowski, K. (2013). Comparison of the nodule vs. root transcriptome of the actinorhizal plant *Datisca glomerata*: actinorhizal nodules contain a specific class of defensins. *PLoS One* 8:e72442. doi: 10.1371/journal.pone.0072442

Desikan, R., Horak, J., Chaban, C., Mira-Rodado, V., Witthoft, J., Elgass, K., et al. (2008). The histidine kinase AHK5 integrates endogenous and environmental signals in *Arabidopsis* guard cells. *PLoS One* 3:e2491. doi: 10.1371/journal.pone.0002491

Dlugosch, K. M., Lai, Z., Bonin, A., Hierro, J., and Rieseberg, L. H. (2013). Allele identification for transcriptome-based population genomics in the invasive plant *Centaurea solstitialis*. *G3*, 359–367. doi: 10.1534/g3.112.003871

Doyle, J. J. (2011). Phylogenetic perspectives on the origins of nodulation. *Mol. Plant Microbe Interact.* 24, 1289–1295. doi: 10.1094/MPMI-05-11-0114

Doyle, J. J. (2016). Chasing unicorns: nodulation origins and the paradox of novelty. *Am. J. Bot.* 103, 1865–1868. doi: 10.3732/ajb.1600260

Foo, E., and Davies, N. W. (2011). Strigolactones promote nodulation in pea. *Planta* 234, 1073–1081. doi: 10.1007/s00425-011-1516-7

Fowler, D., Coyle, M., Skiba, U., Sutton, M. A., Cape, J. N., Reis, S., et al. (2013). The global nitrogen cycle in the twenty-first century. *Phil. Trans. R. Soc. B* 368:20130164. doi: 10.1098/rstb.2013.0164

Gamas, P., Brault, M., Jardinaud, M., and Frugier, F. (2017). Cytokinins in symbiotic nodulation: when, where, what for? *Trends Plant Sci.* 22, 792–802. doi: 10.1016/j.tplants.2017.06.012

Gene Ontology Consortium (2017). Expansion of the gene ontology knowledgebase and resources. *Nucleic Acids Res.* 45, D331–D338. doi: 10.1093/nar/gkw1108

Gomez-Roldan, V., Fermas, S., Brewer, F. B., Puech-Pages, V., Dun, E. A., Pillot, J., et al. (2008). Strigolactone inhibition of shoot branching. *Nature* 455, 189–195. doi: 10.1038/nature07271

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, D1178–D1186. doi: 10.1093/nar/gkr944

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883

Granqvist, E., Sun, J., Op, den Camp, R., Pujic, P., Hill, L., et al. (2015). Bacterial-induced calcium oscillations are common to nitrogen-fixing associations of nodulating legumes and non-legumes. *New Phytol.* 207, 551–558. doi: 10.1111/nph.13464

Griesmann, M., Chang, Y., Liu, X., Song, Y., Haberer, G., Crook, M. B., et al. (2018). Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. *Science* 361:eaat1743. doi: 10.1126/science.aat1743

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/nprot.2013.084

Herridge, D. F., Peoples, M. B., and Boddey, R. M. (2008). Global inputs of biological nitrogen fixation in agricultural systems. *Plant Soil* 311, 1–18. doi: 10.1007/s11104-008-9668-3

Hoagland, D. R., and Arnon, D. I. (1938). *The Water-Culture for Growing Plants Without Soil.* Berkeley, CA: University of California Collage of Agricultural Experiment Station.

Hocher, V., Alloisio, N., Auguy, F., Fournier, P., Doumas, P., Pujic, P., et al. (2011). Transcriptomics of actinorhizal symbioses reveals homologs of the whole common symbiotic signaling cascade. *Plant Physiol.* 156, 700–711. doi: 10.1104/pp.111.174151

Hurst, L. D. (2002). The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet.* 18, 486–487. doi: 10.1016/S0168-9525(02)02722-1

Iwama, A., Yamashino, T., Tanaka, Y., Sakakibara, H., Kakimoto, T., Sato, S., et al. (2007). AHK5 histidine kinase regulates root elongation through an ETR1-dependent abscisic acid and ethylene signaling pathway in *Arabidopsis thaliana*. *Plant Cell Physiol.* 48, 375–380. doi: 10.1093/pcp/pcl065

Jones, P., Binns, D., Chang, H., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031

Joshi, N. A., and Fass, J. N. (2011). *Sickle: A Sliding-Window, Adaptive, Quality-Based Trimming Tool for FastQ Files.* Available at: https://github.com/najoshi/sickle

Kapulnik, Y., and Koltai, H. (2014). Strigolactone involvement in root development, response to abiotic stress, and interactions with the biotic soil environment. *Plant Physiol.* 166, 560–569. doi: 10.1104/pp.114.244939

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kempema, L. A., Cui, X., Holzer, F. M., and Walling, L. L. (2007). *Arabidopsis* transcriptome changes in response to phloem-feeding silver leaf whitefly nymphs. Similarities and distinctions in responses to aphids. *Plant Physiol.* 143, 849–865. doi: 10.1104/pp.106.090662

Kishino, H., and Hasegawa, M. (1989). Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J. Mol. Evol.* 29, 170–179. doi: 10.1007/BF02100115

Koonin, E. V. (2005). Orthologs, paralogs, and evolutionary genomics. *Annu. Rev. Genet.* 39, 309–338. doi: 10.1146/annurev.genet.39.073003.114725

Kouchi, H., Imaizumi-Anraku, H., Hayashi, M., Hakoyama, T., Nakagawa, T., Umehara, Y., et al. (2010). How many peas in a pod? Legume genes responsible for mutualistic symbioses underground. *Plant Cell Physiol.* 51, 1381–1397. doi: 10.1093/pcp/pcq107

Langmead, B., and Salzberg, S. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Larrainzar, E., Riely, B. K., Kim, S. C., Carrasquilla-Garcia, N., Yu, H., Hwang, H., 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. doi: 10.1104/pp.15.00350

Lefebvre, B., Timmers, T., Mbengue, T., Moreau, S., Herve, C., Toth, K., et al. (2010). A remorin protein interacts with symbiotic receptors and regulates bacterial infection. *Proc. Natl. Acad. Sci. U.S.A.* 107, 2343–2348. doi: 10.1073/pnas.0913320107

Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference. *BMC Bioinformatics* 12:323. doi: 10.1186/1471-2105-12-323

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993. doi: 10.1093/bioinformatics/btr509

Li, H., Wang, W., Mortimer, P. E., Li, R., Li, D., Hyde, K. D., et al. (2015). Large-scale phylogenetic analyses reveal multiple gains of actinorhizal nitrogen-fixing symbioses in angiosperms associated with climate change. *Sci. Rep.* 5:14023. doi: 10.1038/srep14023

Li, J., Fu, Y., Pike, S. M., Bao, J., Tian, W., Zhang, Y., et al. (2010). The *Arabidopsis* nitrate transporter NRT1.8 functions in nitrate removal from the xylem sap and mediates cadmium tolerance. *Plant Cell* 22, 1633–1646. doi: 10.1105/tpc.110.075242

LPWG (2017). A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny. *Taxon* 66, 44–77. doi: 10.12705/661.3

Markmann, K., Giczey, G., and Parniske, M. (2008). Functional adaptation of a plant receptor-kinase paved the way for the evolution of intracellular root symbioses with bacteria. *PLoS Biol.* 6:e68. doi: 10.1371/journal.pbio.0060068

McAdam, E. L., Hugill, C., Fort, S., Samain, E., Cottaz, S., Davies, N. W., et al. (2017). Determining the site of action of strigolactones during nodulation. *Plant Physiol.* 175, 529–542. doi: 10.1104/pp.17.00741

Melo, P. M., Silva, L. S., Ribeiro, I., Seabra, A. R., and Carvalho, H. G. (2011). Glutamine synthetase is a molecular target of nitric oxide in root nodules of *Medicago truncatula* and is regulated by tyrosine nitration. *Plant Physiol.* 157, 1505–1517. doi: 10.1104/pp.111.186056

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A., and Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35, W182–W185. doi: 10.1093/nar/gkm321

Nazaire, M. D. (2017). *Picard Team (SAMTools). Picard (Broad Institute).* Available at: http://broadinstitute.github.io/picard/

Nguyen, T. V., Wibberg, D., Battenberg, K., Blom, J., Vanden Heuvel, B., Berry, A. M., et al. (2016). An assemblage of *Frankia* Cluster II strains from California contains the canonical nod genes and also the sulfotransferase gene nodH. *BMC Genomics* 17:796. doi: 10.1186/s12864-016-3140-1

Okazaki, S., Tittabutr, P., Teulet, A., Thouin, J., Fardoux, J., Chaintreuil, C., et al. (2015). Rhizobium-legume symbiosis in the absence of Nod factors: two possible scenarios with or without the T3SS. *ISME* 10, 64–74. doi: 10.1038/ismej.2015.103

Oldroyd, G. E. D. (2013). Speak, friend, and enter: signaling systems that promote beneficial symbiotic associations in plants. *Nat. Rev. Microbiol.* 11, 252–263. doi: 10.1038/nrmicro2990

Pawlowski, K., and Demchenko, K. N. (2012). The diversity of actinorhizal symbiosis. *Protoplasma* 249, 967–979. doi: 10.1007/s00709-012-0388-4

Persson, T., Battenberg, K., Demina, I. V., Vigil-Stenman, T., Vanden Heuvel, B., Pujic, P., et al. (2015). Candidatus *Frankia Datisca* Dg1, the actinobacterial microsymbiont of *Datisca glomerata*, expresses the canonical nod genes nodABC in symbiosis with its host plant. *PLoS One* 10:e0127630. doi: 10.1371/journal.pone.0127630

Pham, J., Liu, J., Bennett, M. H., Mansfield, J. W., and Desikan, R. (2012). *Arabidopsis* histidine kinase 5 regulates salt sensitivity and resistance against bacterial and fungal infection. *New Phytol.* 194, 168–180. doi: 10.1111/j.1469-8137.2011.04033.x

Potter, D., Eriksson, T., Evans, R. C., Oh, S., Smedmark, J. E. E., Morgan, D. R., et al. (2007). Phylogeny and classification of Rosaceae. *Plant Syst. Evol.* 266, 5–43. doi: 10.1007/s00606-007-0539-9

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43:e47. doi: 10.1093/nar/gkv007

Roux, B., Rodde, N., Jardinaud, M., Timmers, T., Sauviac, L., Cottret, L., et al. (2014). An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. *Plant J.* 77, 817–837. doi: 10.1111/tpj.12442

Ruyter-Spira, C., Al-Babili, S., Krol, S., and Bouwmeester, H. (2013). The biology of strigolactones. *Trends Plant Sci.* 18, 72–83. doi: 10.1016/j.tplants.2012.10.003

Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., Kato, T., Nakao, M., et al. (2008). Genome structure of the legume. *Lotus japonicus. DNA Res.* 15, 227–239. doi: 10.1093/dnares/dsn008

Schauser, L., Roussis, A., Stiller, J., and Stougaard, J. (1999). A plant regulator controlling development of symbiotic root nodules. *Nature* 402, 191–195. doi: 10.1038/46058

Shimodaira, H., and Hasegawa, M. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* 16, 1114–1116. doi: 10.1093/oxfordjournals.molbev.a026201

Simao, F. S., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351

Soltis, D. E., Soltis, P. S., Morgan, D. R., Swensen, S. M., Mullin, B. C., Dowdi, J. M., et al. (1995). Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. *Proc. Natl. Acad. Sci. U.S.A.* 92, 2647–2651. doi: 10.1073/pnas.92.7.2647

Sonnhammer, E. L. L., and Koonin, E. V. (2002). Orthology, paralogy and proposed classification for paralog subtypes. *Trends Genet.* 18, 619–620. doi: 10.1016/S0168-9525(02)02793-2

Sorefan, K., Booker, J., Haurogne, K., Goussot, M., Bainbridge, K., Foo, E., et al. (2003). MAX4 and RMS1 are orthologous dioxygenase-like genes that regulate shoot branching in *Arabidopsis* and pea. *Genes Dev.* 17, 1469–1474. doi: 10.1101/gad.256603

Soyano, T., Shimoda, Y., and Hayashi, M. (2014). Nodule inception antagonistically regulates gene expression with nitrate in *Lotus japonicus. Plant Cell Physiol.* 56, 368–376. doi: 10.1093/pcp/pcu168

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690. doi: 10.1093/bioinformatics/btl446

Svistoonoff, S., Benabdoun, F. M., Nambiar-Veetil, M., Imanishi, L., Vaissayre, V., Cesari, S., et al. (2013). The independent acquisition of plant root nitrogen-fixing symbiosis in fabids recruited the same genetic pathway for nodule organogenesis. *PLoS One* 8:e64515. doi: 10.1371/journal.pone.0064515

Swensen, S. M. (1996). The evolution of actinorhizal symbioses: evidence for multiple origins of the symbiotic association. *Am. J. Bot.* 83, 1503–1512. doi: 10.1002/j.1537-2197.1996.tb13943.x

Takei, K., Yamaya, T., and Sakakibara, H. (2004). *Arabidopsis* CYP735A1 and CYP735A2 encode cytokinin hydroxylases that catalyze the biosynthesis of trans-zeatin. *J. Biol. Chem.* 279, 41866–41872. doi: 10.1074/jbc.M406337200

Thomas, K. A., and Berry, A. M. (1989). Effects of continuous nitrogen application and nitrogen preconditioning on nodulation and growth of *Ceanothus griseus* var horizontalis. *Plant Soil* 118, 181–187. doi: 10.1007/BF02232805

Tirichine, L., Sandal, N., Madsen, L. H., Radutoiu, S., Albrektsen, A. S., Sato, S., et al. (2007). A gain-of-function mutation in a cytokinin receptor triggers spontaneous root nodule organogenesis. *Science* 315, 104–107. doi: 10.1126/science.1132397

Toth, K., and Stacey, G. (2015). Does plant immunity play a critical role during initiation of the legume-Rhizobiumsymbiosis? *Front. Plant Sci.* 6:401. doi: 10.3389/fpls.2015.00401

van Velzen, R., Holmer, R., Bu, F., Rutten, L., van Zeijl, A., Liu, W., et al. (2018). Comparative genomics of the nonlegume Parasponia reveals insights into evolution of nitrogen-fixing rhizobium symbioses. *Proc. Natl. Acad. Sci. U.S.A.* 15, E4700–E4709. doi: 10.1073/pnas.1721395115

van Zeijl, A., Liu, W., Xiao, T. T., Kohlen, W., Yang, W., Bisseling, T., et al. (2015a). The strigolactone biosynthesis gene DWARF27 is co-opted in rhizobium symbiosis. *BMC Plant Biol.* 15:260. doi: 10.1186/s12870-015-0651-x

van Zeijl, A., Op, den Camp, R. H. M., Deinum, E. E., Charnikhova, T., Franssen, H., et al. (2015b). Rhizobium lipo-chitooligosaccharide signaling triggers accumulation of cytokinins in *Medicago truncatula* roots. *Mol. Plant* 8, 1213–1226. doi: 10.1016/j.molp.2015.03.010

Vegh, A., Incze, N., Fabian, A., Huo, H., Bradford, K. J., Balazs, E., et al. (2017). Comprehensive analysis of DWARF14-LIKE2 (DLK2) reveals its functional divergence from strigolactone-related paralogs. *Front. Plant Sci.* 8:1641. doi: 10.3389/fpls.2017.01641

Waters, M. T., Nelson, D. C., Scaffidi, A., Flematti, G. R., Sun, Y. K., Dixon, K. W., et al. (2012). Specialisation within the DWARF14 protein family confers distinct responses to karrikins and strigolactones in *Arabidopsis*. *Development* 139, 1285–1295. doi: 10.1242/dev.074567

Werner, G. D. A., Cornwell, W. K., Sprent, J. I., Kattge, J., and Kiers, T. E. (2014). A single evolutionary innovation drives the deep evolution of symbiotic N2-fixation in angiosperms. *Nat. Commun.* 5:4087. doi: 10.1038/ncomms5087

Wheeler, D. L., Church, D. M., Federhen, S., Lash, A. E., Madden, T. L., Pontius, J. U., et al. (2013). Database resources of the national center for biotechnology. *Nucleic Acids Res.* 31, 28–33. doi: 10.1093/nar/gkg033

White, O., Wommack, K., Ravel, J., Mahurkar, A., and Angiuoli, S. (2010). *Acquisition of Data Intensive Academic Grid (DIAG). NSF Major Research Instrumentation Program Award Number 0959894.* Available at: https://www.nsf.gov/awardsearch/showAward?AWD_ID=0959894

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* 13:134. doi: 10.1186/1471-2105-13-134

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Distinctive Patterns of Flavonoid Biosynthesis in Roots and Nodules of *Datisca glomerata* and *Medicago spp.* Revealed by Metabolomic and Gene Expression Profiles

Isaac Gifford[1]*, Kai Battenberg[1†], Arpana Vaniya[2†], Alex Wilson[1], Li Tian[1], Oliver Fiehn[2] and Alison M. Berry[1]

[1] Department of Plant Sciences, University of California, Davis, Davis, CA, United States, [2] West Coast Metabolomics Center, University of California, Davis, Davis, CA, United States

Plants within the Nitrogen-fixing Clade (NFC) of Angiosperms form root nodule symbioses with nitrogen-fixing bacteria. Actinorhizal plants (in Cucurbitales, Fagales, Rosales) form symbioses with the actinobacteria *Frankia* while legumes (Fabales) form symbioses with proteobacterial rhizobia. Flavonoids, secondary metabolites of the phenylpropanoid pathway, have been shown to play major roles in legume root nodule symbioses: as signal molecules that in turn trigger rhizobial nodulation initiation signals and acting as polar auxin transport inhibitors, enabling a key step in nodule organogenesis. To explore a potentially broader role for flavonoids in root nodule symbioses across the NFC, we combined metabolomic and transcriptomic analyses of roots and nodules of the actinorhizal host *Datisca glomerata* and legumes of the genus *Medicago.* Patterns of biosynthetic pathways were inferred from flavonoid metabolite profiles and phenylpropanoid gene expression patterns in the two hosts to identify similarities and differences. Similar classes of flavonoids were represented in both hosts, and an increase in flavonoids generally in the nodules was observed, with differences in flavonoids prominent in each host. While both hosts produced derivatives of naringenin, the metabolite profile in *D. glomerata* indicated an emphasis on the pinocembrin biosynthetic pathway, and an abundance of flavonols with potential roles in symbiosis. Additionally, the gene expression profile indicated a decrease in expression in the lignin/monolignol pathway. In *Medicago sativa*, by contrast, isoflavonoids were highly abundant featuring more diverse and derived isoflavonoids than *D. glomerata.* Gene expression patterns supported these differences in metabolic pathways, especially evident in a difference in expression of cinnamic acid 4-hydroxylase (C4H), which was expressed at substantially lower levels in *D. glomerata* than in a *Medicago truncatula* transcriptome where it was highly expressed. C4H is a major rate-limiting step in phenylpropanoid biosynthesis that separates the pinocembrin pathway from the lignin/monolignol and naringenin-based flavonoid branches. Shikimate *O*-hydroxycinnamoyltransferase, the link between

flavonoid biosynthesis and the lignin/monolignol pathway, was also expressed at much lower levels in *D. glomerata* than in *M. truncatula*. Our results indicate (a) a likely major role for flavonoids in actinorhizal nodules, and (b) differences in metabolic flux in flavonoid and phenylpropanoid biosynthesis between the different hosts in symbiosis.

INTRODUCTION

Root nodule symbioses (RNS) develop as symbiotic associations between nitrogen-fixing bacteria and certain host plants, resulting in the formation of the root nodule, a specialized organ for nitrogen fixation and assimilation. The root nodule provides a number of functions in RNS, primarily serving as a site for the exchange of carbon and energy-containing molecules from the host for nitrogen-containing molecules from the microsymbiont, and also as an environment to help regulate oxygen concentration to protect the nitrogenase enzyme complex. The bacteria capable of establishing these symbioses fall into two distantly related groups: the proteobacterial rhizobia and the actinobacterial genus *Frankia*. The host plants, on the other hand, all belong to a single clade of angiosperms known as the Nitrogen-fixing Clade (NFC) (Soltis et al., 1995), consisting of the order Fabales (nodulated by rhizobia) and three orders that include the actinorhizal plants, Cucurbitales, Fagales, and Rosales (nodulated by *Frankia*).

The establishment of RNS involves a signal–mediated recognition interaction between host and microsymbiont within the rhizosphere, followed by the entry of the microsymbiont into root cells, and ultimately by nodule organogenesis (Gage, 2004). Early stages of organogenesis involve the division of cortical cells and cell expansion during invasion by the microsymbiont, followed by nodule organogenesis and maturation of nitrogen-fixing symbiotic tissue. During the maturation phase, cells within the developing nodule undergo endoreduplication, increasing in volume and becoming more transcriptionally active to promote symbiotic interactions (Vinardell et al., 2003). In both the legume and actinorhizal symbioses several of the initial steps in the internal signaling pathway leading to nodule establishment are conserved. Initial signaling interactions are activated via the Common Symbiotic Pathway, a set of genes shared with the more ancient arbuscular mycorrhizal symbioses (Oldroyd, 2013), indicating a shared evolutionary origin within the NFC (Markmann and Parniske, 2009; Battenberg et al., 2018), followed by a RNS-specific gene expression cascade (Oldroyd, 2013).

Flavonoids are ubiquitous secondary metabolites synthesized by the phenylpropanoid pathway. The flavonoid pathway is one of two major branches in plant phenylpropanoids, the other being monolignol/lignin biosynthesis, and is responsible for producing a wide range of metabolites fundamental for plant structure and function and plant–organism interactions including symbiotic signaling in RNS and nodule organogenesis and development. Plant flavonoids are also key molecules in pigmentation and signaling for pollinator attraction, herbivore or pathogen deterrence, reduction of damage from reactive oxygen species, UV light protection, and regulation of development

(Shirley, 1996; Buer et al., 2010). The reactions linking the flavonoid and monolignol/lignin branches of the pathway and the interconversions among flavonoid classes are illustrated in **Figure 1**.

In the establishment of root nodule symbiosis in many legume genera, flavonoids produced by the host are recognized by rhizobia in the rhizosphere, primarily through the receptor-transcription factor NodD. This, in turn, triggers the expression of the other *nod* genes (*nodA, nodB, nodC*), which synthesize a lipochitooligosaccharide molecule, the Nod factor, which is secreted by the rhizobia (Long, 1996), that in turn triggers host cellular responses leading to root-nodule development (Oldroyd, 2013). A wide range of flavonoid molecules, both aglycones and glycosides, has been identified as nodulation signals in legume symbioses (Peck et al., 2006). To date, no molecule similar to Nod factor has been identified in the *Frankia*-actinorhizal symbioses; however, genomes of some members of the Cluster 2 group of *Frankia* contain homologs of the rhizobial *nodABC* genes that are expressed during symbiosis (Persson et al., 2015; Nguyen et al., 2016), suggesting that a Nod factor may play a role in at least some actinorhizal symbioses. Cluster 2 *Frankia* genomes do not contain any identified homologs of *nodD*, leaving the mechanism of induction of transcription unknown, and the role of flavonoids in actinorhizal-*Frankia* signaling to be determined.

After the initial signaling steps in nodulation, flavonoids play a continuing role in legume nodule development. Flavonoids are known to bind to and inhibit auxin transporters, leading to a disruption of polar auxin transport (Wasson et al., 2006). The accumulation of auxin within certain cortical cells in the root triggers cell division and proliferation, which, in turn, leads to the localized induction of the nodule (Mathesius et al., 1998). Recent studies have suggested that the production of flavonoids during nodule organogenesis is itself a response to increased cytokinin production following the perception of the symbiotic Nod factor (Mathesius et al., 1998). Similar effects in the actinorhizal symbioses have been far less studied; however, it has been shown that inhibition of auxin gradients with an auxin influx inhibitor in the actinorhizal host *Casuarina glauca* led to decreased nodulation, suggesting a similar role for auxin in actinorhizal symbioses (Péret et al., 2007; Champion et al., 2015). Additionally, flavonoids have been suggested to play a role in triggering endoreduplication through DNA breaks resulting in anaphase arrest (Cantero et al., 2006).

In this study, flavonoids from the metabolomes of roots and nodules of the actinorhizal host *Datisca glomerata* were compared with those of the legume *Medicago sativa*. Additionally, the metabolome results were compared with available transcriptomes of *D. glomerata* and *Medicago*
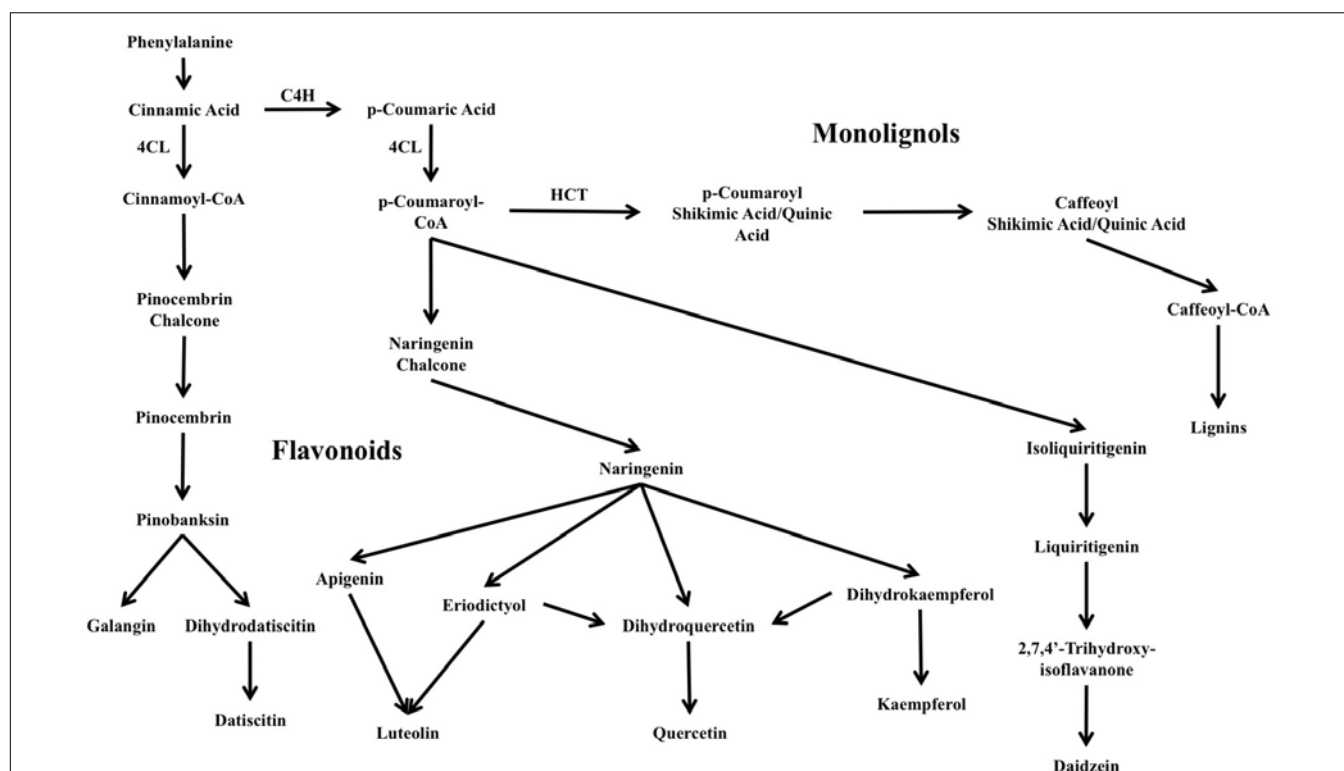
FIGURE 1 | Overview of major flavonoid biosynthesis pathways in *Datisca glomerata*. Also shown is the lignin/monolignol pathway (upper-right). C4H: cinnamic acid 4-hydroxylase. 4CL: 4-coumarate CoA-ligase. HCT: shikimate *O*-hydroxycinnamoyltransferase.

truncatula (Roux et al., 2014; Battenberg et al., 2018). Both hosts were found to synthesize phenylpropanoid derivatives of the flavonoid branch in several different categories including flavones, flavanones, and isoflavonoids, but with different apparent patterns of metabolic flux.

MATERIALS AND METHODS

Growth Conditions and Nodule Sampling

Datisca glomerata seeds were collected from wild plants growing in Gates Canyon, Vacaville, CA, United States, germinated, grown, and inoculated in a greenhouse at University of California, Davis, under conditions as described in Battenberg et al. (2018). The seedlings were inoculated with crushed *Ceanothus thyrsiflorus* nodules containing *Frankia* originally sampled in Sagehen Experimental Forest (Truckee, CA, United States). Until inoculation, one-half-strength Hoagland's solution with nitrogen (Hoagland and Arnon, 1950) was applied weekly. After inoculation, one-half-strength Hoagland's solution without nitrogen was applied weekly.

Uninoculated roots, inoculated roots, and root nodules were collected for analysis from four individual plants per treatment. Inoculated roots and nodules were collected from the same plants; both were harvested 100 days after inoculation. Uninoculated roots were collected from the same plants sampled for inoculated roots and nodules, prior to inoculation. Sampling

methods are described in detail in Battenberg et al. (2018). Collected samples were flash frozen in liquid nitrogen and stored at −80°C until use. For detailed information on samples collected, see **Supplementary Table S1**. Root and nodule samples from individual plants were ground in liquid nitrogen in a mortar and pestle, prior to extraction.

Mature individual *M. sativa* plants were collected from field plantings at the Russell Ranch Sustainable Agriculture Facility, University of California, Davis, and maintained in a greenhouse at University of California, Davis. Roots and root nodules were collected and sampled from four individual plants per treatment, as described above.

Flavonoid Extraction

Root nodules and inoculated roots of *M. sativa* and *Datisca glomerata* were extracted using 80:20 MeOH/H_2O. 40 mg of samples were extracted with 2000 µL of cold solvent. Samples were then mixed for 10 s using Mini Vortexer (VWR, Radnor, PA, United States). Samples were then centrifuged for 5 min at 14,000 relative centrifugal force (RCF) using an Eppendorf Centrifuge 5415D (Hauppaugee, NY, United States). After removing the supernatant, samples were dried using a Labconco CentriVap Concentrator (Kansas City, MO, United States). Dried samples were resuspended in 110 µL of 10:90 ACN/H_2O with 1 µg/mL 12-(cyclohexylcarbamoylamino)dodecanoic acid (CUDA) for LC-MS/MS analysis.

Metabolomic Analysis

For metabolomic analysis of flavonoids and related molecules, a comparison was made between root nodules and inoculated roots collected from the same plants. Chromatography was performed using a Thermo Vanquish UHPLC instrument, a Phenomenex Kinetex C18 column (100 × 2.1 mm, 1.7 μm) with a KrudKatcher Ultra HPLC in-line filter (0.5 μm Depth Filter × 0.004 in ID). The mobile phases were H_2O with 0.1% acetic acid (A) and ACN with 0.1% acetic acid (B). Gradient elution was performed at a flow rate of 0.5 mL/min under the following program: from 0 to 10 min B changed linearly from 10 to 90%, held at 90% B for 2.50 min, returned to 10% B over the next 2.50 min, and held at 10% B for equilibration over 5 min. The column temperature was kept at 45°C. The LC method was modified from Ma et al. (2016). MS/MS data were acquired on a high resolution Thermo Q Exactive HF mass spectrometer in positive electrospray ionization (ESI) mode under the following operating parameters: sheath gas flow rate at 60, auxiliary gas flow rate at 25, sweep gas flow rate at 2, spray voltage at 3.60 kV, capillary temperature at 300°C, S-lens RF level at 50, and auxiliary gas heater temperature at 370°C. Mass spectral data were collected using full scan MS1 and data-dependent MS/MS. Full scan MS1 had the following parameters: scan range from a m/z 150–2000 with the resolution set to 120,000, AGC target set to 1×10^6, and maximum ion injection set to 300 ms. Data-dependent MS2 had the following parameters; scan range from m/z 150–2000 with the resolution set to 15,000, AGC target set to 1×10^5, maximum injection time set to 50 ms, loop count set to 3, and TopN set to trigger the top-3 most abundant ions, with an isolation window of 1.0 m/z, and Higher Energy Collisional Dissociation (HCD) was conducted using three normalized collision energies; 35, 45, and 65%. The observed MS/MS spectra have an HCD collision energy of 48.33%, spectra from the three normalized collision energies are automatically averaged. The injection volume for each sample was 2 μL.

In metabolomics compounds are routinely identified by data processing tools which match MS/MS spectra against mass spectral reference libraries and use cheminformatics to provide spectral interpretation (Neumann and Bocker, 2010; Dunn et al., 2012; Cajka and Fiehn, 2015). Here we have used MS-DIAL software version 2.82 (Tsugawa et al., 2015) was used to process the raw data and metabolites were reported using a 0% peak count filter to keep all detected features. MS-DIAL was used for data deconvolution, peak alignment, and compound identification by searching mass spectral refrence libraries. Compound identifications were made based on an in-house accurate mass and retention time (m/z-RT) library created from the QC reference standard mix and the following tandem mass spectral libraries; MassBank, ReSpect, MetaboBASE, HMDB, GNPS, NIST 17 MS/MS, FAHFA, LipidBlast, and iTree MS/MS only. The tandem mass spectral libraries were downloaded in an msp format from MassBank of North America (MoNA) which was later used in MS-DIAL. MS-FLO (Mass Spectral Feature List Optimizer) was used as post processing tool to optimize the feature list from MS-DIAL to remove duplicate and isotopic features and identifiy ions adduct (DeFelice et al., 2017).

After reduction, annotations were also labeled with Metabolomics Standards Intitaive (MSI) levels and mass error (mDa) to provide confidence in each annotation. Level 1 is the highest level of identification. It is described as using two or more orthogonal data from an authentic standard. Level 2 is when only one set of reference data from an authentic standard has been used, for example, either using an in-house accurate m/z-RT library or using mass spectral library search for MS/MS matching. Level 3, is similar to Level 2 where a match can be made with either a m/z-RT library or a MS2 library, but the match lacks high accuracy. Lastly, Level 4 indicates those metabolites that are unknown (Sumner et al., 2007; Schymanski et al., 2014). The observed MS spectra of identified flavonoid compounds in *D. glomerata* and *M. sativa* are shown in **Supplementary Figures S1**, **S2**, respectively, as head-to-tail comparisons of experimental and reference MS/MS spectra.

Metabolome and Phenylpropanoid Pathway Analysis

Flavonoid molecules detected by LCMS were annotated by their International Chemical Identifier (InChIKey) and separated into subclasses of flavonoids, isoflavonoids, flavones, flavonols, and anthocyanins by ClassyFire (Feunang et al., 2016). For each plant, the average peak height of flavonoids in each plant was calculated and molecules with average peak heights above the mean were considered highly abundant. Significant differences between roots and nodules were identified with two-tailed Welch's t-tests. A comparison of the overall proportions of flavonoids annotated by class in the metabolomes of *D. glomerata* and *M. sativa* was performed with a chi-square test. T-tests and chi-square tests were performed in R using a significance level of $p < 0.05$.

For molecules with multiple annotated isotopes, only the dominant isotope was used, identified by the highest peak height across all samples. Phenylpropanoid biosynthesis pathways were obtained from KEGG (Kanehisa et al., 2016). Maps used included: Flavonoid Biosynthesis, Flavone and Flavonol Biosynthesis, Anthocyanin Biosynthesis, Isoflavonoid Biosynthesis, and Phenylpropanoid Biosynthesis. Within each class molecules that were significantly different between root and nodule LC-MS samples were grouped by their structural similarity using molecular structures acquired from PubChem (Kim et al., 2016) to reference molecules that included: eriodictyol, naringenin, liquiritigenin, daidzein, genistein, glycitein, formononetin, kaempferol quercetin, luteolin, apigenin, cyanidin, pinocembrin, pinobanksin, galangin, and chrysin. In cases where an enzyme for synthesizing a particular molecule could not be identified, putative pathways were inferred, placing molecules together in groups if their chemical structures shared diagnostic structures of the reference molecules including: a 2C–3C carbon double bond, a 3 carbon hydroxyl, 3′ or 4′ hydroxyls, or a 2 or 3 carbon benzene ring.

Abundant Compound Verification by HPLC, UV Absorption and LC-MS

To investigate the effect of *Frankia* interactions with *D. glomerata* on flavonoid production uninoculated roots, roots inoculated

with *Frankia*, and nodules of *D. glomerata* were collected for a second analysis of selected abundant flavonoids. The ground tissue (100 mg) was extracted in 300 μL of 80% methanol, with incubation in an ultrasonic water bath for 20 min at 30°C. The extract was then centrifuged twice for 10 min each at 17,000 × *g*. The supernatant was transferred to an HPLC vial; 10 μL of the supernatant was injected on a reverse phase HPLC and analyzed as previously described (Knollenberg et al., 2018). Major metabolites (i.e., abundant HPLC peaks) that exhibited differential accumulation among roots and nodules were collected and analyzed by mass spectrometry (MS) in negative mode and MS/MS using an established method (Ono et al., 2016). The flavonoid metabolites were tentatively identified based on their retention times, UV absorption spectra, as well as MS and MS/MS data, taking into consideration phenolic metabolites previously reported to accumulate in roots of Datiscaceae (Bohm, 1988). In addition, authentic standards of kaempferol, luteolin, and genistein (Sigma Aldrich, St. Louis, MO, United States) were analyzed in parallel with the *D. glomerata*.

One-way analysis of variation (ANOVA) followed by Tukey's HSD test were performed on the metabolite data using JMP (SAS Institute, Cary, NC, United States).

Transcriptome Analysis of *D. glomerata* and *M. truncatula* Phenylpropanoid Pathways

Transcript annotations from Battenberg et al. (2018) were used for *D. glomerata* and *M. truncatula* (Roux et al., 2014). The transcriptome of *M. truncatula* was chosen because of its depth of coverage and comparable stage of nodulation. Enzyme Commission (EC) number annotations were made with InterProScan v5.21 (Jones et al., 2014) and Trinotate v3.0.1 (Haas et al., 2013). Transcripts annotated with Enzyme Commission (EC) numbers belonging to the KEGG phenylpropanoid biosynthetic pathways listed above were identified in each transcriptome. Expression fold changes between nodules and roots (Log-scale) were used to generate heat maps in Microsoft Excel. To determine genes important for symbiosis and compare relative expression levels of genes between nodules of the two hosts, transcripts in the 90th percentile or above ranked by transcripts per million (TPM) in the full transcriptomes were considered "highly expressed," and transcripts below the 50th percentile were considered "low expression." Statistical significance of expression level differences between *D. glomerata* and *M. truncatula* nodules were determined with one-sample *t*-tests comparing the percentile ranks of the most highly expressed transcript for each gene in the *D. glomerata* transcriptome with the most highly expressed transcript from *M. truncatula* ($p < 0.05$).

RESULTS

Metabolomics Analysis Revealed Abundant and Diverse Flavonoid Accumulation in *D. glomerata* and *M. sativa*

In total, 384 compounds from *D. glomerata* roots and nodules were initially annotated as flavonoids, however, only 60

FIGURE 2 | Distribution of flavonoids in *D. glomerata* and *M. sativa* roots and nodules by class.
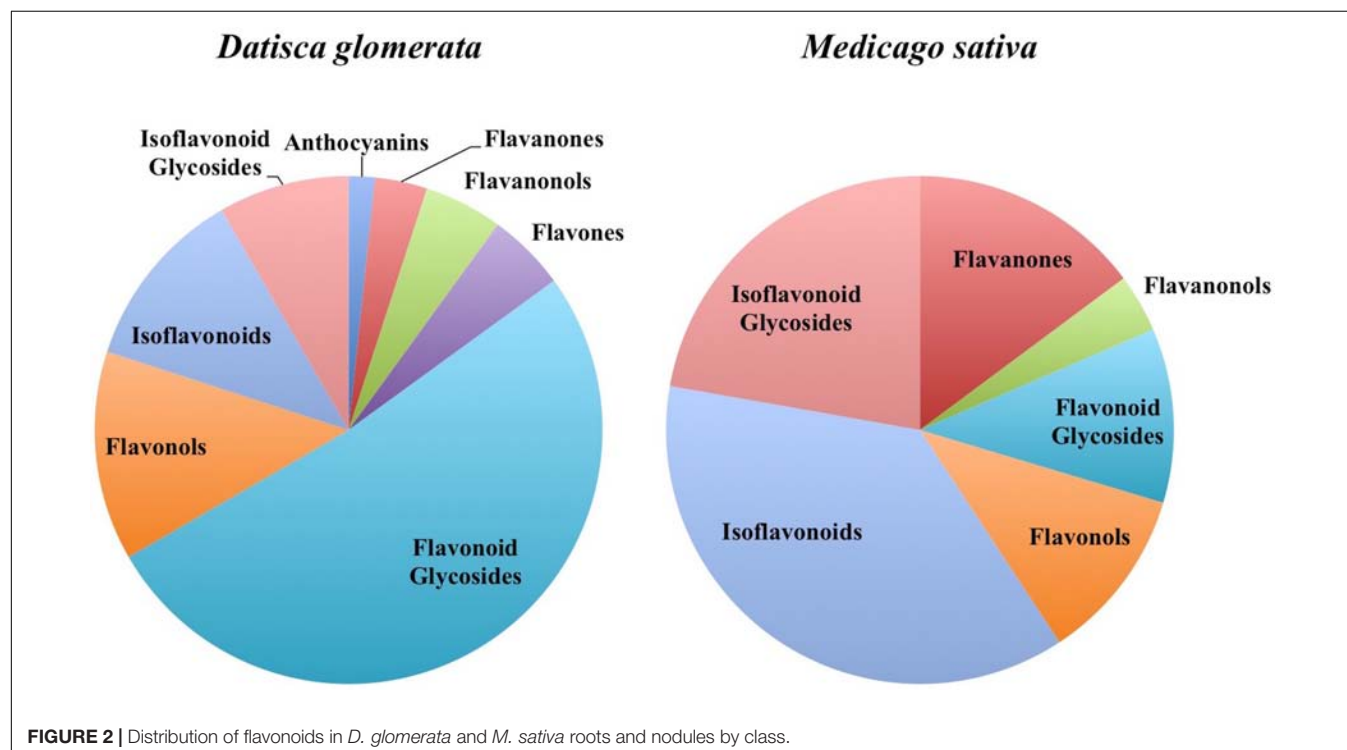
metabolites were matched against spectra from an MS/MS library at high levels of identification (MSI Level 1, 2, or 3) and the rest were reclassified as unknowns (**Supplementary Table S2**). The relative distributions of annotated flavonoids by class in *D. glomerata* and *M. sativa* are presented in **Figure 2** and listed in **Figures 3, 4** for *D. glomerata* and *M. sativa*, respectively. Of the 60 annotated flavonoids in *D. glomerata* 24 were aglycones and 36 were glycosides, the majority of which were flavonols (**Figure 3**). In *M. sativa* 281 compounds were initially annotated as flavonoids but only 27 compounds met the MSI Level 3 or better, of which nine were glycosylated. *M. sativa* root and nodule flavonoids were predominantly isoflavonoids (**Figure 4**). These differences in flavonoid distribution between *D. glomerata* and *M. sativa* were strongly significant ($p < 5 \times 10^{-8}$, **Supplementary Table S3**). With the exception of glycitin in *M. sativa* nodules all flavonoids identified were detected in both the roots and nodules of their respective plant (**Figures 3, 4**).

Ten flavonoids were highly abundant in nodules of *D. glomerata*. Abundant aglycones included isoquercitin (m/z 465.1021), quercetin (m/z 303.0495), galangin (m/z 271.0597), pinocembrin (m/z 257.0805), datiscetin (m/z 287.0545), cirsiliol (m/z 331.0807), and daidzein (m/z 255.0648), while the abundant glycosylated flavonoids included the datiscetin-derivative datiscin (m/z 595.1652) and two putative derivatives of kaempferol (m/z 449.1072 and m/z 639.1916) (**Figure 3**). The relative abundance of these flavonoids in *D. glomerata* roots and nodules, determined by t-tests, was not significantly different between roots and nodules.

In *M. sativa* nodules, the highly abundant flavonoids were exclusively isoflavonoids, including formononetin (m/z 269.0813), the most abundant, and its derivative the second most abundant flavonoid medicarpin (m/z 271.0968), as well as coumestrol (m/z 269.0449) and the genistein-derivative prunetin (m/z 285.0762) (**Figure 4**). Due to sample variability, statistically significant fold changes were not obtained for the majority of the identified *M. sativa* flavonoids between nodules and roots.

Differential Accumulation of Flavonoids in *D. glomerata* Roots and Nodules
Flavonoids
In *D. glomerata* nodules, naringenin and pinocembrin (m/z 257.0805), both of which are flavanones from which other flavonoid classes are derived, were significantly increased over roots (**Figure 3**). Dihydrokaempferol (m/z 271.0597), an aglycone derivative of naringenin that is an intermediate in the synthesis of flavonols from naringenin, was also significantly increased in the nodule.

Flavonols and Flavones
As noted above, the most abundant flavonoids in *D. glomerata* nodules were the flavonols datiscetin and datiscin, as well as galangin, all of which are derivatives of pinocembrin (**Figure 3**). Other flavonols, including kaempferol and quercetin, were found to have several known and putative derivatives significantly more abundant in the nodules than roots as well. Three putative kaempferol glycosides were highly abundant in *D. glomerata* nodules: astragalin (m/z 449.1071), kaempferol 7-*O*-glucoside

(m/z 449.1072), and demethoxycentaureidin 7-*O*-rutinoside (m/z 639.1916); and quercetin and its derivative isoquercetin (m/z 465.1021) were highly abundant in *D. glomerata* nodules.

No flavones, putative derivatives of apigenin and luteolin, were found to be significantly different between *D. glomerata* roots and nodules (**Figure 3**).

Isoflavonoids
Daidzein (m/z 255.0648) was one of the most abundant flavonoids in *D. glomerata* nodules and one of its derivatives, 6′-*O*-acetyldaidzin (m/z 459.1279) was significantly increased in nodules over roots (**Figure 3**). Additionally, genistein (m/z 271.0597), belonging to a separate isoflavonoid pathway, was also significantly more abundant in nodules than roots.

Anthocyanins
Only one anthocyanin was annotated from *D. glomerata*: peonidin 3-*O*-rutinoside (m/z 463.1229) and it was neither abundant nor significantly different in the nodule (**Figure 2**).

Rutinose Glycosides Accumulated in *D. glomerata* Nodules
A considerable number of rutinose glycosides were identified in the flavonoids of *D. glomerata* roots and nodules, across several of the flavonoid classes, including datiscin (m/z 595.1652), one of the most abundant flavonoids identified (**Figure 3**). Other rutinose glycosides detected included rutin (quercetin-3-*O*-rutinoside, m/z 611.1600), narirutin (naringenin-7-*O*-rutinoside, m/z 419.1331), peonidin-3-*O*-rutinoside (m/z 463.1229), demethoxycentaureidin-7-*O*-rutinoside (m/z 639.1916), and kaempferol-3-*O*-rutinoside (m/z 595.1653). None of these were significantly different between roots and nodules, e.g., more abundant in nodules than roots. Of the rutinose glycosides, only narirutin was identified in *M. sativa* roots or nodules (**Figure 3**).

Nodulation Enhanced Flavonoids in Inoculated *D. glomerata* Roots and Nodules
To examine how nodulation may influence flavonoid metabolites, we compared abundant metabolite composition of non-inoculated roots, inoculated roots, and nodules. Six metabolites (peaks 1–6) showed significantly greater accumulation in nodules in comparison to non-inoculated roots (**Figure 5**). Five (peaks 2–6) showed significantly greater accumulation in nodules than in the inoculated roots (**Figure 5**). In addition, two of the metabolites examined (peaks 2 and 3) showed significantly greater accumulation in inoculated roots when compared to non-inoculated roots. These metabolites were tentatively identified as structurally related flavonoids and flavonoid glycosides based on their retention times, UV absorption spectra, as well as MS and MS/MS data. The most abundant compound (peak 3) in all samples was tentatively identified as datiscetin (3,5,7,2′-Tetrahydroxyflavone) with other compounds (peaks 1, 2, and 5) potentially representing methylated or glycosylated derivatives. Peaks 4 and 6 were tentatively identified as kaempferol and galangin, respectively.

FIGURE 3 | Flavonoids identified in *D. glomerata* roots and nodules. All annotations conform to MSI level 2. For each flavonoid the abundance in both the roots and nodules is given as the log average peak height of 10 samples and presented as a heat-map. The names of molecules considered highly abundant are highlighted in blue. Significant changes in abundance between roots and nodules are highlighted in yellow.

The accumulation of these metabolites in roots and nodules are consistent with previous reported flavonoid profiles of *D. glomerata* (Bohm, 1988). These metabolite identifications were further supported by the observations that the retention time and UV absorption spectra for peak 3 (datiscetin; [M-H]⁻ m/z 285.0369) was inconsistent with an authentic standard of luteolin (m/z 286.2390) that has the same MS fragmentation

pattern as datiscetin. Similarly, the retention time and UV absorption spectra for peak 6 (galangin; [M-H]⁻ m/z 269.0434) were inconsistent with authentic standards of apigenin (m/z 270.2400) and genistein (m/z 270.2400) that share the same MS fragmentation pattern as galangin. The data for peak 4 was consistent with an authentic standard of kaempferol analyzed in parallel.

Medicago sativa

Annotation	m/z (Da)	Theoretical m/z (Da)	Mass error (mDa)	Adduct	Exact Mass (Da)	Dot Product	Reverse Dot Product	Retention Time	InChIKey	Root Average Peak Height (Arbitrary Units)	Nodule Average Peak Height (Arbitrary Units)	Fold Change (Root vs Nodule)	p-value (Root vs Nodule)	Class
Isoliquiritin	419.1343	419.1337	0.6	[M+H]+	418.1264	944.5847	943.5583	2.18	YNWXJFQOCHMPCK-LXGDFETPSA-N	5.89	6.03	0.14	0.627	Chalcone Glycoside
7-Hydroxyflavanone	241.0863	241.0859	0.4	[M+H]+	240.0786	981.1453	988.3962	4.56	SWAJPHCXKPCPQZ-UHFFFAOYSA-N	5.56	6.90	1.34	0.339	Flavanone
Liquiritigenin	257.0812	257.0808	0.4	[M+H]+	256.0735	905.5894	927.9195	3.31	FURUXTVZLHCCNA-UHFFFAOYSA-N	6.60	7.31	0.71	0.296	Flavanone
Pinocembrin	257.0813	257.0808	0.5	[M+H]+	256.0735	787.3966	907.5052	3.92	URFCJEUYXNAHFI-UHFFFAOYSA-N	5.63	6.11	0.48	0.495	Flavanone
Naringenin	273.0762	273.0758	0.4	[M+H]+	272.0685	-1	-1	2.72	FTVWIRXFELQLPI-ZDUSSCGKSA-N	5.73	6.13	0.39	0.408	Flavanone
Narirutin	419.1332	419.1337	0.5	[M+H-C6H10O5]+	580.1792	864.0896	811.4224	1.78	HXTFHSYLYXVTHC-AJHDJQPGSA-N	4.67	4.77	0.10	0.683	Flavanone Glycoside
Dihydrokaempferol	271.0606	271.0596	1.0	[M+H-H2O]+	288.0634	840.4518	822.2188	0.58	PADQINQHPQKXNL-LSDHHAIUSA-N	6.97	6.72	-0.25	0.058	Flavanonol
Naringin	581.1872	581.1865	0.7	[M+H]+	580.1792	-1	-1	1.85	DFPMSGMNTNDNHN-CSIAVLANSA-N	3.59	4.21	0.62	0.293	Flavanone Glycoside
Quercetin 3'-methyl ether	317.0661	317.0656	0.5	[M+H]+	316.0583	924.9923	975.8453	3.1	IZQSVPBOUDKVDZ-UHFFFAOYSA-N	4.03	4.26	0.23	0.575	Flavonol
Datiscetin	287.0555	287.0550	0.5	[M+H]+	286.0477	911.2479	889.0267	2.95	WCNLFPKXBGWWDS-UHFFFAOYSA-N	4.31	4.29	-0.02	0.956	Flavonol
3-Hydroxy-3'-methoxyflavone	269.0812	269.0808	0.4	[M+H]+	268.0735	783.5383	962.6492	4.56	GYLGASXCHPNKHD-UHFFFAOYSA-N	7.25	5.63	-1.62	0.322	Flavonol
Coumestrol	269.0449	269.0445	0.4	[M+H]+	268.0372	972.4412	959.3188	2.9	ZZIALNLLNHEQPJ-UHFFFAOYSA-N	6.69	7.66	0.97	0.165	Isoflavonoid
Daidzein	255.0656	255.0652	0.4	[M+H]+	254.0579	-1	-1	2.18	ZQSURDFPHDXJC-UHFFFAOYSA-N	7.23	7.06	-0.17	0.493	Isoflavonoid
Medicarpin	271.0968	271.0965	0.3	[M+H]+	270.0892	940.9747	825.0352	3.75	NSRJSISNDPOJOP-BBRMVZONSA-N	8.35	8.03	-0.31	0.228	Isoflavonoid
Genistein	271.0605	271.0601	0.4	[M+H]+	270.0528	-1	-1	2.79	TZBJGXHYKVUXJN-UHFFFAOYSA-N	5.35	4.97	-0.38	0.104	Isoflavonoid
Prunetin	285.0762	285.0758	0.4	[M+H]+	284.0685	742.4944	674.7249	2.93	KQMVAGISDHMXJJ-UHFFFAOYSA-N	8.06	7.66	-0.41	0.101	Isoflavonoid
Glycitein	285.0762	285.0758	0.4	[M+H]+	284.0685	-1	-1	2.36	DXYUAIFZCFRPTH-UHFFFAOYSA-N	5.54	5.13	-0.41	0.063	Isoflavonoid
Glycitin	447.1293	447.1286	0.7	[M+H]+	446.1213	-1	-1	1.09	OZBAVEKZGSOMOJ-MIUGBVLSSA-N	4.13	0.00	0.00	0.066	Isoflavonoid Glycoside
Sissotrine	469.1113 / 447.1294	469.1105 / 447.1286	0.8 / 0.8	[M+Na]+ / [M+H]+	446.1213	751.821	788.7471	1.72	LFEUICHQZGNOHD-UHFFFAOYSA-N	6.31	6.12	-0.18	0.271	Isoflavonoid Glycoside
Ononin	453.1163 / 431.1343	453.1156 / 431.1337	0.7 / 0.6	[M+Na]+ / [M+H]+	430.1264	965.7795	839.9406	2.16	MGJLSBDCWOSMHL-MIUGBVLSSA-N	7.80	7.36	-0.44	0.036	Isoflavonoid Glycoside
6''-O-Malonylglycitin	533.1301	533.1290	1.1	[M+H]+	532.1217	836.0235	702.8371	2.31	OWMHCYFEIJPHFB-GOZZSVHWSA-N	6.79	5.98	-0.81	0.134	Isoflavonoid Glycoside
Formononetin 7-O-glucoside-6''-O-malonate	517.1349	517.1341	0.8	[M+H]+	516.1268	864.2612	773.4846	2.67	RDTAGQKYPGLCBK-GOZZSVHWSA-N	8.09	7.16	-0.93	0.030	Isoflavonoid Glycoside
6''-O-Malonyldaidzin	503.1194	503.1184	1.0	[M+H]+	502.1111	864.9066	739.4436	1.86	MTXMHWSVSZKYBT-ASDZUOGYSA-N	6.24	5.07	-1.18	0.076	Isoflavonoid Glycoside
Formononetin	269.0813	269.0808	0.5	[M+H]+	268.0735	914.7	967.7452	3.38	HKQYGTCOTHHOMP-UHFFFAOYSA-N	9.40	8.95	-0.45	0.038	Isoflavonoid, O-methylated
3',4'-Dimethoxy-7-hydroxyisoflavone	299.0919	299.0914	0.5	[M+H]+	298.0841	836.9082	916.3702	3	VFZJLPRJAQGFO-UHFFFAOYSA-N	6.54	5.90	-0.64	0.083	Isoflavonoid, O-methylated
6,4'-Dimethoxy-7-hydroxyisoflavone	299.0919	299.0914	0.5	[M+H]+	298.0841	932.1259	870.8793	3.55	KJGPBYUQZLUKLL-UHFFFAOYSA-N	6.51	5.86	-0.65	0.043	Isoflavonoid, O-methylated
Biochanin A	285.0762	285.0758	0.4	[M+H]+	284.0685	881.004	911.9885	4.15	WUADCCWRTIWANL-UHFFFAOYSA-N	7.56	6.91	-0.65	0.154	Isoflavonoid, O-methylated

| Scale	Log Root	Log Nodule	Log Fold Change
	10.00	10.00	2.00
	5.00	5.00	1.00
	0.00	0.00	0.00
			-1.00
			-2.00

FIGURE 4 | Flavonoids identified in *M. sativa* roots and nodules. All annotations conform to MSI level 2. For each flavonoid the abundance in both the roots and nodules is given as the log average peak height of four samples and presented as a heat-map. The names of molecules considered highly abundant are highlighted in blue. Significant changes in abundance between roots and nodules are highlighted in yellow.

Expression of Phenylpropanoid Pathway Genes in Transcriptomes of *D. glomerata* and *M. truncatula* Roots and Nodules

To understand the expression of the phenylpropanoid and flavonoid biosynthetic genes in roots and nodules, transcriptome data from roots and nodules of *D. glomerata* (Battenberg et al., 2018) and *M. truncatula* (Roux et al., 2014) were analyzed by differential expression between tissues and by comparing relative expression levels (as percentiles) of transcripts within each transcriptome. The reactions of the phenylpropanoid pathway, including the flavonoid branch, and the lignin-monolignol branch, are summarized in **Figure 1** and the relative expression of the genes comprising each branch are depicted in **Figure 6**. Early steps in the phenylpropanoid pathway preceding the flavonoid branch, e.g., phenylalanine ammonia-lyase (EC 4.3.1.24), and 4-coumarate-CoA ligase (4CL, EC 6.2.1.12) were highly expressed, above the 95th percentile in roots and nodules of both *D. glomerata* and *M. truncatula*, but were not significantly up-regulated or down-regulated in nodules of either host (**Figure 6**). The enzyme 4CL is responsible for the conversion of cinnamic acid to cinnamoyl-CoA and p-coumaric acid to p-coumaroyl-CoA, the precursor

to flavonoid chalcones (**Figure 1**). In *D. glomerata* nodules, of the nine annotated 4CL transcripts, seven were not up-regulated relative to roots, while two (DgTrNR01535_a1_i1 and DgTrNR01535_a1_i2) were up-regulated over four-fold. In *M. truncatula* nodules, several annotated transcripts of this gene were also up-regulated between four- and six-fold, while one transcript was up-regulated greater than 300-fold (**Figure 6**).

Two of the four annotated transcripts of cinnamic acid 4-hydroxylase (C4H, EC 1.14.13.11), also known as *trans*-cinnamate 4-monooxygenase, were down-regulated in *D. glomerata* nodules almost 10-fold, while the other two were not statistically different (**Figure 6**) and none were expressed above the 87th percentile. C4H is the enzyme that catalyzes the last step in phenylpropanoid biosynthesis that precedes the separation of the naringenin flavonoid branch and the lignin-monolignol branch (**Figure 1**). In *M. truncatula*, by contrast, no C4H transcript was down-regulated in the nodule and one transcript (Medtr5g075450) was extremely highly expressed, above the 98th percentile of all genes in the transcriptome (**Figure 6**). The most highly expressed C4H transcripts expressed at significantly different levels in the
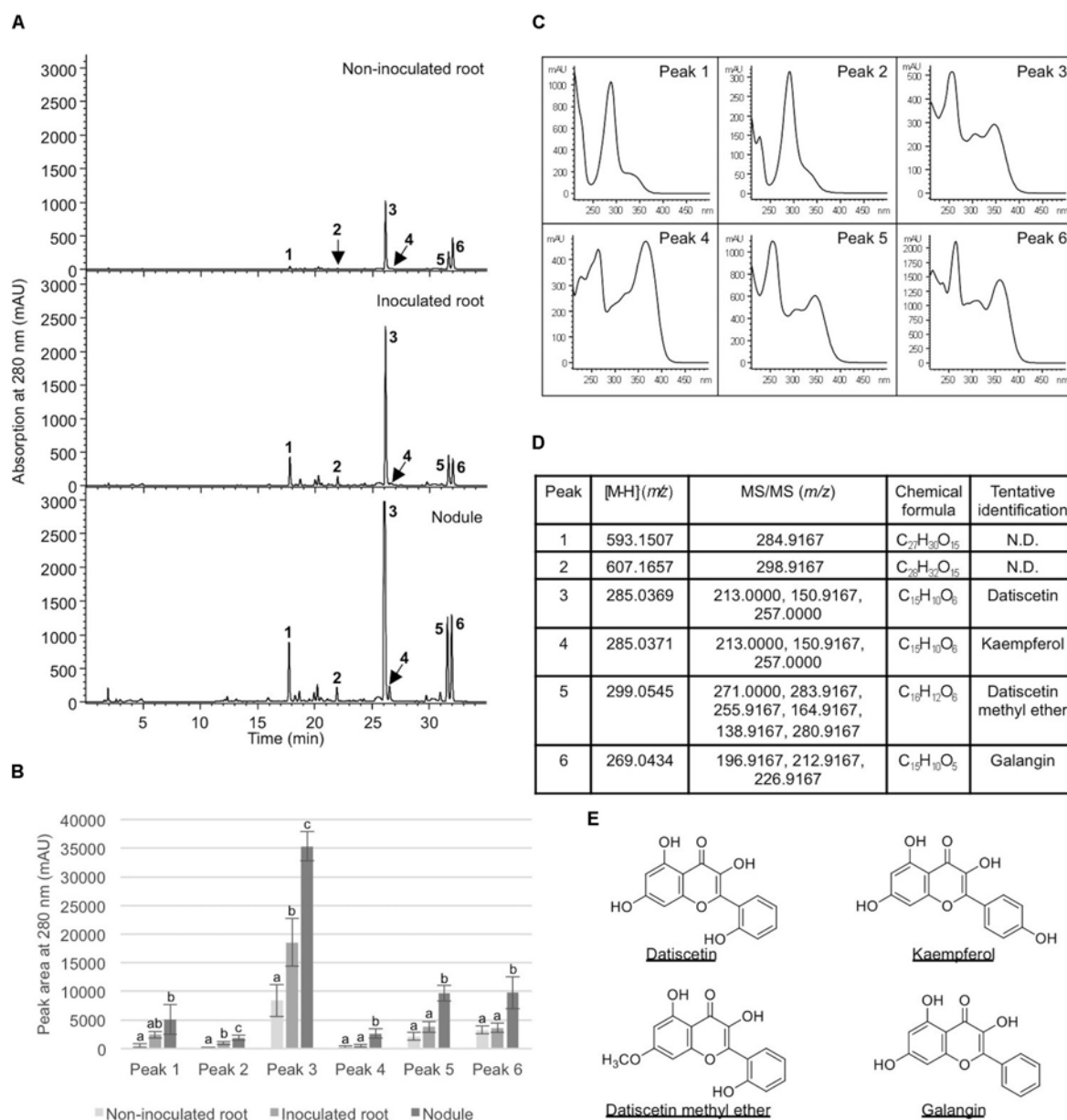
FIGURE 5 | Analysis of metabolites in *Datisca glomerata* roots and nodules. **(A)** Representative HPLC elution profiles of phenolic metabolites extracted from non-inoculated and inoculated *D. glomerata* roots as well as nodules. Peaks (1–6) that show differential accumulation in roots and nodules are indicated. **(B)** Peak areas of differentially accumulated phenolic metabolites. Data shown are the average of four biological replicates ± standard deviation. Different letters indicate significant ($P < 0.05$) differences in metabolite levels among non-inoculated root, inoculated root, and nodule for each peak. **(C)** Absorption spectra of peaks 1–6. **(D)** MS and MS/MS analyses of peaks 1–6. **(E)** Chemical structures of tentatively identified phenolic metabolites in *D. glomerata* roots and nodules.

transcriptomes of *D. glomerata* and *M. truncatula* nodules ($p < 5 \times 10^{-4}$).

As shown in **Figure 4**, in the flavonoid branch, one transcript of chalcone isomerase (EC 5.5.1.6), which catalyzes the synthesis of flavonoids from chalcones, was highly up-regulated in *D. glomerata* nodules (approximately 50-fold) while the second was not significantly different. Of the 10 chalcone isomerase transcripts in *M. truncatula*, on the other hand, none were up-regulated in the nodule, and

half of them were significantly down-regulated. Transcripts annotated as encoding flavanone 3-dioxygenase (EC number 1.14.13.21), the enzyme responsible for the synthesis of both eriodictyol from naringenin and dihydroquercetin from dihydrokaempferol (precursors to flavones and flavanols, respectively), were also among the most up-regulated flavonoid biosynthesis genes identified in the *D. glomerata* nodule (approximately 20- and 100-fold). By contrast, *M. truncatula* showed down-regulation of these genes in

the nodule or no significant difference between nodules and roots.

Four transcripts were annotated as flavonol synthase (flavonol biosynthesis, EC 1.14.11.23) in *D. glomerata*, one of which was significantly up-regulated in the nodule and was expressed in the 99th percentile in the nodule (**Figure 6**). In *M. truncatula* the corresponding transcripts were not significantly different between roots and nodules and were expressed in the 57th percentile at most, a marked contrast. Three flavonol-3-*O*-glucosyltransferases (EC 2.4.1.91) were annotated in *D. glomerata*; one of which showed high expression (above the 93rd percentile) with no significant change in expression between nodules and roots. In the transcriptome of *M. truncatula*, no transcripts were annotated as flavonol-3-*O*-glucosyltransferases.

Strikingly, no transcripts in the isoflavonoid pathway were annotated in the *D. glomerata* transcriptome, whereas, in the *M. truncatula* transcriptome, there were multiple transcripts annotated as encoding several enzymes in isoflavonoid biosynthesis (**Figure 6**). One of the eight transcripts of 2,7,4′-trihydroxyisoflavanone 4′-*O*-methyltransferase (EC 2.1.1.212) was highly expressed, above the 93rd percentile while four others were up-regulated over 10-fold in *M. truncatula* nodules. Enzymes that catalyze the synthesis of daidzein derivatives including formononetin and daidzein 7-*O*-glucoside were expressed but not significantly different between roots and nodules. Three transcripts of isoflavone 7-*O*-glucoside-6″-*O*-malonyltransferase (EC 2.3.1.115), an enzyme known to synthesize malonated daidzein derivatives, were annotated in *M. truncatula*. One was down-regulated over 10-fold in the *M. truncatula* nodule relative to the roots, falling to around the 45th percentile in the nodule while the other two remained around the 72nd percentile.

The gene encoding the earliest enzyme in anthocyanin biosynthesis, anthocyanidin 3-*O*-glucosyltransferase (EC 2.4.1.115), was up-regulated more than 10-fold in *D. glomerata* nodules and two other transcripts were expressed above the 90th percentile (**Figure 6**). By contrast, no anthocyanidin 3-*O*-glucosyltransferase transcripts were annotated in the nodule or root transcriptomes of *M. truncatula*.

Genes encoding the major enzymes in the lignin-monolignol branch of the phenylpropanoid pathway showed generally low expression in the transcriptome of *D. glomerata* (**Figure 6**). The most highly expressed transcript of the first enzyme in this branch, shikimate *O*-hydroxycinnamoyltransferase (HCT, EC 2.3.1.133) was expressed in the 48th percentile. Coumaroyl-shikimate 3′-monooxygenase (EC 1.14.13.36), the next gene in the pathway, was expressed in the 61st percentile. Caffeoyl CoA-*O*-methyltransferase (EC 2.1.104), converting caffeoyl-CoA to feruloyl-CoA, was the exception. Two transcripts were expressed in the 90th percentile in *D. glomerata* nodules, although transcripts were expressed up to 10-fold higher in the roots. In *M. truncatula* nodules, the transcripts encoding HCT were expressed much more highly than HCT in *D. glomerata* ($p < 5 \times 10^{-5}$), around the 80th percentile. No transcripts of coumaroyl-shikimate 3′-monooxygenase were annotated in *M. truncatula*.
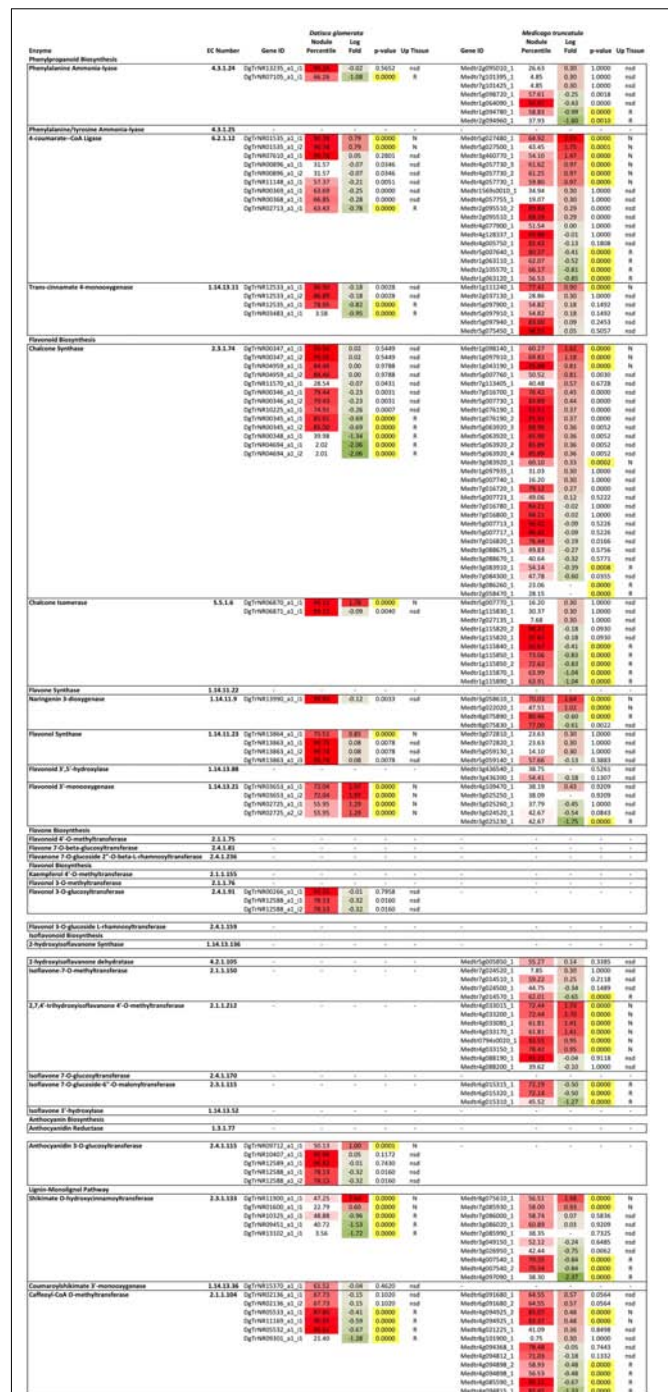
FIGURE 6 | Heat map of flavonoid biosynthesis gene expression in *D. glomerata* and *M. truncatula* including relative expression in the nodule transcriptome based on percentile of all genes in their respective transcriptome (Percentile) and fold change between nodules and roots (Log fold change). Significance level used for differences in expression between roots and nodules was $p < 0.001$. Up-regulated tissues: N, nodules; R, roots; and nsd, no significant difference.

DISCUSSION

Metabolite and Expression Analyses Provide New Insights Into Flavonoid Metabolism in *D. glomerata* Roots and Nodules

A range of flavonoids were synthesized in roots and nodules of *D. glomerata*, including flavones, flavonols, flavanones, anthocyanins, and isoflavonoids. A greater number of flavonoids were more abundant in the nodule relative to roots than vice versa (**Figure 3**), suggesting an overall increase in flavonoid biosynthesis in the nodule.

For several highly-abundant metabolites analyzed, there was a general trend of increasing concentration, when comparing uninoculated roots with either roots post-inoculation, or root nodules (**Figure 5**). This suggests that inoculation with *Frankia* induces some change in flavonoid metabolism in roots, either systemically in nodulated plants, or locally by association with *Frankia* in the rhizosphere. In a split-root experiment on *M. sativa,* initiation of nodulation by application of either the symbiont *Sinorhizobium meliloti* or its Nod factor to roots on one side of the plant led to increased amounts of daidzein on the uninoculated side (Catford et al., 2006), supporting an interpretation that flavonoid biosynthesis and distribution is under global regulation in RNS, similar to, or part of, autoregulation (Reid et al., 2011).

Derivatives of the pinocembrin-derived subclass of flavonoids, especially datiscetin and datiscin, represented some of the most abundant molecules in *D. glomerata* nodules (**Figure 3**). The biosynthesis of datiscetin has been proposed to proceed by a reaction similar to the synthesis of galangin from pinobanksin through the addition of a 2C–3C double bond to dihydrodatiscetin that is itself synthesized from pinobanksin (Grambow and Grisebach, 1971). Galangin biosynthesis utilizes flavonol synthase (EC 1.14.11.23), which was found to be up-regulated four-fold in *D. glomerata* nodules (**Figure 6**). Because flavonol synthase has been shown to catalyze multiple reactions including both kaempferol and galangin biosynthesis (Miyahisa et al., 2006), it seems likely that datiscetin biosynthesis could be performed by this enzyme as well. Dihydrodatiscetin itself, however, was not identified in roots or nodules of *D. glomerata* in this study (**Figure 3**).

The prevalence of pinocembrin and its derivatives in *D. glomerata* is likely directly related to the relatively low expression of C4H (EC 1.14.13.11) (**Figure 5**), which appears to be a pivotal enzyme in the phenylpropanoid pathway in *D. glomerata* whose expression impacts two separate branch points (**Figure 1**): first, the enzyme catalyzes the conversion of the pinocembrin-precursor cinnamic acid to naringenin-precursor p-coumaric acid, and thus controls the balance between the pinocembrin and naringenin flavonoid branches. Because the early enzymes in the flavonoid branch, including chalcone synthase (EC 2.3.1.74) and naringenin 3-dioxygenase (EC 1.14.11.9) are multi-functional, catalyzing reactions with multiple substrates (Martens et al., 2010), the altered flux favoring cinnamoyl-CoA in *D. glomerata* likely directs the

flow of flavonoid biosynthesis more toward pinocembrin and ultimately datiscetin and datiscin. Second, because expression of C4H is diminished relative to *M. truncatula,* the synthesis of naringenin-based flavonoids is likely aided in *D. glomerata* by lower expression of HCT (EC 2.3.1.133), which decreases metabolic flux to lignins in favor of flavonoid biosynthesis, a pattern similar to what was shown to occur when HCT was down-regulated in *M. sativa* by Gallego-Giraldo et al. (2014).

C4H has been shown to be the major rate-limiting step in lignin biosynthesis (Anterola and Lewis, 2002) suggesting that it also functions to control the relative flux of phenylpropanoids between flavonoid and lignin biosynthesis. Control of flux between flavonoid and lignin branches of the phenylpropanoid pathway in plants could be useful for crop improvement to enhance flavonoid content, particularly the pinocembrin pathway. Flavonoids in general are nutritional antioxidants, and pinocembrin specifically has shown both antitumor and neuroprotective capabilities (Rasul et al., 2013).

In addition to datiscetin, a flavonol synthesized from pinocembrin as discussed above, other flavonols were abundant in *D. glomerata* roots and nodules, particularly quercetin and its glycosides (**Figure 3**). Flavonol synthase (EC 1.14.11.23) was very highly expressed in *D. glomerata* roots and nodules, with one transcript expressed above the 99th percentile, compared to 57th percentile at the highest in *M. truncatula* nodules; and the other transcript was significantly up-regulated in the *D. glomerata* nodule seven-fold (**Figure 6**), reflecting the great abundance of flavonols, which may perform a variety of roles in nodules. Quercetin has been shown to regulate auxin gradients in roots during nodule formation as measured by a *gusA* gene-auxin response promoter construct (Mathesius et al., 1998). Quercetin showed a higher level of auxin transport inhibiting activity than kaempferol, apigenin, naringenin, or genistein; glycosylation decreased the auxin-inhibitory effect (Mathesius et al., 1998, 2015). Interestingly, in the laser-capture microdissection study of developmental gene expression in *M. truncatula* performed by Roux et al. (2014), 100% of flavonol synthase transcripts (Medtr5g059140) in nodules were found in zone FIID, the second distal fraction of the nodule. Cells in this zone are undergoing expansion and rhizobial infection (Roux et al., 2014), making it the zone where auxin gradient inhibition is likely required during the nodulation process (Mathesius et al., 2015). Additionally, quercetin has been shown to arrest cell division and cause DNA breaks leading to endoreduplication in eukaryotic cells *in vitro* (Cantero et al., 2006). This may suggest another potential role for flavonols in the formation of symbiotic cells in nodules (Vinardell et al., 2003), in addition to CCS52A-mediated endoreduplication, as described by Adachi et al. (2011). Finally, flavonols have been shown to protect enzymes from deactivation by nitric oxide and peroxynitrite (Heijnen et al., 2001). Nitric oxide is a signaling molecule required for nodule formation, however, it also deactivates enzymes important for symbiosis, including nitrogenase and glutamine synthetase, by nitration of tyrosine residues (Melo et al., 2011). Flavonoids in general have been found to protect enzymes by scavenging nitric oxide and peroxynitrite. Heijnen et al. (2001), found that galangin had strong scavenging activity and correlated this activity with the

hydroxyl group on the third carbon; this would suggest that datiscetin is a strong peroxynitrite scavenger as well.

A feature of the flavonoid glycosides in *D. glomerata* roots and nodules was the frequent occurrence of rutinose glycosides. Six flavonoids were identified with rutinose glycosylations. Rutinose is a common glycosylation of flavonoids in many plants (Cuyckens et al., 2001), and was previously shown to occur in leaves of *Datisca cannabina* (Bohm, 1988). Rutinose was reported to occur as an exceptionally abundant free sugar in roots, nodules, and leaves of *D. glomerata* and *D. cannabina* (Schubert et al., 2010) that could not be hydrolyzed in cell extracts. Thus it is hypothesized that rutinose is synthesized as a flavonoid glycosylation and then released as a free sugar (Schubert et al., 2010).

Potential Nod Gene Inducing Flavonoids in *Datisca glomerata*

In legume symbioses a range of flavonoids, predominantly aglycones, have been shown to induce the expression of *nod* genes in rhizobia, including flavones (luteolin), flavanones (eriodictyol and naringenin), and chalcones in *Medicago* and isoflavonoids (daidzein and genistein) in *Glycine* (Phillips and Tsai, 1992). In this study all of these metabolites were identified in roots and nodules of both *D. glomerata* and *M. sativa*, with the exception of chalcones, which were not detected in either host (**Figures 3**, **4**). The identified molecules could play potentially similar roles in the *D. glomerata* symbiosis as in *Medicago*. *D. glomerata* is nodulated by Cluster 2 *Frankia* strains whose genomes have been shown to contain *nodABC* gene homologs that are expressed in symbiosis (Persson et al., 2015; Nguyen et al., 2016); however, it is unknown whether flavonoids are involved in the regulation of these genes since no clear *nodD* homolog has been identified (Persson et al., 2015). Our results indicate that molecules in the pinocembrin pathway, including galangin and datiscetin, are possible candidates for a similar role in *D. glomerata*, because they are synthesized in great abundance (**Figure 3**) and are synthesized by uninoculated roots as well as inoculated roots and nodules (**Figure 5**) suggesting they are present in roots before the roots are infected.

Pinocembrin, the precursor to galangin and datiscetin, has been reported in several actinorhizal hosts. In the Fagales, it was found in the leaves and flowers of members of the genus *Alnus* (Betulaceae; Ren et al., 2017), as well as in leaves of *Myrica* and *Comptonia* (Myricaceae), all hosts nodulated by *Frankia* belonging to Cluster 1 (Wollenweber et al., 1985; Normand et al., 1996). However, flavonoids from leaf exudates of eight species of *Ceanothus* (Rhamnaceae, in Rosales), another genus that, like *D. glomerata*, is nodulated by Cluster 2 *Frankia* (Normand et al., 1996), were not reported to include pinocembrin or its derivatives (Wollenweber et al., 2004).

D. glomerata and *M. sativa* Share Similar Classes of Flavonoids but Differ in Abundance

Both *D. glomerata* and *M. sativa* metabolomes contained flavonoids in a range of classes. Both hosts contained similar

classes of flavonoids, including flavanones, flavonols, flavonoid glycosides, isoflavonoids, and isoflavonoid glycosides, but within each class they varied significantly in diversity (**Figure 2**). In both hosts the abundance of the majority of flavonoids was not statistically different between their roots and nodules.

The largest differences between the two plants were in the amount within particular flavonoid classes produced, most notably, in the pinocembrin pathway as discussed above, and in the isoflavonoids. All four of the highly abundant flavonoids in *M. sativa* nodules were isoflavonoids (**Figure 4**). This conforms to earlier reports highlighting isoflavonoids as most abundant in nodules of *M. sativa* (Tiller et al., 1994) and *M. truncatula* (Modolo et al., 2007; Staszkow et al., 2011). In *D. glomerata*, however, only one of the 10 highly abundant flavonoids (daidzein) was an isoflavonoid (**Figure 3**). Isoflavonoids identified in *D. glomerata* were found early in the biosynthetic pathway (**Figure 3**) whereas *M. sativa* included highly abundant derived isoflavonoids, including the pterocarpin medicarpin and its precursor formononetin (**Figure 4**). Pterocarpins are primarily found in legumes (Dewick, 1982) and have been shown to function as antimicrobials that show much greater inhibition of gram-positive bacteria than gram-negatives (Gnanamanickam and Smith, 1980). However, Auguy et al. (2011) reported expression of isoflavone reductase in the actinorhizal plant *C. glauca*, suggesting the synthesis of the more derived isoflavonoids similar to the *Medicago*. This leaves the distribution and role of isoflavonoids in actinorhizal plants unresolved.

CONCLUSION

We present the first comparison of metabolic profiles of flavonoids from both roots and nodules of two host plants within the NFC, *D. glomerata* and *M. sativa*, with transcriptomes obtained from roots and nodules, in the context of phenylpropanoid biosynthetic pathways. The most abundant flavonoids in *D. glomerata* were derivatives of pinocembrin as well as naringenin whereas flavonoids from *M. sativa* were isoflavonoids and derivatives of naringenin. These findings correlate with the pattern of expression of cinnamic acid 4-hydroxylase (C4H), in the transcriptomes of the two hosts. *D. glomerata* showed relatively low expression of C4H in nodules compared to *M. truncatula*, suggesting a role for this enzyme in directing the flow of the phenylpropanoid pathway between the pinocembrin branch and the naringenin branch. Similarly, shikimate *O*-hydroxycinnamoyltrasferase (HCT), the link between the flavonoid and monolignol branches of the phenylpropanoid pathway, also showed lower expression in *D. glomerata*, supporting a difference in metabolic flux between the two hosts that favors flavonoids over monolignol/lignin production in *D. glomerata*.

Flavonoids of the same classes were present in roots and nodules of both *D. glomerata* and *M. sativa*, including flavanones, flavonols, and isoflavonoids, suggesting similar roles for flavonoids during nodule development and symbiosis across lineages in the NFC. Common roles may include

symbiotic signaling, protection of enzymes from nitration, nodule organogenesis including phytohormone regulation, and cell-cycle modification. To identify symbiotically important flavonoids, further higher resolution transcriptome studies including spatio-temporal sampling (as in Roux et al., 2014, or Larrainzar et al., 2015) in combination with metabolomics profiling are needed. Secondly, responses of *Frankia* in culture to purified flavonoids identified as unique or amplified in their respective hosts should be measured at the transcriptomic level, and in terms of nodulation patterns, to evaluate a broader role for flavonoids as signaling molecules in the actinorhizal symbioses.

DATA AVAILABILITY STATEMENTS

Metabolome data from *D. glomerata* and *M. sativa* generated in this study are presented in **Supplementary Table S2**. Transcriptome data for *D. glomerata* and *M. truncatula* used were obtained from Roux et al. (2014) and Battenberg et al. (2018), respectively.

AUTHOR CONTRIBUTIONS

IG developed metabolomic and transcriptomic profiles, determined patterns of flavonoid metabolism, and substantially wrote the manuscript; each author contributed to the writing of the manuscript for their, respective, section. In addition, KB provided annotated transcriptomes and was responsible for the plant material. AV developed analytical methods for and performed metabolomic analyses. AW performed LC-MS (MS/MS) analyses on peaks identified by HPLC and analyzed glycosyltransferase transcriptome data. LT carried out HPLC analyses and provided data interpretation. OF contributed to metabolomics project design and manuscript editing. AB provided project oversight and contributed to manuscript construction.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2018.01463/full#supplementary-material

FIGURE S1 | Head-to-tail comparisons of MS/MS spectra of the most abundant flavonoids identified in *D. glomerata*. Reference library spectra are shown in red, experimental spectra are given in blue. Metadata include the similarity dot product and reverse dot product scores, InChIKey, and the structure of annotated compounds.

FIGURE S2 | Head-to-tail comparisons of MS/MS spectra of the most abundant flavonoids identified in *M. sativa*. Reference library spectra are shown in red, experimental spectra are given in blue. Metadata include the similarity dot product and reverse dot product scores, InChIKey, and the structure of annotated compounds.

TABLE S1 | Collection data for root and nodule samples of *D. glomerata* and *M. sativa*.

TABLE S2 | Flavonoid metabolome data obtained from LCMS analyses of *D. glomerata* and *M. sativa* nodules and roots.

TABLE S3 | Number of flavonoids annotated in each class in *D. glomerata* and *M. sativa* roots and nodules and chi-square test comparing the distribution of classes by proportion of total flavonoids in each host.

REFERENCES

Adachi, S., Minamisawa, K., Okushima, Y., Inagaki, S., Yoshiyama, K., Kondou, Y., et al. (2011). Programmed induction of endoreduplication by DNA double-strand breaks in Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 108, 10004–10009. doi: 10.1073/pnas.1103584108

Anterola, A. M., and Lewis, N. G. (2002). Trends in lignin modification: a comprehensive analysis of the effects of genetic manipulations/mutations on lignification and vascular integrity. *Phytochemistry* 61, 221–294. doi: 10.1016/S0031-9422(02)00211-X

Auguy, F., Abdel-Lateif, K., Doumas, P., Badin, P., Guerin, B., Bogusz, D., et al. (2011). Activation of the isoflavonoid pathway in actinorhizal symbioses. *Funct. Plant Biol.* 38, 690–696. doi: 10.1071/FP11014

Battenberg, K., Potter, D., Tabuloc, C., Chiu, J. C., and Berry, A. M. (2018). Comparative transcriptomics of two actinorhizal plants and the legume Medicago truncatula support the homology of root nodule symbioses and is congruent with a two-step process of evolution in the nitrogen-fixing clade of angiosperms. *Front. Plant Sci.* doi: 10.3389/fpls.2018.01256

Bohm, B. A. (1988). Flavonoid systematics of the datiscaceae. *Biochem. Syst. Ecol.* 16, 151–155. doi: 10.1016/0305-1978(88)90088-9

Buer, C. S., Imin, N., and Djordjevic, M. A. (2010). Flavonoids: new roles for old molecules. *J. Integr. Plant Biol.* 52, 98–111. doi: 10.1111/j.1744-7909.2010.00905.x

Cajka, T., and Fiehn, O. (2015). Toward merging untargeted and targeted methods in mass spectrometry-based metabolomics and lipidomics. *Anal. Chem.* 88, 524–545. doi: 10.1021/acs.analchem.5b04491

Cantero, G., Campanella, C., Mateos, S., and Cortés, F. (2006). Topoisomerase II inhibition and high yield of endoreduplication induced by the flavonoids luteolin and quercetin. *Mutagenesis* 21, 321–325. doi: 10.1093/mutage/gel033

Catford, J. G., Staehelin, C., Larose, G., Piche, Y., and Vierheilig, H. (2006). Systemically suppressed isoflavonoids and their stimulating effects on nodulation and mycorrhization in alfalfa split-root systems. *Plant Soil* 285, 257–266. doi: 10.1007/s11104-006-9072-8

Champion, A., Lucas, M., Tromas, A., Vaissayre, V., Crabos, A., Diédhiou, I., et al. (2015). Inhibition of auxin signaling in *Frankia* species-infected cells in *Casuarina glauca* nodules leads to increased nodulation. *Plant Physiol.* 167, 1149–1157. doi: 10.1104/pp.114.255307

Cuyckens, F., Rozenberg, R., de Hoffman, E., and Claeys, M. (2001). Structure characterization of flavonoid O-diglycosides by positive and negative nano-electrospray ionization ion trap mass spectrometry. *J. Mass Spectrom.* 36, 1203–1210. doi: 10.1002/jms.224

DeFelice, B. C., Mehta, S. S., Samra, S., Čajka, T., Wancewicz, B., Fahrmann, J. F., et al. (2017). mass spectral feature list optimizer (MS-FLO): a tool to minimize false positive peak reports in untargeted liquid chromatography-mass spectroscopy (LC-MS) data processing. *Anal. Chem.* 89, 3250–3255. doi: 10.1021/acs.analchem.6b04372

Dewick, P. M. (1982). "Isoflavonoids," in *The Flavonoids: Advances in Research*, eds J. B. Harborne and T. J. Mabry (Cambridge: Cambridge University Press), 535–640. doi: 10.1007/978-1-4899-2915-0_10

Dunn, W. B., Erban, A., Weber, R. J. M., Creek, D. J., Brown, M., Breitling, R., et al. (2012). Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics* 9, 44–66. doi: 10.1007/s11306-012-0434-4

Feunang, Y. D., Eisner, R., Knox, C., Chepelev, L., Hastings, J., Owen, G., et al. (2016). ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. *J. Cheminform.* 8, 1–20.

Gage, D. J. (2004). Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes. *Microbiol. Mol. Biol. Rev.* 68, 280–300. doi: 10.1128/MMBR.68.2.280-300.2004

Gallego-Giraldo, L., Bhattarai, K., Pislariu, C. I., Nakashima, J., Jikumaru, Y., Kamiya, Y., et al. (2014). Lignin modification leads to increased nodule numbers in alfalfa. *Plant Physiol.* 164, 1139–1150. doi: 10.1104/pp.113.232421

Gnanamanickam, S. S., and Smith, D. A. (1980). Selective toxicity of isoflavonoid phytoalexins to gram-positive bacteria. *Phytopathology* 70, 894–896. doi: 10.1094/Phyto-70-894

Grambow, H. K., and Grisebach, H. (1971). Further studies on the biosynthesis of flavonoids in *Datisca cannabina. Phytochemistry* 10, 789–796. doi: 10.1016/S0031-9422(00)97148-6

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/nprot.2013.084

Heijnen, C. G., Haene, G. R., van Acker, F. A., van der Vijgh, W. J., and Bast, A. (2001). Flavonoids as peroxynitrite scavengers: the role of the hydroxyl groups. *Toxicol. Vitro* 15, 3–6. doi: 10.1016/S0887-2333(00)00053-9

Hoagland, D. R., and Arnon, D. I. (1950). The water-culture method for growing plants without soil. *Calif. Agric. Exp. Stn. Circ.* 347.

Jones, P., Binns, D., Chang, H., Fraser, M., Li, W., AcAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457–D462. doi: 10.1093/nar/gkv1070

Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., et al. (2016). PubChem substance and compound databases. *Nucleic Acids Res.* 44, D1202–D1213. doi: 10.1093/nar/gkv951

Knollenberg, B. J., Liu, J., Yu, S., Lin, H., and Tian, L. (2018). Cloning and functional characterization of a p-coumaroyl quinate/shikimate 3'-hydroxylase from potato (Solanum tuberosum). *Biochem. Biophys. Res. Commun.* 496, 462–467. doi: 10.1016/j.bbrc.2018.01.075

Larrainzar, E., Riely, B., Kim, S. C., Carrasquilla-Garcia, N., Yu, H., Hwang, H., et al. (2015). Deep sequencing of the *Medicago truncatula* root transcriptome reveals a massive and early interaction between Nod factor and ethylene signals. *Plant Physiol.* 169, 233–265. doi: 10.1104/pp.15.00350

Long, S. R. (1996). Rhizobium symbiosis: nod factors in perspective. *Plant Cell* 8, 1885–1898. doi: 10.1105/tpc.8.10.1885

Ma, Y., Tanaka, N., Vaniya, A., Kind, T., and Fiehn, O. (2016). Ultrafast polyphenol metabolomics of red wines using MicroLC-MS/MS. *J. Agric. Food Chem.* 64, 505–512. doi: 10.1021/acs.jafc.5b04890

Markmann, K., and Parniske, M. (2009). Evolution of root endosymbiosis with bacteria: how novel are nodules? *Trends Plant Sci.* 14, 77–86. doi: 10.1016/j.tplants.2008.11.009

Martens, S., Preub, A., and Matern, U. (2010). Multifunctional flavonoid dioxygenases: flavonol and anthocyanin biosynthesis in *Arabidopsis thaliana* L. *Phytochemistry* 71, 1040–1049. doi: 10.1016/j.phytochem.2010.04.016

Mathesius, U., Jin, J., van Noorden, G. E., Ng, P. J., and Wasson, A. P. (2015). "Regulation of nodule development by short- and long-distance auxin transport control," in *Biological Nitrogen Fixation*, Vol. 1, ed. F. J. de Bruijn (Hoboken, NJ: John Wiley & Sons), 465–473.

Mathesius, U., Schlaman, H. R. M., Spaink, H. P., Sautter, C., Rolfe, B. G., and Djordjevic, M. A. (1998). Auxin transport inhibition precedes root nodule formation in white clover roots and is regulated by flavonoids and derivatives of chitin oligosaccharides. *Plant J.* 14, 23–34. doi: 10.1046/j.1365-313X.1998.00090.x

Melo, P. M., Silva, L. S., Ribeiro, I., Seabra, A. R., and Carvalho, H. G. (2011). Glutamine synthetase is a molecular target of nitric oxide in root nodules of *Medicago truncatula* and is regulated by tyrosine nitration. *Plant Physiol.* 157, 1505–1517. doi: 10.1104/pp.111.186056

Miyahisa, I., Funa, N., Ohnishi, Y., Martens, S., Moriguchi, T., and Horinouchi, S. (2006). Combinatorial biosynthesis of flavones and flavonols in *Escherichia coli. Appl. Microbiol. Biotechnol.* 71, 53–58. doi: 10.1007/s00253-005-0116-5

Modolo, L. V., Blount, J. W., Achnine, L., Naoumkina, M. A., Wang, X., and Dixon, R. A. (2007). A functional genomics approach to (iso)flavonoid glycosylation in the model legume *Medicago truncatula. Plant Mol. Biol.* 64, 499–518. doi: 10.1007/s11103-007-9167-6

Neumann, S., and Bocker, S. (2010). Computational mass spectrometry for metabolomics: identification of metabolites and small molecules. *Anal. Bioanal. Chem.* 398, 2779–2788. doi: 10.1007/s00216-010-4142-5

Nguyen, T. V., Wibberg, D., Battenberg, K., Blom, J., Vanden Heuvel, B., Berry, A. M., et al. (2016). An assemblage of *Frankia* cluster II strains from California contains the canonical nod genes and also the sulfotransferase gene nodH. *BMC Genomics* 17:796. doi: 10.1186/s12864-016-3140-1

Normand, P., Orso, S., Cournoyer, B., Jeannin, P., Chapelon, C., Dawson, J., et al. (1996). Molecular phylogeny of the genus Frankia and related genera and emendation of the family *Frankiaceae. Int. J. Syst. Evol. Microbiol.* 46, 1–9.

Oldroyd, G. E. D. (2013). Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nat. Rev. Microbiol.* 11, 252–263. doi: 10.1038/nrmicro2990

Ono, N., Qin, X., Wilson, A., Li, G., and Tian, L. (2016). Two UGT84 family glycosyltransferases catalyze a critical reaction of hydrolyzable tannin biosynthesis in pomegranate (*Punica granatum). PLoS One* 11:e0156319. doi: 10.1371/journal.pone.0156319

Peck, M. C., Fisher, R. F., and Long, S. R. (2006). Diverse flavonoids stimulate NodD1 binding to nod gene promoters in *Sinorhizobium meliloti. J. Bacteriol.* 188, 5417–5427. doi: 10.1128/JB.00376-06

Péret, B., Swarup, R., Jansen, L., Devos, G., Auguy, F., Collin, M., et al. (2007). Auxin influx activity is associated with *Frankia* infection during actinorhizal nodule formation in *Casuarina glauca. Plant Physiol.* 144, 1852–1862. doi: 10.1104/pp.107.101337

Persson, T., Battenberg, K., Demina, I. V., Vigil-Stenman, T., Vanden Heuvel, B., Pujic, P., et al. (2015). Candidatus Frankia datiscae Dg1, the actinobacterial microsymbiont of *Datisca glomerata*, expresses the canonical nod genes nodABC in symbiosis with its host plant. *PLoS One* 10:e0127630. doi: 10.1371/journal.pone.0127630

Phillips, D. A., and Tsai, S. M. (1992). Flavonoids as plant signals to rhizosphere microbes. *Mycorrhiza* 1, 55–58. doi: 10.1007/BF00206136

Rasul, A., Millimouno, F. M., Wltayb, W. A., Ali, M., Li, J., and Li, X. (2013). Pinocembrin: a novel natural compound with versatile pharmacological and biological activities. *Biomed. Res. Int.* 2013:379850. doi: 10.1155/2013/379850

Reid, D. E., Ferguson, B. J., Hayashi, S., Lin, Y., and Gresshoff, P. M. (2011). Molecular mechanisms controlling legume autoregulation of nodulation. *Ann. Bot.* 108, 789–795. doi: 10.1093/aob/mcr205

Ren, X., He, T., Chang, Y., Zhao, Y., Chen, X., Bai, S., et al. (2017). The genus Alnus, a comprehensive outline of its chemical constituents and biological activities. *Molecules* 22:e1383. doi: 10.3390/molecules22081383

Roux, B., Rodde, N., Jardinaud, M., Timmers, T., Sauviac, L., Cottret, L., et al. (2014). An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. *Plant J.* 77, 817–837. doi: 10.1111/tpj.12442

Schubert, M., Melnikova, A. N., Mesecke, N., Zubkova, E. K., Fortte, R., Batashev, D. R., et al. (2010). Two novel disaccharides, rutinose and methylrutinose, are involved in carbon metabolism in *Datisca glomerata. Planta* 231, 507–521. doi: 10.1007/s00425-009-1049-5

Schymanski, E. L., Jeon, J., Guide, R., Fenner, K., Ruff, M., Singer, H. P., et al. (2014). Identifying small molecules via high resolution mass spectrometry: communicating confidence. *Environ. Sci. Technol.* 48, 2097–2098. doi: 10.1021/es5002105

Shirley, B. W. (1996). Flavonoid biosynthesis: 'new' functions for an 'old' pathway. *Trends Plant Sci.* 1, 377–382.

Soltis, D. E., Soltis, P. S., Morgan, D. R., Swensen, S. M., Mullin, B. C., Dowd, J. M., et al. (1995). Chloroplast gene sequence data suggest a single orgin of the predisposition for symbiotic nitrogen fixation in angiosperms. *Proc. Natl. Acad. Sci. U.S.A.* 92, 2647–2651. doi: 10.1073/pnas.92.7.2647

Staszkow, A., Swarcewicz, B., Banasiak, J., Muth, D., Jasinski, M., and Stobiecki, M. (2011). LC/MS profiling of flavonoid glycoconjugates isolated from hairy roots, suspension root cell cultures and seedling roots of Medicago truncatula. *Metabolomics* 7, 604–613. doi: 10.1007/s11306-011-0287-2

Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., et al. (2007). Proposed minimum reporting standards for chemical analysis. *Metabolomics* 3, 211–221. doi: 10.1007/s11306-007-0082-2

Tiller, S. A., Parry, A. D., and Edwards, R. (1994). Changes in the accumulation of flavonoid and isoflavonoid conjugates associated with plant age and nodulation in alfalfa (*Medicago sativa*). *Physiol. Plant.* 91, 27–36. doi: 10.1111/j.1399-3054.1994.tb00655.x

Tsugawa, H., Cajka, T., Kind, T., Ma, Y., Higgins, B., Ikeda, K., et al. (2015). MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat. Methods* 12, 523–526. doi: 10.1038/nmeth.3393

Vinardell, J. M., Fedorova, E., Cebolla, A., Kevei, Z., Horvath, G., Kelemen, Z., et al. (2003). Endoreduplication mediated by the anaphase-promoting complex activator CCS523A is required for symbiotic cell differentiation in *Medicago truncatula* nodules. *Plant Cell* 15, 2093–2105. doi: 10.1105/tpc.014373

Wasson, A. P., Pellerone, F. I., and Mathesius, U. (2006). Silencing the flavonoid pathway in *Medicago truncatula* inhibits root nodule formation and prevents auxin transport regulation by rhizobia. *Plant Cell* 18, 1617–1629. doi: 10.1105/tpc.105.038232

Wollenweber, E., Dorr, M., Bohm, B. A., and Roitman, J. N. (2004). Exudate flavonoids of eight species of *Ceanothus* (Rhamnaceae). *Z. Naturforsch.* 59, 459–462. doi: 10.1515/znc-2004-7-801

Wollenweber, E., Kohorst, G., Mann, K., and Bell, J. M. (1985). Leaf gland flavonoids in *Comptonia peregrina* and *Myrica pensylvanic* (Myricaceae). *J. Plant Physiol.* 117, 423. doi: 10.1016/S0176-1617(85)80049-3

Actinorhizal Signaling Molecules: *Frankia* Root Hair Deforming Factor Shares Properties With NIN Inducing Factor

Maimouna Cissoko[1,2,3,4], Valérie Hocher[4], Hassen Gherbi[4], Djamel Gully[4], Alyssa Carré-Mlouka[4,5], Seyni Sane[6], Sarah Pignoly[1,2,4], Antony Champion[1,2,7], Mariama Ngom[1,2], Petar Pujic[8], Pascale Fournier[8], Maher Gtari[9], Erik Swanson[10], Céline Pesce[10], Louis S. Tisa[10], Mame Oureye Sy[3] and Sergio Svistoonoff[1,2,4*]

[1] Laboratoire Commun de Microbiologie, Institut de Recherche pour le Développement/Institut Sénégalais de Recherches Agricoles/Université Cheikh Anta Diop, Centre de Recherche de Bel Air, Dakar, Senegal, [2] Laboratoire Mixte International Adaptation des Plantes et Microorganismes Associés Aux Stress Environnementaux, Centre de Recherche de Bel Air, Dakar, Senegal, [3] Laboratoire Campus de Biotechnologies Végétales, Département de Biologie Végétale, Faculté des Sciences et Techniques, Université Cheikh Anta Diop, Dakar, Senegal, [4] Laboratoire des Symbioses Tropicales et Méditerranéennes, Institut de Recherche pour le Développement/INRA/CIRAD, Université Montpellier/SupAgro, Montpellier, France, [5] UMR 7245, Molécules de Communication et Adaptation des Microorganismes, Muséum National d'Histoire Naturelle, Centre National de la Recherche Scientifique, Sorbonne Universités, Paris, France, [6] Laboratoire de Botanique et de Biodiversité Végétale, Département de Biologie Végétale, Faculté des Sciences et Techniques, Université Cheikh Anta Diop, Dakar, Senegal, [7] UMR Diversité Adaptation et Développement des Plantes (DIADE), Institut de Recherche pour le Développement, Montpellier, France, [8] Ecologie Microbienne, UMR 5557 CNRS, Université Lyon 1, Villeurbanne, France, [9] Institut National des Sciences Appliquées et de Technologie, Université Carthage, Tunis, Tunisia, [10] Department of Molecular, Cellular, and Biomedical Sciences, University of New Hampshire, Durham, NH, United States

Actinorhizal plants are able to establish a symbiotic relationship with *Frankia* bacteria leading to the formation of root nodules. The symbiotic interaction starts with the exchange of symbiotic signals in the soil between the plant and the bacteria. This molecular dialog involves signaling molecules that are responsible for the specific recognition of the plant host and its endosymbiont. Here we studied two factors potentially involved in signaling between *Frankia casuarinae* and its actinorhizal host *Casuarina glauca:* (1) the Root Hair Deforming Factor (CgRHDF) detected using a test based on the characteristic deformation of *C. glauca* root hairs inoculated with *F. casuarinae* and (2) a NIN activating factor (CgNINA) which is able to activate the expression of *CgNIN*, a symbiotic gene expressed during preinfection stages of root hair development. We showed that CgRHDF and CgNINA corresponded to small thermoresistant molecules. Both factors were also hydrophilic and resistant to a chitinase digestion indicating structural differences from rhizobial Nod factors (NFs) or mycorrhizal Myc-LCOs. We also investigated the presence of CgNINA and CgRHDF in 16 *Frankia* strains representative of *Frankia* diversity. High levels of root hair deformation (RHD) and activation of Pro*CgNIN* were detected for *Casuarina*-infective strains from clade Ic and closely related strains from clade Ia unable to nodulate *C. glauca*. Lower levels were present for distantly related strains belonging to clade III. No CgRHDF or CgNINA could be detected for *Frankia coriariae* (Clade II) or for uninfective strains from clade IV.

Keywords: symbioses, nodulation factors, nodule inception, *Casuarina*, *Alnus*, *Discaria*

INTRODUCTION

Legumes and actinorhizal plants form a N_2-fixing root nodule symbiosis in association with rhizobia and *Frankia* bacteria, respectively (Vessey et al., 2005). The establishment of these beneficial bacterial-plant relationships requires communication between the partners. Rhizobial symbioses have received considerable attention because several legumes are important crop species. However, actinorhizal symbioses, which play an important ecological role (Dawson, 2008), have been less well studied and the molecular dialog between *Frankia* and their host plants is still poorly understood. One reason is that most actinorhizal plants are woody shrubs or trees for which genetic approaches are very difficult (Wall, 2000; Perrine-Walker et al., 2011). In addition, the genetics of the bacterial partner, *Frankia*, is not fully developed and up to now *Frankia* cells remain recalcitrant to stable genetic transformation (Kucho et al., 2009, 2017). Recent progress including the sequencing of several *Frankia* genomes (Normand et al., 2007; Tisa et al., 2016), transcriptomic studies (Alloisio et al., 2010; Benson et al., 2011), proteomic studies (Mastronunzio and Benson, 2010; Ktari et al., 2017) together with functional studies on several actinorhizal species (Svistoonoff et al., 2014) have opened new avenues for identifying components involved in the initial symbiotic dialog between the two partners.

The interaction of rhizobia with model legumes begins with the production and recognition of signal molecules by their respective eukaryotic and prokaryotic symbiotic partners (Oldroyd, 2013). Early events leading to nodule formation involve bacterial penetration into their hosts *via* root hairs. Bacteria elicit the stimulation and reorientation of root hair cell wall growth. This rhizobia-induced tip growth results first in the entrapment of the bacteria within curled root hairs and then in the initiation and development of infection threads (ITs), tubular structures through which bacteria pass on their way down the root hair and into the underlying cortical cell layers (Lhuissier et al., 2001). Ahead of the advancing threads, cells in the inner cortex are induced to dedifferentiate and divide, and a nodule primordium is formed. In the first part of the signal exchange, the plant roots secrete flavonoids that lead to the activation of a set of rhizobial genes (the *nod* genes), which are essential for infection, nodule development and the control of host specificity (Masson-Boivin et al., 2009; Oldroyd, 2013). These genes are responsible for the synthesis of lipo-chito-oligosaccharides (LCOs) called Nod factors (NFs) that signal back to the plant (Oldroyd, 2013). NF biosynthesis is dependent on *nodABC* genes which are present in all rhizobia able to synthetize NFs and strain-specific combinations of other nodulation genes responsible for the addition of various decorations to the core structure. (Masson-Boivin et al., 2009). In model legumes, NFs perception elicits a range of responses including ion fluxes, calcium oscillations, changes in gene expression patterns, and extensive deformation of roots hairs, which has been used as a bioassay to identify the chemical nature of NFs (Lerouge et al., 1990; Oldroyd, 2013).

Much less is known about signaling molecules involved in the actinorhizal symbioses. Canonical *nodABC* genes are not found in the sequenced genomes of 36 *Frankia* strains including *Frankia alni* and *Frankia casuarinae* (Tisa et al., 2016) confirming a previous report showing that *F. alni* DNA will not complement rhizobial *nod* mutants (Cérémonie et al., 1998). Only distant homologs of *nodB* and *nodC* are found in *F. alni* genome. Unlike rhizobial *nod* genes, they are not organized into a cluster together with other symbiotic genes and their expression is not induced under symbiotic conditions (Normand et al., 2007; Alloisio et al., 2010). These findings are consistent with experiments showing that chitin oligomers similar to rhizobial NFs are not be detected in *F. alni* culture supernatant (Cérémonie et al., 1999) suggesting structural differences between the *Frankia* symbiotic signals and rhizobial NFs. Recently, canonical *nodABC* genes have been found in the genome of two uncultured *Frankia* strains: Candidatus *Frankia datiscae* Dg1 and Candidatus *Frankia californicae* Dg2 (Persson et al., 2015; Nguyen et al., 2016), and in one isolated strain, *Frankia* sp. NRRL B-16219 (Ktari et al., 2017). *F. datiscae* Dg1 *nodABC* genes are arranged in two operons which are expressed in *Datisca glomerata* nodules, but their involvement in symbiotic signaling is still not known (Persson et al., 2015).

Frankia is able to infect their host root either through intracellular (root hair) or intercellular modes. In the first case, one of the earliest visible plant response to *Frankia* is an extensive deformation of root hairs. This response occurs in actinorhizal plants belonging to the order Fagales (*Betulaceae, Casuarinaceae*) that display a range of relatively advanced features reminiscent of model legumes: a complex root hair infection process involving the formation of ITs and the implication of cortical cell divisions at the initial stages of infection (Svistoonoff et al., 2014). *Frankia* culture supernatants also cause root hair deformation (RHD) and a *Frankia* root hair deforming factor in *Alnus* (AgRHDF) was identified (Prin and Rougier, 1987; Ghelue et al., 1997; Cérémonie et al., 1999; Gabbarini and Wall, 2011). Using RHD as a bioassay, partial purification was achieved. AgRHDF is a relatively small (< 3 kDa), heat stable, hydrophilic molecule that is resistant to a chitinase treatment, but its chemical structure remains unknown (Cérémonie et al., 1999).

In recent years, we have developed complementary bioassays using plant genes that are specifically expressed in response to interaction with a compatible *Frankia*. This approach is particularly well suited for *C. glauca* where transgenic plants containing promoters of symbiotic genes fused to either *GUS* or *GFP* can be generated (Svistoonoff et al., 2010a). Expressed Sequence Tag (EST) libraries of *C. glauca* and *Alnus glutinosa* (Hocher et al., 2006, 2011) provide extensive lists of genes potentially involved in the actinorhizal symbiosis. Among the candidate genes, we identified *CgNIN,* the putative ortholog of legume *NIN* genes, which encodes a transcription factor playing a central role in rhizobial nodulation (Schauser et al., 1999; Marsh et al., 2007; Soyano et al., 2013, 2014; Yoro et al., 2014). In *C. glauca, CgNIN* also has an important role in nodulation particularly at early steps of infection (Clavijo et al., 2015). After contact with either *Frankia* cells or cell-free *Frankia* supernatants, the *CgNIN* promoter is strongly

activated at 12 to 48 h (Clavijo et al., 2015). This property was used to establish a new bioassay leading to the partial purification and characterization of a NIN activating factor, called CgNINA. While rhizobial NFs are amphiphilic chitin-based molecules, CgNINA, like AgRHDF, is hydrophilic and resistant to chitinase (Chabaud et al., 2016). However, it is not known to what extent CgNINA is related to factors able to deform root hairs.

Further experiments concerning these *Frankia* symbiotic factors are reported here. We show that *C. glauca* was able to perceive a root hair deforming factor secreted by *F. casuarinae* (CgRHDF) and the properties of CgRHDF were compared to those previously identified with CgNINA. The presence of CgNINA and CgRHDF in strains representative of *Frankia* diversity was investigated.

MATERIALS AND METHODS

Plant Material and Growth Conditions

Casuarina glauca seeds (seed lot 15.934, ref.086-5929) were provided by the Australian Tree Seed Centre[1]. *Ochetophila trinervis* (= *Discaria trinervis*) seeds were collected from plants growing in Pampa de Huenuleo (Bariloche, Argentina). *A. glutinosa* seeds were harvested from a tree situated in the left bank of Rhône River in Lyon, France. *C. glauca* and *O. trinervis* seeds were disinfected and germinated in a sterilized substrate for three weeks and transferred into glass tubes filled with a modified Broughton and Dilworth (BD) medium as described previously (Ngom et al., 2016). *A. glutinosa* seeds were washed in distilled sterile water for 30 min before sterilization in 96% ethanol for 30 min followed by 3% solution of calcium hypochlorite for 30 min. Seeds were germinated on 1.5% plant agar for 10 days at 20°C, and transferred in 5 mL tubes containing liquid Fahraeus medium without nitrogen (Fahraeus, 1957). Plants were grown for 6 weeks in growth chamber at 25° C at 75% relative air humidity and 16 h light cycle/day. Transgenic *C. glauca* plants containing a Pro*CgNIN:GFP* construct described previously (Clavijo et al., 2015) were grown in hydroponics in pots containing the modified BD medium and vegetatively propagated as described previously (Svistoonoff et al., 2010b).

Preparation of Cell-Free Supernatants and Inoculation

The bacterial strains used in this study are listed in **Supplementary Table S1** and were grown for twenty-one days in modified basic propionate (BAP) media described previously (Ngom et al., 2016) according to conditions listed in **Supplementary Table S1**. Bacterial cultures were exposed to plant root exudates (RE) for five days as described previously (Beauchemin et al., 2012; Clavijo et al., 2015; Chabaud et al., 2016). Cell-free supernatant fluids were purified from cultures showing an absorbance of 0.3 at 595 nm. Cultures were collected by centrifugation at 4,000 g for 5 min and the supernatant

fluids were filtered through a 0.22 μm filter as described in Chabaud et al., 2016. Unless otherwise indicated, experiments were performed with the supernatant fluids of a *F. casuarinae* culture induced with RE from its host plant *C. glauca* and referred to as FCS for *Frankia casuarinae supernatant*. FCS were concentrated fifty times (FCS 50X) using an R-210/215 evaporator (BÜCHI Labortechnik AG, Switzerland). Nodulation experiments with the different *Frankia* strains were performed as described previously (Alloisio et al., 2010; Svistoonoff et al., 2010b; Imanishi et al., 2011).

Characterization of *F. casuarinae* Supernatant Fluids

Temperature and pH Sensitivity

To test heat inactivation, FCS 50X was autoclaved at 120°C for 20 min. Cold sensitivity was determined by freezing FCS at -80°C for 1 h. The effects of pH on FCS activity were determined as follows: the initial pH of FCS (6.7) was adjusted to pH 3, 5, 7, 8 or 10 by adding either HCl or KOH solutions. The FCS mixtures that these pH values were incubated for 1h at room temperature and neutralized back to pH 6.7 by adding either HCl or KOH solutions before performing the bioassays described below. Samples that lost CgNINA or CgRHDF activities were sonicated for 30 min using a Branson 2510 sonicator.

Size Fractionation

To estimate the size of signaling molecules, the FCS samples were dialyzed as described in Chabaud et al., 2016. Ten mL of FCS 50X were dialyzed for 12h at 4°C against 5 L of ultrapure water with stirring and using either a 100–500, 500–1000 or 3500–5000 Da cutoff membrane (Float-A-Lyzer G2 dialysis devices, Spectrum Laboratories, CA, United States). The dialyzed solutions were tested using the CgRHD and CgNINA bioassays described below. The size of active compounds was also estimated using centrifugal filters with cutoffs of 30, 10, and 3 kDa (Amicon Ultra-4 centrifugal filters; Merck-Millipore, Cork, Ireland). Four mL of 50X FCS were loaded on a 30 kDa cell which was spun at 4,000 g for 20 min. The filtrate recovered from the 30 kDa filtration was treated similarly using a 10 kDa cell, and the resulting 10 kDa filtrate was added to a 3 kDa Cell and spun.

Phase Extraction and Sensitivity to Chitinase

Two sequential butanol extractions were performed on FCS 50X with a ratio 1-butanol / water (1:3; v/v) described previously (Chabaud et al., 2016). The butanol phase was evaporated at 80°C under a nitrogen flow and the residue was dissolved in 20% acetonitrile as described in (Chabaud et al., 2016). Chitinase digestions were performed on the aqueous phase extract as described previously (Chabaud et al., 2016). Chitinase activity was assessed using a colorimetric method to estimate the amount of p-nitrophenol (p-NP) released from a reaction mixture containing the substrate p-nitrophenyl N-acetyl glucosaminide (p-NP-NAG) (Parham and Deng, 2000). A solution containing 1 mg mL^{-1} of *Streptomyces griseus* chitinase (C6137; Sigma-Aldrich) in a 50 mM phosphate buffer pH 6.0 was prepared. A portion of this solution (100 μl) was mixed with 50 μl of 5X

FCS aqueous extract, 100 µl of p-NP-NAG solution at 10 mM and 250 µl of acetate buffer (pH 5.5 0,1 M). The reaction was incubated at 37°C for 1 h under stirring and was terminated by the addition of 250 µl of $CaCl_2$ at 0.5 M and 1000 µl of NaOH at 0.5 M. The amount of p-NP released was evaluated by measuring the absorbance at 400 nm with a spectrophotometer. Enzyme activity was expressed in µg of liberated p-NP per hour of incubation. Control reactions were performed without p-NP-NAG, without chitinase and without the aqueous FCS extract.

Root Hair Deformation and NINA Bioassays in *Casuarina glauca*

Unless otherwise indicated, CgRHD and CgNINA bioassays were performed on aliquots of the same solution and the amount needed to achieve a final concentration equivalent a 10^{-2} dilution of raw (no diluted) *Frankia* culture supernatant was added to the nitrogen-free BD medium. All experiments were performed on at least 4 plants. At least two independent experiments were carried out for each tested solution.

The deformation of *C. glauca* root hairs (CgRHD) was evaluated using 3 week-old non transgenic plants grown in glass tubes exposed to nitrogen starvation for one week as described previously (Ngom et al., 2016). Treatments were performed by replacing the medium with fresh nitrogen-free BD medium containing the assayed solution. Deformation of root hairs situated on small lateral roots was scored as described by Clavijo et al. (2015) using micrographs taken with a BX50F microscope (Olympus) equipped with a Micro Publisher 3.3 RTV (Qimaging) digital camera. A blind evaluation of each micrograph was performed to determine the deformation level of observed root hairs using the following scale based on Clavijo et al. (2015): 0a: no deformation; 0b: straight root hair with tip swelling; 1 weak deformation, only one change in growth direction; 2: intermediate deformation, more than one change in growth direction but no bifurcation; and 3: strong deformation: one or more bifurcations (**Figure 1**). Deformation levels 0a and 0b were considered non-symbiotic. For each experiment at least four plants were analyzed per treatment and 6 small lateral roots were analyzed per plant. The total number of root hairs scored for each level of symbiotic deformation was used for the statistical analyzes described below. Each experiment was repeated four times independently.

For each treatment the percentage of symbiotic deformation (%SyD) defined as the proportion of root hairs showing a symbiotic response was calculated and used to determine a deformation index using the following scale: level 1: SyD < 15%; level 2: 16% < SyD < 25%; level 3: 26% < SyD < 40%; level 4: SyD > 41%.

The activation of *ProCgNIN* in response to tested solutions was evaluated using transgenic Pro*CgNIN:GFP* plants that were grown in hydroponics deprived of nitrogen for one week as previously described (Clavijo et al., 2015; Chabaud et al., 2016). After 24 h of contact with tested solutions, GFP fluorescence was monitored in the short lateral root hairs using an AZ100 macroscope (Nikon) equipped with a 5X objective, a GFP filter (Excitation filter 470 nm ± 40 nm; Barrier filter 535 nm ± 50 nm;

Nikon) and a digital camera Sight D5 RI1 (Nikon). For each observation, a blind evaluation of GFP fluorescence levels was performed using the following scale: 0: no detectable fluorescence; 1: weak fluorescence; 2: intermediate fluorescence; and 3: strong fluorescence (**Figure 1**). For each experiment the number of plants with a given fluorescence level was used for the statistical analysis described below. Each experiment was repeated at least four times independently.

Root Hair Deformation Bioassay in *A. glutinosa*

The deformation of *A. glutinosa* root hairs (AgRHD) was evaluated using 7 week-old plants that had at least four well developed secondary roots. Biological tests were performed at 10^{-2} final dilutions of *Frankia* culture supernatant fluids on plants growing in 5 ml Fahraeus media without nitrogen. Evaluation of deformation was done in the region located about 1.5 cm from the root tip and five levels of RHD were recorded for each observed root: 0a: no deformation; 0b: swelling; 1: branching; 2: branching and partial deformation; 3: total RHD and retracting. Deformation levels 0a and 0b were considered as non-symbiotic. All experiments were performed on at least 3 plants and 5 roots were observed for each plant.

Statistical Analyses

Statistical analyses were performed on raw data: the number of hairs counted in each level of deformation using the R software package (R Core Team, 2013). A Shapiro-Wilk normality test was performed followed by a non-parametric Kruskal-Wallis multiple comparison test and a pairwise Wilcoxon test. These tests were used to compare the symbiotic response obtained for each treatment.
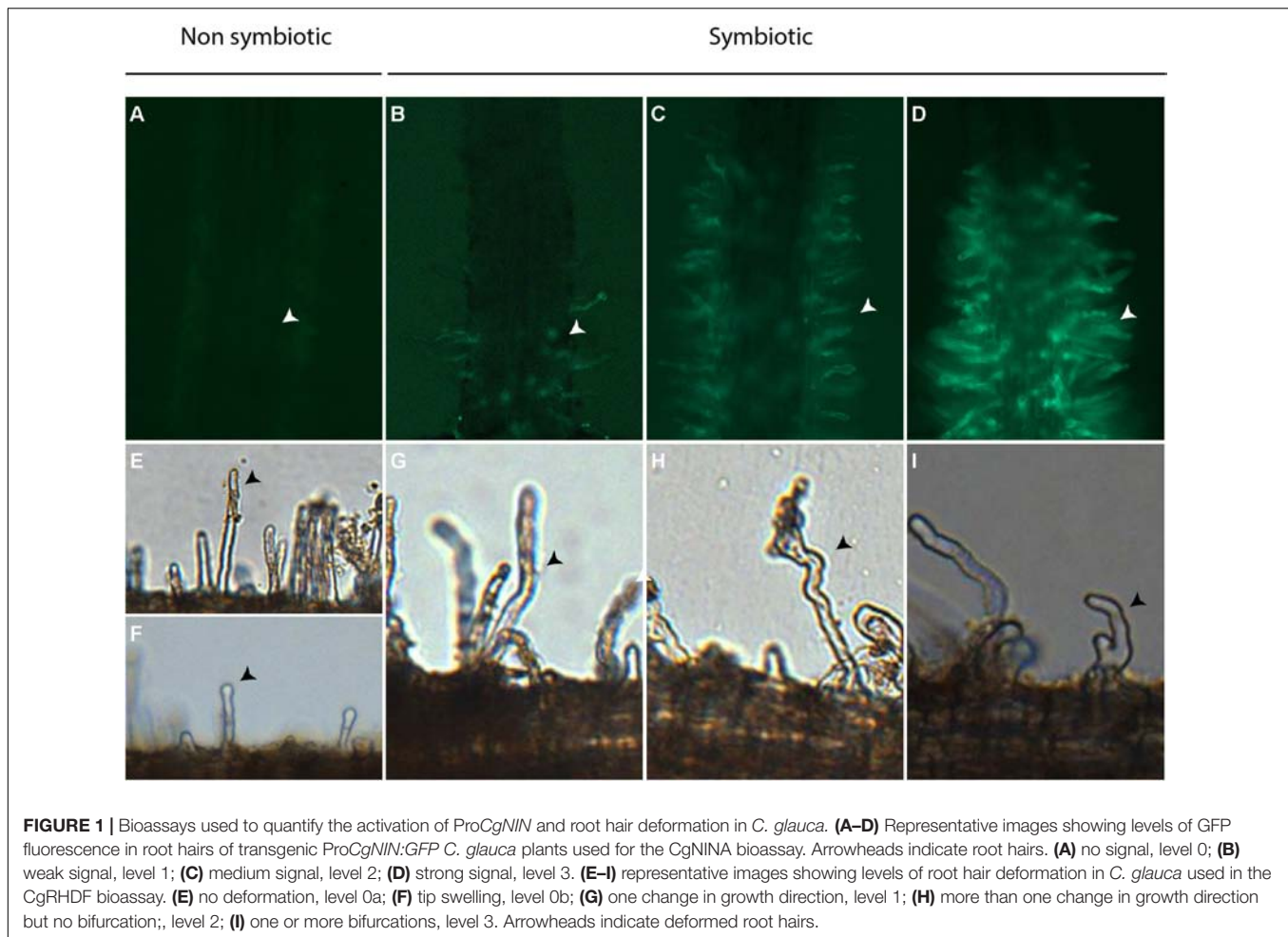
Phylogenetic Tree

The strict core genome of 17 *Frankia* strains was determined with the Get_Homologs package (Contreras-Moreira and Vinuesa, 2013). Out of 150,000 amino acid sequences in the *Frankia* pan genome, 420 proteins were identified as orthologs and part of the strict *Frankia* core genome. A concatenated phylogenetic tree was constructed. These concatenations were aligned using Clustal W (Larkin et al., 2007). The distance matrix was computed by Jukes-Cantor method (Jukes and Cantor, 1969). The Neighboring-joining method (Saitou and Nei, 1987) was used to build the phylogeny. The percentage of replicate trees in which the associated taxa clustered together was determined using a bootstrap test (1000 replicates) (Felsenstein, 1985). *Streptomyces coelicolor* was used as an outgroup.

RESULTS

CgRHD and CgNINA Activities Are Present in *F. casuarinae* Supernatant Fluid

Previous studies have shown that factors inducing RHD in *A. glutinosa* (AgRHDF) are present in several *Alnus*-infective
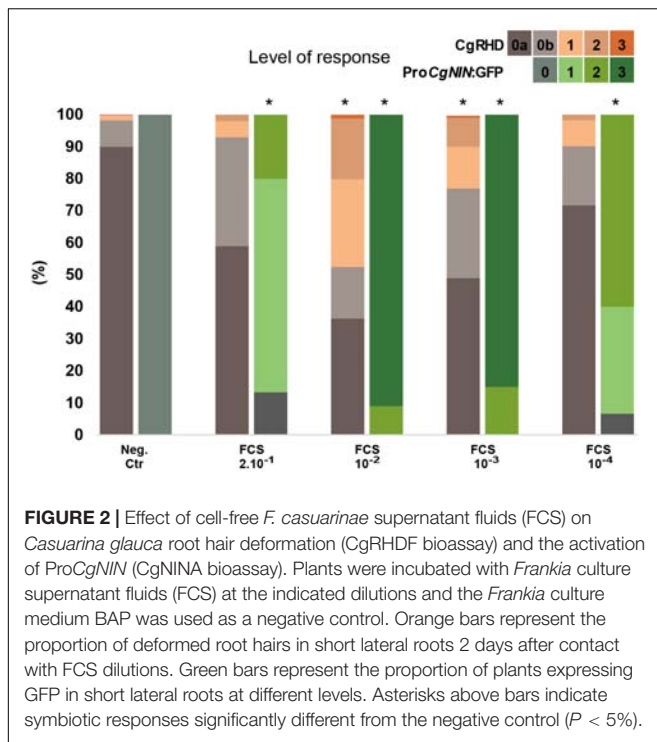
FIGURE 1 | Bioassays used to quantify the activation of Pro*CgNIN* and root hair deformation in *C. glauca*. **(A–D)** Representative images showing levels of GFP fluorescence in root hairs of transgenic Pro*CgNIN:GFP C. glauca* plants used for the CgNINA bioassay. Arrowheads indicate root hairs. **(A)** no signal, level 0; **(B)** weak signal, level 1; **(C)** medium signal, level 2; **(D)** strong signal, level 3. **(E–I)** representative images showing levels of root hair deformation in *C. glauca* used in the CgRHDF bioassay. **(E)** no deformation, level 0a; **(F)** tip swelling, level 0b; **(G)** one change in growth direction, level 1; **(H)** more than one change in growth direction but no bifurcation;, level 2; **(I)** one or more bifurcations, level 3. Arrowheads indicate deformed root hairs.

Frankia strains (Prin and Rougier, 1987; Ghelue et al., 1997; Cérémonie et al., 1999). We recently found that *F. casuarinae* produces an extracellular factor, named CgNINA, which is able to induce the expression of the early symbiotic gene *CgNIN* in small lateral roots (Clavijo et al., 2015; Chabaud et al., 2016). To investigate whether a root hair deforming factor, hereafter named CgRHDF, was also produced by *F. casuarinae,* we incubated wild type *C. glauca* plants with $2 \cdot 10^{-1}$, 10^{-2} 10^{-3}, and 10^{-4}-fold dilutions of *F. casuarinae* supernatant fluids (FCS) and scored the deformation of root hairs situated in small lateral roots. In parallel, transgenic plants containing the Pro*CgNIN:GFP* construct were incubated with the same solutions and the activation of Pro*CgNIN* was recorded. As shown in **Figure 2** and **Supplementary Table S2**, no RHD and no GFP fluorescence were detected in negative control roots treated with diluted BAP medium. RHD and GFP expression were maximal for the 10^{-2} dilution and lower levels of deformation and GFP fluorescence was observed with higher dilutions. The more concentrated dilution had a decreased response for both RHD and GFP fluorescence suggesting that the receptor could be saturated or inhibitory compounds may be associated with the extracts. At the dilutions 10^{-1} and 10^{-4}, only the CgNINA bioassay showed a significant difference with the negative control suggesting that

the CgNINA bioassay is more sensitive that the one based on CgRHD. The last three dilutions (10^{-2} to 10^{-4}) appear to show dose-dependent responses for the CgNINA bioassay that may be used to quantify this factor.

CgRHDF and CgNINA Share Physio-Chemical Properties

CgRHDF and CgNINA properties were further compared by using similar treatments to those used to characterize CgNINA (Chabaud et al., 2016). First, the effects of temperature and pH sensitivity were analyzed. As shown in **Figures 3A,B** and **Supplementary Table S2**, CgRHD was not affected by elevated temperatures (autoclaving) or treatment at pH values ranging from 5 to 10. However, cold treatment (freezing) or acidic pH conditions severely decreased CgRHD levels. Sonication of the inactivated fractions (frozen or acid-treated) resulted in a partial recovery of CgRHD activity (**Figures 3A,B** and **Supplementary Table S2**). Similar results were obtained with the CgNINA bioassay. These results suggesting that both factors are thermoresistant and possibly precipitate at low pH or upon freezing but this aggregate can be resuspended using sonication. Both dialysis membranes and centrifugal filters were used to

FIGURE 2 | Effect of cell-free *F. casuarinae* supernatant fluids (FCS) on *Casuarina glauca* root hair deformation (CgRHDF bioassay) and the activation of Pro*CgNIN* (CgNINA bioassay). Plants were incubated with *Frankia* culture supernatant fluids (FCS) at the indicated dilutions and the *Frankia* culture medium BAP was used as a negative control. Orange bars represent the proportion of deformed root hairs in short lateral roots 2 days after contact with FCS dilutions. Green bars represent the proportion of plants expressing GFP in short lateral roots at different levels. Asterisks above bars indicate symbiotic responses significantly different from the negative control ($P < 5\%$).

determine the approximate size of CgRHDF and CgNINA. As shown in **Figure 3C** and **Supplementary Table S2**, CgRHDF and CgNINA were detected inside the 100–500 Da and the 500-1000 Da cut-off dialysis tubing, but only a residual activity was found in the 3.5–5 kDa) suggesting that both factors correspond to small molecules with a molecular mass between 1 and 3.5 kDa. Experiments performed using centrifugal filters with 30, 10, and 3 kDa cut-offs yielded similar results for the CgNINA bioassay (**Figure 3D** and **Supplementary Table S2**). However, maximum activity using the CgRHD bioassay was detected in the 3 kDa and 10 kDa retentates and only residual CgRHD was detected in the 3kDa flow though suggesting that both factors correspond to small but distinct molecules. Taken together these experiments suggest that the size of CgNINA is between 1 and 3 kDa while CgRHDF is between 3 and 3.5 kDa. The polarity of the two factors was investigated using a butanol extraction. CgRHDF and CgNINA were only detected in the aqueous phase, indicating that both factors are hydrophilic (**Figure 3E** and **Supplementary Table S2**). Finally, we investigated whether CgRHDF contains a chitin backbone by performing the chitinase digestion experiment described in Chabaud et al. (2016). As shown in **Figure 3F** and **Supplementary Table S2**, the incubation with chitinase had no significant effect on CgRHDF or CgNINA activities. To rule out any inhibitory effect by FCS, we quantified the chitinase activity in the FCS/chitinase solution. As shown in **Supplementary Figure S1**, chitinase activity was not decreased by the addition of FCS to the chitinase solution. We conclude that *F. casuarinae* secretes two factors, CgNINA and CgRHDF, which are possibly two distinct molecules. Both factors sharing similar biochemical properties and correspond to small hydrophilic and thermoresistant molecules lacking a chitin backbone.

Presence of RHDF and CgNINA in *Frankia* Strains Representative of *Frankia* Diversity

We were interested in determining whether other *Frankia* strains had CgNINA and RHDF activities and tested different strains representing *Frankia* diversity. Cell-free supernatant fluids corresponding to 16 *Frankia* strains and another Actinobacteria, *S. coelicolor,* were tested for their capacity to deform *C. glauca* root hairs. As shown in **Figure 4**, **Supplementary Table S3**, and **Supplementary Figure S2**, CgRHD was present in all the *Frankia* belonging to clades I and III. The strongest activities were detected in strains that nodulate the genus *Casuarina* (clade Ic). Remarkably, RHD activity was also detected for several *Frankia* strains unable to form nodules on *Casuarina* such as *F. alni* (Clade Ia, *Alnus*-infective), *Frankia eleagni* and EANIpec (both clade III, *Elaeagnus*-infective strains). No deformation was detected for *Frankia coriariae* (clade II). Supernatant fluids from the atypical strains from clade IV or from the non-*Frankia* actinobacterium *S. coelicolor* did not induce RHD on *Casuarina*. Similar results were obtained when the same supernatant fluids were tested with the NIN bioassay (**Figure 4**, **Supplementary Table S3**, and **Supplementary Figure S2**) except for *F. discariae.*
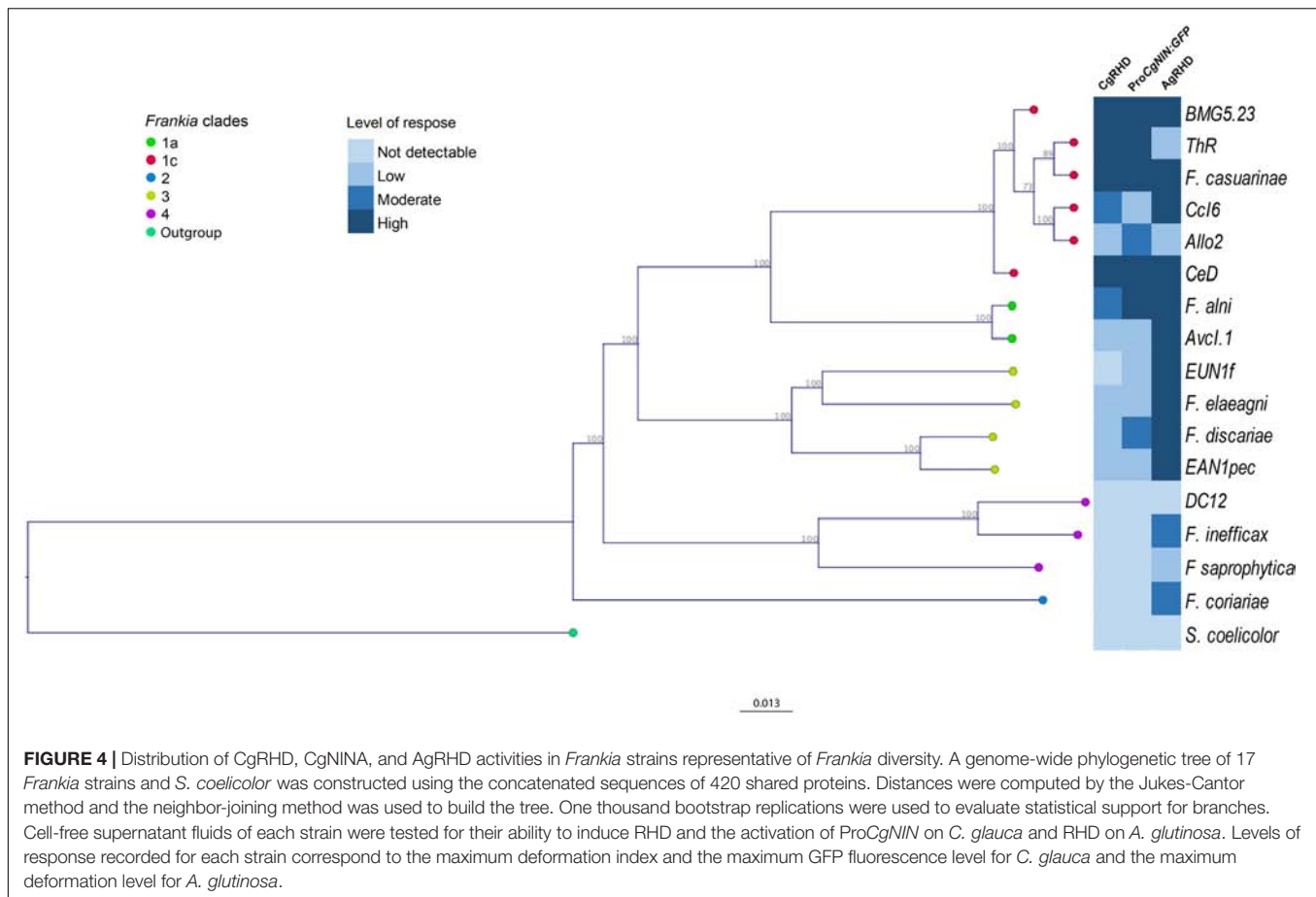
The strong RHD and NINA activities obtained with the *Alnus*-strain *F. alni* prompted us to investigate whether the same culture supernatant fluids could induce RHD on *A. glutinosa*. As shown in **Figure 4** and **Supplementary Figure S2**, responses observed in *A. glutinosa* were generally similar or stronger compared to the ones recorded in *C. glauca* particularly for strains belonging to group III. Interestingly, strains that were not able to activate Pro*CgNIN* or to deform root hairs of *C. glauca* such as *F. coriariae* and the two atypical strains from group IV, EuI1c and CN3 were able to deform *A. glutinosa* root hairs. To confirm the absence of unintentional contaminations with a compatible strain, *C. glauca, A. glutinosa*, and *O. trinervis* were inoculated with the bacterial pellet obtained while preparing the supernatant fluids described above. As shown in **Supplementary Table S3**, nodules were only obtained with compatible strains, thus demonstrating that responses recorded with incompatible strains are not the result of any contamination; incompatible strains are thus probably able to synthetize CgRHDF and CgNINA.

Surprisingly for *F. discariae,* we detected not only weak levels of RHD in *C. glauca* but also a weak activation of Pro*CgNIN* and a strong deformation of root hairs in *A. glutinosa* (**Figure 4** and **Supplementary Figure S2**). These results are in apparent contradiction with our previous work, in which we could not detect any activation of Pro*CgNIN* in response to *F. discariae* supernatants. However, at that time *F. discariae* cultures were induced with RE from *O. trinervis* instead of *C. glauca* (Chabaud et al., 2016). We therefore repeated the experiments for *F. alni* and *F. discariae* using RE from *A. glutinosa* and *O. trinervis*, respectively. As shown in **Figure 4**, **Supplementary Table S3**, and **Supplementary Figure S2**, incubation of *F. alni* with RE from the host plant (*A. glutinosa*) did not change the response of *C. glauca,* but slightly increased the response of *A. glutinosa*. For *F. discariae*, incubation with exudates from *O. trinervis* reduced the level of Pro*CgNIN* activation compared to the
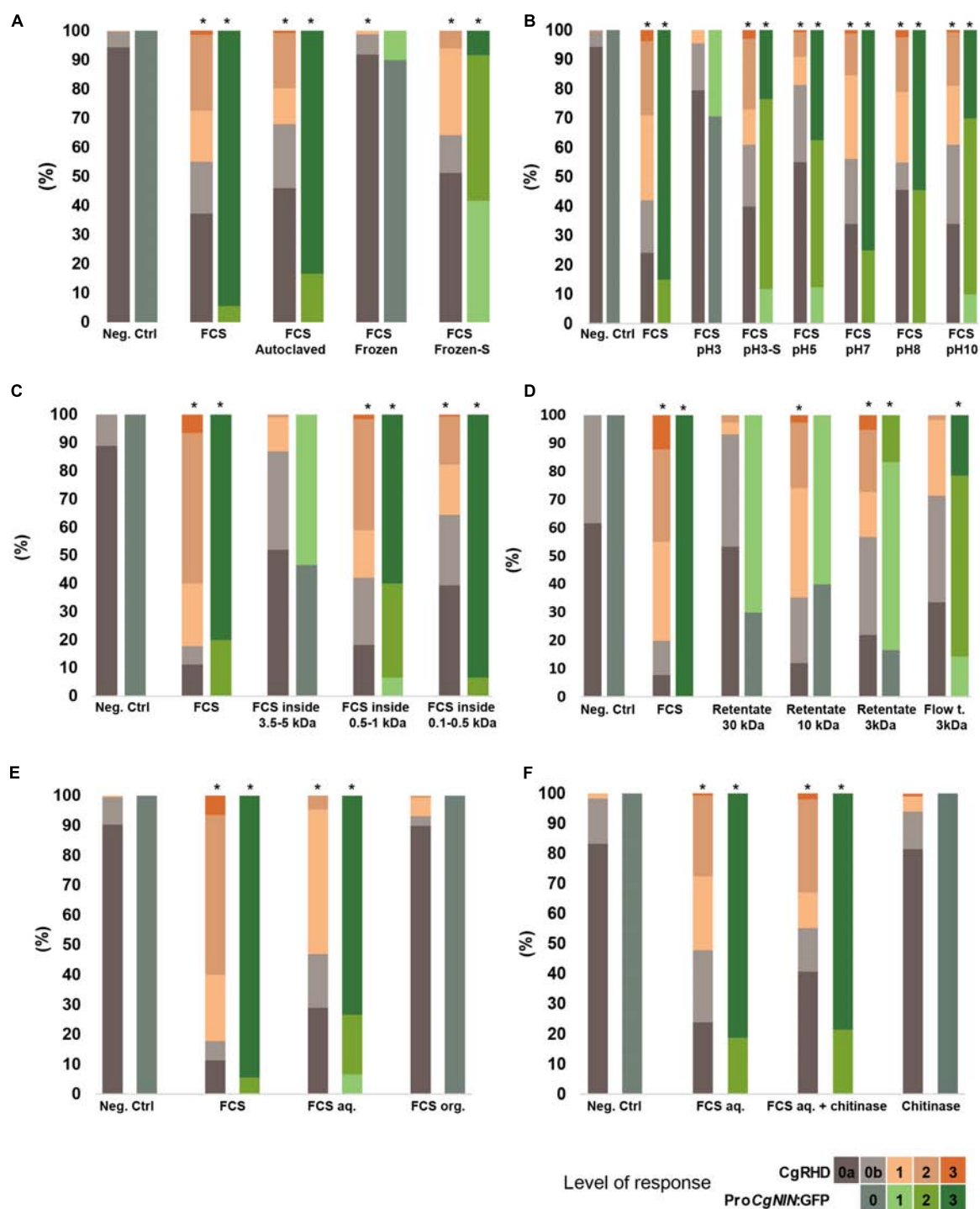
FIGURE 3 | Physico-chemical properties of CgRHDF and CgNINA. Orange bars show the proportion of deformed root hairs in short lateral roots 2 days after contact with FCS submitted to the different treatments. BAP medium diluted 100 times was used as a negative control. Asterisks above bars indicate symbiotic responses significantly different from the negative control (P < 5%). **(A)** Temperature sensitivity. High levels of CgRHD and GFP were detected in autoclaved FCS but no significant activity was present in FCS that was previously frozen. Sonication of frozen FCS (Frozen-S) allowed partial recovery of CgRHD and CgNINA activities. **(B)** pH sensitivity: FCS were incubated at different pH for one hour at the indicated pH. Sonication of FCS incubated at pH3 (pH3-S) restored CgRHD and CgNINA levels similar to the untreated control. **(C)** Size estimation using a dialysis tubing. CgRHD and CgNINA activities inside dialysis tubings with the indicated cutoffs was scored **(D)** Size estimation using centrifugal filters. FCS were submitted to successive filtrations using filters with the indicated cut-offs. CgRHD and CgNINA were evaluated on the retentate or the flow through (Flow t.). **(E)** CgRHD and CgNINA activity after 1-butanol extraction. Significant activities were only detected in the aqueous fraction (FCSaq) and not in the organic fraction (FCSorg). **(F)** Sensitivity to Chitinase. FCSaq incubated with chitinase showed similar CgRHD and CgNINA activities compared to untreated FCS.

FIGURE 4 | Distribution of CgRHD, CgNINA, and AgRHD activities in *Frankia* strains representative of *Frankia* diversity. A genome-wide phylogenetic tree of 17 *Frankia* strains and *S. coelicolor* was constructed using the concatenated sequences of 420 shared proteins. Distances were computed by the Jukes-Cantor method and the neighbor-joining method was used to build the tree. One thousand bootstrap replications were used to evaluate statistical support for branches. Cell-free supernatant fluids of each strain were tested for their ability to induce RHD and the activation of Pro*CgNIN* on *C. glauca* and RHD on *A. glutinosa*. Levels of response recorded for each strain correspond to the maximum deformation index and the maximum GFP fluorescence level for *C. glauca* and the maximum deformation level for *A. glutinosa*.

experiment performed with RE from *C. glauca* or without RE and the results obtained were not significantly different from the negative control (**Supplementary Figure S2**). Together, these results showed that *F. discariae* is able to synthetize molecules able to induce CgRHD, AgRHD, and the activation of Pro*CgNIN*. The activation of Pro*CgNIN* was enhanced when *F. discariae* was incubated with RE from *C. glauca* and was possibly below detection limits using the less sensitive equipment described in (Chabaud et al., 2016).

DISCUSSION

Presence of CgRHDF and Comparison to CgNINA

In legumes, RHD is one of the earliest visible responses induced upon recognition of rhizobial NFs by the host plant, and the development of a bioassay based on RHD was crucial to identify the chemical nature of NFs (Lerouge et al., 1990). The perception of NFs also provokes significant alterations of gene expression and notably the expression of symbiosis-induced genes such as *MtEnod11* (Journet et al., 2001) and *NIN* (Schauser et al., 1999; Radutoiu et al., 2003). In actinorhizal plants infected intracellularly such as *C. glauca* and *A. glutinosa*, RHD is also one of the first visible responses to *Frankia*

inoculation and factors able to induce RHD in *Alnus* (AgRHDF) have been partially purified and characterized (Cérémonie et al., 1999). In *C. glauca*, we have used transgenic plants expressing a Pro*CgNIN:GFP* fusion to characterize CgNINA, a factor present in cell free *F. casuarinae* supernatant fluids able to activate the *CgNIN* promoter in *C. glauca* root hairs. Here we have shown that a CgRHDF is also present in cell-free *F. casuarinae* supernatant fluids. Furthermore, we have shown that CgRHDF and CgNINA share similar physico-chemical properties listed in **Table 1**. Interestingly the experiment performed with centrifugal filters suggests that CgNINA and CgRHDF are both small molecules with slightly different sizes, CgNINA being probably smaller than CgRHDF. This observation is intriguing if CgNINA or CgRHDF are the actinorhizal analogs of rhizobial NFs because NFs are known to induce both RHD and the expression of early nodulins such as NIN. Responses obtained with centrifugal filters were however, less contrasted compared to the other experiments shown here and residual CgRHD activity was still present in the 3kDa flow through. We therefore cannot exclude that CgNINA also possess a small RHD activity and additional experiments are needed to confirm this hypothesis. If CgRHDF and CgNINA can indeed be separated, it would be interesting to know if those molecules are able to induce the high frequency nuclear Ca^{2+} spiking in growing *C. glauca* root hairs as described previously (Chabaud et al.,

	CgRHDF	CgNINA	AgRHDF	Rhizobial NF
Induction	Inducible	Inducible	Inducible	Inducible
	(root exudates)	(root exudates)	(root exudates)	(flavonoids)
Size	1–3 kDa	1–3 kDa	1.2–3 kDa	1 kDa
Thermal stability	Thermoresistant	Thermoresistant	Thermoresistant	Thermoresistant
Active concentration	10^{-2}–10^{-3} (supernatant)	10^{-1}–10^{-4} (supernatant)	10^{-3} 10^{-5} (supernatant)	Down to 10^{-12} M
Hydrophilicity	Hydrophilic	Hydrophilic	Hydrophilic	Amphiphilic
Chitinase action	resistant	resistant	resistant	sensitive

2016). The ability to induce Ca^{2+} oscillations in response to symbiotic bacteria is a common feature of nodulating species within the nitrogen-fixing clade (Granqvist et al., 2015). Alternatively we hypothesize that CgRHDF and CgNINA are the same molecule but a cofactor or a specific decoration is needed to enhance CgRHD activity without affecting its ability to activate Pro*CgNIN*. The 3 kDa centrifugal filter possibly eliminated decorated molecules with higher mass or cofactors and therefore strong activity was only detected with the CgNINA bioassay.

Comparison With AgRHDF and Rhizobial Nod Factors

CgRHDF and CgNINA also share many characteristics with AgRHDF, the corresponding factor characterized using *A. glutinosa*/*F. alni* (**Table 1**; Ghelue et al., 1997; Cérémonie et al., 1999). However, both factors appear to be structurally different from the rhizobial NFs because unlike NFs they are not found in the organic phase following a butanol extraction and are not sensitive to the endochitinase from *Aeromonas hydrophila* (Cérémonie et al., 1999) or the exochitinase from *S. griseus* (Cérémonie et al., 1999; Chabaud et al., 2016). This difference is in agreement with (1) the lack of *nodA* genes in the sequenced genomes of *F. casuarinae* and *F. alni* (Normand et al., 2007). (2) the absence of chitin oligomers in *F. alni* supernatant fluids (Cérémonie et al., 1999), and (3) the failure of NFs from the broad host-range rhizobia NGR234 to elicit RHD or Ca^{2+} spiking in *A. glutinosa* or *C. glauca* (Cérémonie et al., 1999; Granqvist et al., 2015; Chabaud et al., 2016). The possibility that actinorhizal recognition is mediated by molecules that are not hydrolyzed by tested chitinases is unexpected because downstream components of the NF signaling pathway are conserved not only between actinorhizal and rhizobial symbioses (Svistoonoff et al., 2014; Griesmann et al., 2018), but also between rhizobial and arbuscular-mycorrhizal symbioses where chitin-derived Myc-LCOs and COs play an important role as signaling molecules (Camps et al., 2015). Putative orthologs of NF receptors are present in *C. glauca* and *A. glutinosa* (Hocher et al., 2011). We are currently studying whether these genes play a role in actinorhizal symbioses. Because LysM receptor kinases have been shown to recognize not only chitin-derived molecules but also peptidoglycans and exopolysaccharides (Willmann et al., 2011; Kawaharada et al., 2015), orthologs of genes encoding NF receptors could be involved in the recognition of

CgRHDF/NINA or AgRHDF, even if their chemical backbone is not chitin-based.

Presence of NINA and CgRHDF in Other *Frankia* Strains

In most legumes, NFs allow the specific recognition between the host plant and its symbiotic rhizobia (Masson-Boivin et al., 2009; Oldroyd, 2013). Changes in specific decorations often result in host incompatibility (Dénarié et al., 1996) but NFs from incompatible strains can induce symbiotic responses such as RHD and activation of symbiotic genes when applied at increased concentrations (Roche et al., 1991). These results can be explained by changes of affinity between NFs and the cognate NF receptors able to recognize the chitin backbone and also the modified backbone structure. A misrecognition leads to decreased affinity but this can be compensated by increased amounts of substrate (Dénarié et al., 1996). The symbiotic responses in non-host plants reported here point to a similar mechanism in *C. glauca*: strains from clades I and III possibly synthesize molecules sharing a common molecular backbone that is recognized by *C. glauca* receptors inducing RHD and the activation of Pro*CgNIN* promoter. Optimal recognition is achieved for compatible strains (clade Ic) and some related strains (*F. alni* from clade Ia) but only the backbone would be recognized for more distant strains (clade III). This recognition is not detectable for non-infective strains (clade IV) and the distantly related strain *F. coriariae* suggesting that those strains do not produce sufficient amounts of this recognized backbone under the tested conditions.

Comparison Between CgRHDF and AgRHDF

Compared to CgRHDF and CgNINA, the distribution of AgRHDF seems less related to phylogeny. Generally, AgRHDF levels were stronger for clade III strains and several strains without any CgRHDF or CgNINA activity (*F. coriariae* and two uninfective strains from clade IV). These differences suggest that the AgRHDF assay detects smaller concentrations of deforming factors or that root hairs of *A. glutinosa* are deformed by a wider range of molecules compared to *C. glauca*. This second hypothesis is in agreement with RHD detected in *Alnus* roots incubated with non-*Frankia* bacteria or fungi (Berry and Torrey, 1983; Knowlton and Dawson, 1983; Prin and Rougier, 1987; Sequerra et al., 1994).

Impact of Root Exudates

We also found that the nature of RE used to incubate *Frankia* cultures could have an impact on RHDF and CgNINA activities. Unexpectedly lower CgRHDF and CgNINA activities were found when *F. discariae* was cultivated with RE from *O. trinervis* compared to RE from *C. glauca*. In legumes, specific flavonoids secreted by the host plant induce the expression of *nod* genes and the synthesis of NFs (Oldroyd, 2013). RE secreted by the host plant probably also play a role in actinorhizae formation because the incubation with RE induces morphological changes in *Frankia* and accelerate the nodulation process (Gabbarini and Wall, 2008, 2011; Beauchemin et al., 2012). Myricaceae seed extracts also influence *Frankia* growth (Bagnarol et al., 2007). In *Alnus* AgRHDF is reported to be produced either constitutively (McEwan et al., 1992; Ghelue et al., 1997; Cérémonie et al., 1999) or upon induction with RE (Prin and Rougier, 1987). Different plant RE have different effects on *Frankia* physiology. Information about CgRHDF is scarce but flavonoids isolated from *Casuarina* seeds have been shown to induce the production of CgRHDF by the *Casuarina*-infective BR strain (Selim, 1995). Increased CgRHD activity in *F. discariae* incubated with *C. glauca* RE could be due to increased amounts of flavonoids in *Casuarina* RE compared to *O. trinervis*.

AUTHOR CONTRIBUTIONS

MC, VH, PP, MS, and SSv conceived and designed the experiments. MC, PP, PF, VH, and AC-M performed the experiments. MC, SSv, PP, PF, and VH analyzed the data. SS and MC statistical analyzed the data. MS, MG, MC, VH, HG, DG, AC-M, SP, AC, MN, ES, CP, and SSv contributed reagents, materials, analysis, and tools. MC, SSv, and LT wrote the paper. All authors read the final version of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2018.01494/full#supplementary-material

REFERENCES

Alloisio, N., Queiroux, C., Fournier, P., Pujic, P., Normand, P., Vallenet, D., et al. (2010). The *Frankia alni* symbiotic transcriptome. *Mol. Plant. Microbe Interact.* 23, 593–607. doi: 10.1094/MPMI-23-5-0593

Bagnarol, E., Popovici, J., Alloisio, N., Maréchal, J., Pujic, P., Normand, P., et al. (2007). Differential Frankia protein patterns induced by phenolic extracts from *Myricaceae* seeds. *Physiol. Plant.* 130, 380–390. doi: 10.1111/j.1399-3054.2007.00875.x

Beauchemin, N. J., Furnholm, T., Lavenus, J., Svistoonoff, S., Doumas, P., Bogusz, D., et al. (2012). Casuarina root exudates alter the physiology, surface properties, and plant infectivity of *Frankia* sp. strain CcI3. *Appl. Environ. Microbiol.* 78, 575–580. doi: 10.1128/AEM.06183-11

Benson, D. R., Brooks, J. M., Huang, Y., Bickhart, D. M., and Mastronunzio, J. E. (2011). The biology of *Frankia* sp. strains in the post-genome Era. *Mol. Plant. Microbe Interact.* 24, 1310–1316. doi: 10.1094/MPMI-06-11-0150

Berry, A. M., and Torrey, J. G. (1983). Root hair deformation in the infection process of *Alnus rubra*. *Can. J. Bot.* 61, 2863–2876. doi: 10.1139/b83-319

Camps, C., Jardinaud, M.-F., Rengel, D., Carrère, S., Hervé, C., Debellé, F., et al. (2015). Combined genetic and transcriptomic analysis reveals three major signalling pathways activated by Myc-LCOs in Medicago truncatula. *New Phytol.* 208, 224–240. doi: 10.1111/nph.13427

Cérémonie, H., Cournoyer, B., Maillet, F., Normand, P., and Fernandez, M. P. (1998). Genetic complementation of rhizobial nod mutants with Frankia DNA: artifact or reality? *Mol. Gen. Genet. MGG* 260, 115–119. doi: 10.1007/s004380050877

Cérémonie, H., Debellé, F., and Fernandez, M. P. (1999). Structural and functional comparison of Frankia root hair deforming factor and rhizobia Nod factor. *Can. J. Bot.* 77, 1293–1301. doi: 10.1139/b99-060

Chabaud, M., Gherbi, H., Pirolles, E., Vaissayre, V., Fournier, J., Moukouanga, D., et al. (2016). Chitinase-resistant hydrophilic symbiotic factors secreted by

Frankia activate both Ca2 + spiking and NIN gene expression in the actinorhizal plant *Casuarina glauca*. *New Phytol.* 209, 86–93. doi: 10.1111/nph.13732

Clavijo, F., Diedhiou, I., Vaissayre, V., Brottier, L., Acolatse, J., Moukouanga, D., et al. (2015). The Casuarina NIN gene is transcriptionally activated throughout Frankia root infection as well as in response to bacterial diffusible signals. *New Phytol.* 208, 887–903. doi: 10.1111/nph.13506

Contreras-Moreira, B., and Vinuesa, P. (2013). GET_Homologues, a versatile software package for scalable and robust microbial pangenome analysis. *Appl. Environ. Microbiol.* 79, 7696–7701. doi: 10.1128/AEM.02411-13

Dawson, J. O. (eds). (2008). "Ecology of actinorhizal plants," in *Nitrogen-fixing Actinorhizal Symbioses* (Berlin: Springer), 199–234. doi: 10.1007/978-1-4020-3547-0_8

Dénarié, J., Débelle, F., and Prome, J.-C. (1996). Rhizobium lipo-chitooligosaccharide nodulation factors: signaling molecules mediating recognition and morphogenesis. *Annu. Rev. Biochem.* 65, 503–535. doi: 10.1146/annurev.bi.65.070196.002443

Fahraeus, G. (1957). The infection of clover root hairs by nodule bacteria studied by a simple glass slide technique. *Microbiology* 16, 374–381. doi: 10.1099/00221287-16-2-374

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791. doi: 10.1111/j.1558-5646.1985.tb00420.x

Gabbarini, L., and Wall, L. (2008). Analysis of nodulation kinetics in *Frankia–Discaria trinervis* symbiosis reveals different factors involved in the nodulation process. *Physiol. Plant.* 133, 776–785. doi: 10.1111/j.1399-3054.2008.01096.x

Gabbarini, L., and Wall, L. (2011). Diffusible factors involved in early interactions of actinorhizal symbiosis are modulated by the host plant but are not enough to break the host range barrier. *Funct. Plant Biol.* 38, 671–681. doi: 10.1071/FP11003

Ghelue, M. V., Løvaas, E., Ringø, E., and Solheim, B. (1997). Early interactions between *Alnus glutinosa* and *Frankia* strain ArI3. Production and specificity

of root hair deformation factor (s). *Physiol. Plant.* 99, 579–587. doi: 10.1111/j. 1399-3054.1997.tb05360.x

Granqvist, E., Sun, J., Op den Camp, R., Pujic, P., Hill, L., Normand, P., et al. (2015). Bacterial-induced calcium oscillations are common to nitrogen-fixing associations of nodulating legumes and nonlegumes. *New Phytol.* 207, 551–558. doi: 10.1111/nph.13464

Griesmann, M., Chang, Y., Liu, X., Song, Y., Haberer, G., Crook, M. B., et al. (2018). Phylogenomics reveals multiple losses of nitrogen–fixing root nodule symbiosis. *Science* 361:eaat1743. doi: 10.1126/science.aat1743

Hocher, V., Alloisio, N., Auguy, F., Fournier, P., Doumas, P., Pujic, P., et al. (2011). Transcriptomics of actinorhizal symbioses reveals homologs of the whole common symbiotic signaling cascade. *Plant Physiol.* 156, 700–711. doi: 10.1104/pp.111.174151

Hocher, V., Auguy, F., Argout, X., Laplaze, L., Franche, C., and Bogusz, D. (2006). Expressed sequence-tag analysis in *Casuarina glauca* actinorhizal nodule and root. *New Phytol.* 169, 681–688. doi: 10.1111/j.1469-8137.2006.01644.x

Imanishi, L., Vayssières, A., Franche, C., Bogusz, D., Wall, L., and Svistoonoff, S. (2011). Transformed hairy roots of *Discaria trinervis*: a valuable tool for studying actinorhizal symbiosis in the context of intercellular infection. *Mol. Plant. Microbe Interact.* 24, 1317–1324. doi: 10.1094/MPMI-03-11-0078

Journet, E. P., El-Gachtouli, N., Vernoud, V., de Billy, F., Pichon, M., Dedieu, A., et al. (2001). Medicago truncatula ENOD11: a novel RPRP-encoding early nodulin gene expressed during mycorrhization in arbuscule-containing cells. *Mol. Plant. Microbe Interact.* 14, 737–748. doi: 10.1094/MPMI.2001.14.6.737

Jukes, T. H., and Cantor, C. R. (1969). *Evolution of Protein Molecules*. Cambridge, MA: New York Academic Press, 21–132.

Kawaharada, Y., Kelly, S., Nielsen, M. W., Hjuler, C. T., Gysel, K., Muszyński, A., et al. (2015). Receptor-mediated exopolysaccharide perception controls bacterial infection. *Nature* 523, 308–312. doi: 10.1038/nature14611

Knowlton, S., and Dawson, J. O. (1983). Effects of *Pseudomonas cepacia* and cultural factors on the nodulation of *Alnus rubra* roots by Frankia. *Can. J. Bot.* 61, 2877–2882. doi: 10.1139/b83-320

Ktari, A., Gueddou, A., Nouioui, I., Miotello, G., Sarkar, I., Ghodhbane-Gtari, F., et al. (2017). Host plant compatibility shapes the proteogenome of *Frankia coriariae*. *Front. Microbiol.* 8:720. doi: 10.3389/fmicb.2017.00720

Kucho, K., Kakoi, K., Yamaura, M., Higashi, S., Uchiumi, T., and Abe, M. (2009). Transient transformation of Frankia by fusion marker genes in liquid culture. *Microbes Environ.* 24, 231–240. doi: 10.1264/jsme2.ME09115

Kucho, K.-I., Tamari, D., Matsuyama, S., Nabekura, T., and Tisa, L. S. (2017). Nitrogen fixation mutants of the actinobacterium *Frankia casuarinae* CcI3. *Microbes Environ.* 32, 344–351. doi: 10.1264/jsme2.ME17099

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404

Lerouge, P., Roche, P., Faucher, C., Maillet, F., Truchet, G., Promé, J. C., et al. (1990). Symbiotic host-specificity of *Rhizobium meliloti* is determined by a sulphated and acylated glucosamine oligosaccharide signal. *Nature* 344, 781–784. doi: 10.1038/344781a0

Lhuissier, F. G. P., De Ruijter, N. C. A., Sieberer, B. J., Esseling, J. J., and Emons, A. M. C. (2001). Time course of cell biological events evoked in legume root hairs by rhizobium Nod Factors: state of the art. *Ann. Bot.* 87, 289–302. doi: 10.1006/anbo.2000.1333

Marsh, J. F., Rakocevic, A., Mitra, R. M., Brocard, L., Sun, J., Eschstruth, A., et al. (2007). Medicago truncatula NIN is essential for rhizobial-independent nodule organogenesis induced by autoactive calcium/calmodulin-dependent protein kinase. *Plant Physiol.* 144, 324–335. doi: 10.1104/pp.106. 093021

Masson-Boivin, C., Giraud, E., Perret, X., and Batut, J. (2009). Establishing nitrogen-fixing symbiosis with legumes: how many rhizobium recipes? *Trends Microbiol.* 17, 458–466. doi: 10.1016/j.tim.2009.07.004

Mastronunzio, J. E., and Benson, D. R. (2010). Wild nodules can be broken: proteomics of Frankia in field-collected root nodules. *Symbiosis* 50, 13–26. doi: 10.1007/s13199-009-0030-1

McEwan, N. R., Green, D. C., and Wheeler, C. T. (1992). Utilisation of the root hair curling reaction in *Alnus glutinosa* for the assay of nodulation signal compounds. *Acta Oecol.* 13, 509–510.

Ngom, M., Gray, K., Diagne, N., Oshone, R., Fardou, J., Gherbi, H., et al. (2016). Performance of diverse Frankia strains on salt-stressed *Casuarina glauca* and

Casuarina equisetifolia Plants. *Front. Plant Sci.* 7:1331. doi: 10.3389/fpls.2016. 01331

Nguyen, T. V., Wibberg, D., Battenberg, K., Blom Brian, J. B., Heuvel, V., Berry, A. M., et al. (2016). An assemblage of Frankia Cluster II strains from California contains the canonical nod genes and also the sulfotransferase gene nodH. *BMC Genomics* 17:796. doi: 10.1186/s12864-016-3140-1

Normand, P., Lapierre, P., Tisa, L. S., Gogarten, J. P., Alloisio, N., Bagnarol, E., et al. (2007). Genome characteristics of facultatively symbiotic *Frankia* sp. strains reflect host range and host plant biogeography. *Genome Res.* 17, 7–15. doi: 10.1101/gr.5798407

Oldroyd, G. E. D. (2013). Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nat. Rev. Microbiol.* 11, 252–263. doi: 10.1038/nrmicro2990

Parham, J. A., and Deng, S. P. (2000). Detection, quantification and characterization of β-glucosaminidase activity in soil. *Soil Biol. Biochem.* 32, 1183–1190. doi: 10.1016/S0038-0717(00)00034-1

Perrine-Walker, F., Gherbi, H., Imanishi, L., Hocher, V., Ghodhbane-Gtari, F., Lavenus, J., et al. (2011). Symbiotic signaling in actinorhizal symbioses. *Curr. Protein Pept. Sci.* 12, 156–164. doi: 10.2174/1389203117956 84896

Persson, T., Battenberg, K., Demina, I. V., Vigil-Stenman, T., Vanden Heuvel, B., Pujic, P., et al. (2015). *Candidatus* Frankia Datiscae Dg1, the actinobacterial microsymbiont of *Datisca glomerata*, expresses the canonical nod genes nodABC in symbiosis with its host plant. *PLoS One* 10:e0127630. doi: 10.1371/ journal.pone.0127630

Prin, Y., and Rougier, M. (1987). Preinfection events in the establishment of *Alnus-frankia* symbiosis: study of the root hair deformation step. *Life Sci. Adv.* 6, 98–106.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna: R Core Team.

Radutoiu, S., Madsen, L. H., Madsen, E. B., Felle, H. H., Umehara, Y., Grønlund, M., et al. (2003). Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *Nature* 425, 585–592. doi: 10.1038/nature02039

Roche, P., Debellé, F., Maillet, F., Lerouge, P., Faucher, C., Truchet, G., et al. (1991). Molecular basis of symbiotic host specificity in *Rhizobium meliloti*: nodH and nodPQ genes encode the sulfation of lipo-oligosaccharide signals. *Cell* 67, 1131–1143. doi: 10.1016/0092-8674(91)90290-F

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425. doi: 10.1093/ oxfordjournals.molbev.a040454

Schauser, L., Roussis, A., Stiller, J., and Stougaard, J. (1999). A plant regulator controlling development of symbiotic root nodules. *Nature* 402, 191–195. doi: 10.1038/46058

Selim, S. M. A. (1995). *La Symbiose Casuarina-Frankia. Optimisation de la Croissance et Approche Biologique de la Reconnaissance Microorganisme-Plante Hote*. Available at: http://www.theses.fr/1995PA112288 [accessed February 27, 2018].

Sequerra, J., Capellano, A., Faure-Raynard, M., and Moiroud, A. (1994). Root hair infection process and myconodule formation on *Alnus incana* by *Penicillium nodositatum*. *Can. J. Bot.* 72, 955–962. doi: 10.1139/b94-121

Soyano, T., Hirakawa, H., Sato, S., Hayashi, M., and Kawaguchi, M. (2014). NODULE INCEPTION creates a long-distance negative feedback loop involved in homeostatic regulation of nodule organ production. *Proc. Natl. Acad. Sci. U.S.A.* 111, 14607–14612. doi: 10.1073/pnas.14127 16111

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 Genet.* 9:e1003352. doi: 10.1371/ journal.pgen.1003352

Svistoonoff, S., Gherbi, H., Nambiar-Veetil, M., Zhong, C., Michalak, Z., Laplaze, L., et al. (2010a). Contribution of transgenic Casuarinaceae to our knowledge of the actinorhizal symbioses. *Symbiosis* 50, 3–11. doi: 10.1007/ s13199-009-0036-8

Svistoonoff, S., Sy, M. O., Diagne, N., Barker, D. G., Bogusz, D., and Franche, C. (2010b). Infection-specific activation of the Medicago truncatula Enod11 early nodulin gene promoter during actinorhizal root nodulation. *Mol. Plant. Microbe Interact.* 23, 740–747. doi: 10.1094/MPMI-23-6-0740

Svistoonoff, S., Hocher, V., and Gherbi, H. (2014). Actinorhizal root nodule symbioses: what is signalling telling on the origins of nodulation? *Curr. Opin. Plant Biol.* 20, 11–18. doi: 10.1016/j.pbi.2014.03.001

Tisa, L. S., Oshone, R., Sarkar, I., Ktari, A., Sen, A., and Gtari, M. (2016). Genomic approaches toward understanding the actinorhizal symbiosis: an update on the status of the Frankia genomes. *Symbiosis* 70, 5–16. doi: 10.1007/s13199-016-0390-2

Vessey, K. J., Pawlowski, K., and Bergman, B. (2005). Root-based N 2-fixing symbioses: legumes, actinorhizal plants, *Parasponia* sp. and cycads. *Root Physiol. Gene Funct.* 266, 51–78. doi: 10.1007/1-4020-4099-7_3

Wall, L. G. (2000). The actinorhizal symbiosis. *J. Plant Growth Regul.* 19, 167–182.

Willmann, R., Lajunen, H. M., Erbs, G., Newman, M.-A., Kolb, D., Tsuda, K., et al. (2011). Arabidopsis lysin-motif proteins LYM1 LYM3 CERK1 mediate bacterial peptidoglycan sensing and immunity to bacterial infection. *Proc. Natl. Acad. Sci. U.S.A.* 108, 19824–19829. doi: 10.1073/pnas.1112862108

Yoro, E., Suzaki, T., Toyokura, K., Miyazawa, H., Fukaki, H., and Kawaguchi, M. (2014). a positive regulator of nodule organogenesis, nodule inception, acts as a negative regulator of rhizobial infection in *Lotus japonicus*. *Plant Physiol.* 165, 747–758. doi: 10.1104/pp.113.233379

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Comparative Analysis of the Nodule Transcriptomes of *Ceanothus thyrsiflorus* (Rhamnaceae, Rosales) and *Datisca glomerata* (Datiscaceae, Cucurbitales)

Marco G. Salgado[1], Robin van Velzen[2], Thanh Van Nguyen[1], Kai Battenberg[3], Alison M. Berry[3], Daniel Lundin[4,5] and Katharina Pawlowski[1]*

[1] Department of Ecology, Environment and Plant Sciences, Stockholm University, Stockholm, Sweden, [2] Laboratory of Molecular Biology, Department of Plant Sciences, Wageningen University, Wageningen, Netherlands, [3] Department of Plant Sciences, University of California, Davis, Davis, CA, United States, [4] Centre for Ecology and Evolution in Microbial Model Systems, Linnaeus University, Kalmar, Sweden, [5] Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden

Two types of nitrogen-fixing root nodule symbioses are known, rhizobial and actinorhizal symbioses. The latter involve plants of three orders, Fagales, Rosales, and Cucurbitales. To understand the diversity of plant symbiotic adaptation, we compared the nodule transcriptomes of *Datisca glomerata* (Datiscaceae, Cucurbitales) and *Ceanothus thyrsiflorus* (Rhamnaceae, Rosales); both species are nodulated by members of the uncultured *Frankia* clade, cluster II. The analysis focused on various features. In both species, the expression of orthologs of legume Nod factor receptor genes was elevated in nodules compared to roots. Since arginine has been postulated as export form of fixed nitrogen from symbiotic *Frankia* in nodules of *D. glomerata,* the question was whether the nitrogen metabolism was similar in nodules of *C. thyrsiflorus*. Analysis of the expression levels of key genes encoding enzymes involved in arginine metabolism revealed up-regulation of arginine catabolism, but no up-regulation of arginine biosynthesis, in nodules compared to roots of *D. glomerata,* while arginine degradation was not upregulated in nodules of *C. thyrsiflorus*. This new information corroborated an arginine-based metabolic exchange between host and microsymbiont for *D. glomerata,* but not for *C. thyrsiflorus.* Oxygen protection systems for nitrogenase differ dramatically between both species. Analysis of the antioxidant system suggested that the system in the nodules of *D. glomerata* leads to greater oxidative stress than the one in the nodules of *C. thyrsiflorus,* while no differences were found for the defense against nitrosative stress. However, induction of nitrite reductase in nodules of *C. thyrsiflorus* indicated that here, nitrite produced from nitric oxide had to be detoxified. Additional shared features were identified: genes encoding enzymes involved in thiamine biosynthesis were found to be upregulated in the nodules of both species. Orthologous nodule-specific subtilisin-like proteases that have been linked to the infection process in actinorhizal Fagales, were also upregulated in the nodules of *D. glomerata* and

C. thyrsiflorus. Nodule-specific defensin genes known from actinorhizal Fagales and Cucurbitales, were also found in *C. thyrsiflorus.* In summary, the results underline the variability of nodule metabolism in different groups of symbiotic plants while pointing at conserved features involved in the infection process.

INTRODUCTION

Nitrogen is the element that most often limits plant growth. Members of four different plant orders can form root nodule symbioses with nitrogen-fixing soil bacteria (Mylona et al., 1995). In the root nodules, bacteria fix ambient dinitrogen while being hosted within plant cells. There are two types of root nodule symbioses; (i) most legume species and the non-legume genus *Parasponia* (Cannabaceae, Rosales) interact with a polyphyletic group of Gram-negative proteobacteria collectively known as rhizobia and (ii) actinorhizal plants interact with Gram-positive actinobacteria from the genus *Frankia*. The latter encompass 24 genera distributed over eight families across three orders: Cucurbitales (Datiscaceae and Coriariaceae), Fagales (Betulaceae, Casuarinaceae, and Myricaceae), and Rosales (Elaeagnaceae, Rhamnaceae, and Rosaceae) (Soltis et al., 1995; Pawlowski and Demchenko, 2012). Together with Fabales, they form a monophyletic group within the Fabid clade (Soltis et al., 1995; Angiosperm Phylogeny Group, 2016). Despite sharing a relatively recent ancestor (ca. 100 mya; Bell et al., 2010), actinorhizal species show high diversity in nodule anatomy, physiology, and metabolism (Swensen, 1996; Pawlowski and Demchenko, 2012).

Phylogenetic analyses have shown that symbiotic *Frankia* strains comprise three clusters that mostly correlate with host specificity (Normand et al., 1996; Sen et al., 2014). Cluster I strains nodulate members of three families in the Fagales [Betulaceae, Casuarinaceae (except for the genus *Gymnostoma*), and Myricaceae (except for the genus *Morella*)]. Cluster II strains display a wide host range nodulating all actinorhizal Cucurbitales and members of two Rosales families (Rosaceae and the genus *Ceanothus* in the Rhamnaceae). Cluster III strains nodulate species in Rosales (Elaeagnaceae and Rhamnaceae, except for the genus *Ceanothus*), and Fagales (the genera *Gymnostoma* and *Morella*). Cluster II, representing the earliest divergent clade within *Frankia,* comprises strains of which so far only two have been cultured (Gtari et al., 2015; Persson et al., 2015; Nguyen et al., 2016; Gueddou et al., 2018). This study focuses on two host plant species of cluster II *Frankia* strains, one from the Cucurbitales (*Datisca glomerata,* Datiscaceae) and one from the Rosales (*Ceanothus thyrsiflorus,* Rhamnaceae).

Several aspects of root nodule symbioses well researched in legumes are still understudied in actinorhizal plants. For example, nodule organogenesis and uptake of bacteria by plant roots are the result of the exchange of diffusible signals between host and bacteria (Oldroyd, 2013). Rhizobia signal to their host plants *via* lipochitooligosaccharide (LCO) Nod factors.

In rhizobia, the synthesis of the LCO common backbone requires the enzymes encoded by the canonical *nod* genes *nodABC*. Plants perceive these LCO Nod factors by a heterodimer of LysM receptors, which then signal *via* the common symbiotic pathway that is shared with, and recruited from, arbuscular mycorrhizal symbioses (Oldroyd, 2013). Signaling in actinorhizal symbioses is less well examined. *Frankia* genome analysis led to the assumption that genes from clusters I and III do not signal via LCO Nod factors (Normand et al., 2007). Nevertheless, recent sequencing of *Frankia* cluster II genomes showed that they contain the canonical *nod* genes *nodABC*, and these genes are expressed *in planta* (Persson et al., 2015; Nguyen et al., 2016). So the first question concerns the presence and expression of genes orthologous to the known legume Nod factor receptor genes in roots and nodules of both plant species.

The second question to address is the mechanism of stable intracellular accommodation of *Frankia* within nodule cells. Cytological analysis of *D. glomerata* as representative of actinorhizal Cucurbitales have shown that the infection thread growth mechanism seems to differ from that used in Fagales and from that of Rosales (Berg et al., 1999; reviewed by Pawlowski and Demchenko, 2012). Genes encoding products relevant for infection thread growth have been identified in legumes (Suzaki et al., 2015), and the expression of their homologs could be examined in both actinorhizal species.

The third question concerns the nitrogen metabolism in nodules, specifically, the nitrogen source exported by the intracellular nitrogen-fixing microsymbionts to the plant cytoplasm. In legumes, it is generally accepted that this nitrogen source is ammonia, which is then assimilated in the cytosol of the infected nodule cells *via* the glutamine synthetase (GS)/glutamate synthase (GOGAT) pathway. High levels of plant GS expression in infected nodule cells have been shown for various legumes, such as *Phaseolus vulgaris* (Forde et al., 1989), *Glycine max* (Miao et al., 1991), and *Medicago sativa* (Temple et al., 1995). In similar fashion, ammonia has also been shown to represent the probable export product of *Frankia* in the actinorhizal tree *Alnus glutinosa* (Betulaceae, Fagales) based on the localization of plant GS (Hirel et al., 1982; Guan et al., 1997) and on the low levels of expression of *Frankia GS* in symbiosis (Alloisio et al., 2010). However, results of Berry et al. (2004, 2011) indicate that in *D. glomerata* nodules, *Frankia* exports an assimilated form of nitrogen, probably arginine, which is broken down in uninfected, rather than infected nodule cells, whereupon the ammonia is reassimilated in the GS/GOGAT pathway. Accordingly, plant GS expression is enhanced in nodules compared to roots, but not in infected cells (Berry et al., 2004). Thus, it remains unclear whether the export of an assimilated form of nitrogen from *Frankia*

rather than ammonia is a feature of the nodules of actinorhizal Cucurbitales, or a feature of nodules induced by cluster II *Frankia* strains. Furthermore, the principal form of nitrogen transported in the xylem, and thus the N metabolite pattern in nodules, differs among actinorhizal plants. In actinorhizal Rosales, asparagine was identified as the principal xylem amino acid in *Hippophae, Elaeagnus, Ceanothus,* and *Discaria* (Schubert, 1986; Valverde and Wall, 2003), while in *D. glomerata,* glutamine, and glutamate were identified as the major xylem nitrogen transport forms (Berry et al., 2004, 2011; Persson et al., 2016). In actinorhizal Fagales (*Alnus* and *Casuarina*), the ureide citrullin was identified as the principal xylem nitrogen transport form (Walsh et al., 1984; Schubert, 1986).

The fourth question concerns the side effects of the oxygen protection system for nitrogenase, namely, oxidative and nitrosative stress. The enzyme complex nitrogenase is irreversibly deactivated by oxygen (Shah and Brill, 1973); however, the high energy demands of nitrogen fixation require optimal respiratory activity. In rhizobial symbioses the host reconciles this conflict; i.e., an oxygen barrier in the nodule cortex results in microaerobic conditions in the area of infected cells, in which oxygen carriers (class II hemoglobins, "leghemoglobins," in legumes, and a class I hemoglobin in *Parasponia* spp.) enable optimal respiration (Minchin, 1997; Garrocho-Villegas et al., 2007). In contrast with rhizobia, *Frankia* strains can provide oxygen protection for nitrogenase themselves by differentiating specialized cell types, vesicles, surrounded by multi-layered envelopes containing hopanoids, bacterial steroid lipids. In these vesicles, nitrogenase is protected from oxygen (Parsons et al., 1987; Berry et al., 1993). As in actinorhizal nodules both host and bacteria can contribute to oxygen protection, the corresponding processes are diverse (Silvester et al., 1990). The process established in legumes leads to the production of high amounts of reactive oxygen species (ROS), first as a side product of leghemoglobin activity (Becana et al., 2000; Ott et al., 2005; Günther et al., 2007) and second as a side effect of mitochondrial respiration under hypoxia (Fukao and Bailey-Serres, 2004). This oxidative stress requires efficient detoxification. Furthermore, nodules are also exposed to nitrosative stress since nitric oxide (NO) is produced throughout the symbiosis in nodules of legumes (Hichri et al., 2015) and also in those of *Alnus firma* (Sasakura et al., 2006). NO can be synthesized by NO synthase, arise from non-enzymatic conversion of nitrite to NO in the apoplast (Bethke et al., 2004), or be produced by reductive pathways involving nitrate reductase (NR) or nitrite:NO reductase (Gupta and Igamberdiev, 2011). Detoxification of NO can be performed by class I hemoglobins (Igamberdiev et al., 2006; Perazzolli et al., 2006) and also by truncated hemoglobins (Sanz-Luque et al., 2015).

In nodules of *Ceanothus* spp., spherical vesicles resemble those found in *Frankia* liquid culture (Strand and Laetsch, 1977). Thus, similar to nodules of *A. incana* ssp. *rugosa,* the plant does not seem to contribute to prevent nitrogenase denaturation from oxygen (Silvester et al., 1988; Kleemann et al., 1994). In contrast, in nodules of *D. glomerata,* lanceolate *Frankia* vesicles arranged in radial orientation form a sphere around the central vacuole of the infected cells. As a second barrier at the place of oxygen access to this sphere, a thick layer of blanket mitochondria

provides physiological oxygen protection (Silvester et al., 1999). The expression of a truncated hemoglobin in these cells indicates nitrosative stress (Pawlowski et al., 2007; Sanz-Luque et al., 2015).

Another question concerns the role of vitamins. The induction of genes encoding enzymes involved in thiamine (vitamin B1) biosynthesis has long been known for nodules of actinorhizal Fagales (Ribeiro et al., 1996; Hocher et al., 2011) and recently was also reported for the model legume *Lotus japonicus* (Nagae et al., 2016a). Despite its increasing interest, the role of thiamine in root nodule symbioses has yet to be understood. However, some data are available; e.g., the loss of function for a key gene involved on its biosynthesis has led to reduced nodule number in *L. japonicus* (Nagae et al., 2016a,b). Thus, nodule function in legumes and actinorhizal Fagales seems to require high amounts of thiamine. In this context, it is interesting that in *Arabidopsis thaliana* thiamine participates in the adaptation to oxidative (Tunc-Ozdemir et al., 2009) as well as to abiotic stresses (Ribeiro et al., 2005; Rapala-Kozik et al., 2012). However, it remains unclear whether the requirement for increased thiamine biosynthesis also extends to nodules of actinorhizal Cucurbitales and Rosales. The synthesis of folic acid (vitamin B9) which modulates root architecture (Ayala-Rodríguez et al., 2017) was included in the analysis.

Light-induced gene expression is not expected in subterranean plant organs, so the transcription of the plastidic gene encoding ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) and the nuclear gene encoding RuBisCO activase in nodules of *D. glomerata* was a surprising result (*rbcL* and *rca;* Okubara et al., 1999). Yet, plants contain a variety of photoreceptors – phytochromes, cryptochromes and phototropins – to respond to light of different wavelengths, and these photoreceptors are not restricted to aerial parts, they also occur in roots (van Gelderen et al., 2018). Root development including nodulation is affected by light (Gundel et al., 2014; Shimomura et al., 2016). Moreover, some photoreceptors integrate light- and temperature signals (Casal and Qüesta, 2018). That being said, it is generally accepted that photosynthesis does not take place in subterranean organs, so the expression of *rbcL* and *rca* in *D. glomerata* nodules remained unexplained. Therefore, the analysis of expression of photosynthesis-related genes in nodules vs. roots was included in this study.

Subtilases are involved in many signaling pathways, e.g., pathways associated with organogenesis and senescence (Taylor and Qiu, 2017). In particular, subtilases are induced in plant-microbe interactions, e.g., arbuscular mycorrhiza (Takeda et al., 2009). In actinorhizal symbioses, nodule-specific subtilases were identified in Fagales (*A. glutinosa;* Ribeiro et al., 1995; *Casuarina glauca;* Laplaze et al., 2000) and recently also in Rosales (*Discaria trinervis;* Fournier et al., 2018). Therefore, their presence was also examined for nodules of *D. glomerata* and *C. thyrsiflorus.*

A mutualistic interaction has to guard against parasitic individuals that extract benefits without paying costs ("cheaters") and thus proliferate more efficiently than their mutualistic brethren. In natural environments, many rhizobia have very low nitrogen fixation capacity (Oono et al., 2009). Plants

from two groups of legumes, namely the Dalbergioids and the Inverted Repeat-Lacking Clade (IRLC), form cysteine-rich peptides (NCRs) in nodules which affect the differentiation of the microsymbionts. Their effects include amplification of the rhizobial genome, inhibition of rhizobial cell division, rhizobial cell elongation and -branching, as well as modification of the rhizobial plasma membrane (Van de Velde et al., 2010), a mechanism originally thought to have evolved for the control of rhizobial "cheaters." However, recent results show that NCRs can have positive as well as negative effects on rhizobia (Pan and Wang, 2017). Nodule-specific expression of cysteine-rich peptides – in this case, defensins – was also found for actinorhizal Fagales (*A. glutinosa;* Carro et al., 2015) and of Cucurbitales (*D. glomerata*; Demina et al., 2013). So the question remains whether other actinorhizal plants, especially from the Rosales, form NCRs.

In summary, in this study, transcriptome sequencing combined with Reverse Transcription – quantitative PCR (RT-qPCR) was employed to compare nodules induced by *Frankia* cluster II strains on *D. glomerata* (Cucurbitales) and *C. thyrsiflorus* (Rosales). We addressed the following research questions: (1) Does expression of legume LCO Nod factor receptor orthologs support their role in the perception of bacterial signal factors in both host plant species? (2) Is the expression of homologs of legume genes encoding proteins required for infection thread formation induced in nodules compared to roots? (3) What can the differential expression of genes encoding enzymes in plant nitrogen metabolism tell us about the nitrogen source exported by the intracellular nitrogen-fixing microsymbionts to the plant cytoplasm? (4) What can the expression of genes encoding enzymes of the antioxidant defense system and of globins involved in NO detoxification tell us about the oxidative and nitrosative stress associated with the different oxygen protection systems for nitrogenase realized in both types of nodules? (5) Are genes involved in thiamine and/or folate biosynthesis induced in either type of nodule? (6) Are photosynthesis-associated genes expressed in both types of nodules? (7) Are subtilases expressed in both types of nodules, and are these subtilases orthologs of the nodule-specific subtilases identified for *A. glutinosa, C. glauca*, and *D. trinervis*? (8) Are nodule-specific cysteine rich peptides (defensins) expressed in nodules of *C. thyrsiflorus,* i.e., in a representative of the actinorhizal Rosales? Contributions to answering these questions will shed light on the commonalities and differences in root nodule symbiosis in relation to plant phylogeny.

MATERIALS AND METHODS

Plant Material and Growth Conditions

Datisca glomerata seedlings were grown in a greenhouse under a photoperiod of 14 h light/8 h dark cycle with 22°C/19°C, respectively, and a relative humidity of 70%. They were fertilized with 1/4 strength Hoagland's solution with 10 mM nitrogen (Hoagland and Arnon, 1938). 8-Week-old seedlings were inoculated with ground nodules of *D. glomerata* infected

with *Candidatus* Frankia datiscae Dg1. After inoculation, plants were supplied with 1/4 strength Hoagland's solution without nitrogen. Whole root nodules were harvested 12 weeks post inoculation.

Ceanothus thyrsiflorus plants were purchased from a nursery, Corn Flower Farm (Elkgrove, CA, United States) as cuttings, and were grown in a greenhouse in University of California, Davis. For successful nodulation, the plants were repotted into new medium (UC mix/perlite 1:1) to remove any fertilizer given to the plants by the nursery. Nodulation status of each plant was checked at this point to ensure that no plants were nodulated prior to any further treatment. After 4 weeks, the plants were inoculated using nodules from *C. thyrsiflorus* that had been inoculated with soil collected in Sagehen Experimental Forest (Truckee, CA, United States). Plants were maintained in a greenhouse and watered with deionized water and Hoagland's solution without nitrogen. Plants were kept under natural daylight except during winter when they were kept under extended artificial daylight. The young (non-lignified) parts of nodules were cut off with a scalpel and flash frozen in liquid nitrogen.

Preparation of RNA-Seq Libraries From *D. glomerata* and *C. thyrsiflorus* Nodules

Total RNA was isolated from five (*D. glomerata*) and three (*C. thyrsiflorus*) independent nodules as described previously (Demina et al., 2013). RNA-Seq libraries were prepared in strand-specific mode and sequenced with an Illumina HiSeq2500 platform (Illumina) yielding 256,841,770 (*D. glomerata;* detailed in **Supplementary Table S1**) and 157,205,694 (*C. thyrsiflorus*) paired-end reads (100 nt each). For *C. thyrsiflorus* an extra library was prepared and sequenced with an Illumina MiSeq instrument (Illumina). This new library was strand-specific, poly(A)-enriched and yielded a total of 36,740,942 (300 nt) paired-end reads that were exclusively used for assembly.

Raw data have been archived in BioProject PRJNA454374 (*D. glomerata*) and BioProject PRJNA454377 (*C. thyrsiflorus*).

Filtering, Trimming and *de novo* Transcriptome Assemblies

For both species, *de novo* assemblies were generated following the same approach. Prokaryotic sequences were removed by mapping the raw reads against the genome of *Candidatus* Frankia datiscae Dg1 (GenBank accession NC_015656.1) using TopHat 2.0.12 (Trapnell et al., 2009). The quality of the retained reads was evaluated by FastQC 0.10.1 (Andrews, 2010). Illumina sequencing adapters and low quality reads were removed by Fastq-mcf 1.04.636 (Aronesty, 2013); next, low quality bases were trimmed at the 3'-end (10 nt) by Seqtk[1]. The resulting quality-filtered dataset, i.e., 200,795,069 paired-end reads (ca. 100 nt) for *D. glomerata* (summarized in **Supplementary Table S1**) and 36,589,680 paired-end reads (ca. 290 nt) for *C. thyrsiflorus*, was assembled with Trinity version

---

[1]https://github.com/lh3/seqtk

r20140717 using default parameters (-K 25, -L 25; Grabherr et al., 2011).

For both species, the completeness of the transcripts in terms of expected orthology was evaluated by BUSCO 2.0.1 (Simão et al., 2015) against the reference plant dataset, embryophyta_odb9, which contained 1440 protein sequences and orthogroup annotations for major clades.

The *D. glomerata* Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GGXR00000000. The version described in this paper is the first version, GGXR01000000. The *C. thyrsiflorus* Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GGXO00000000. The version described in this paper is the first version, GGXO01000000.

Transcriptome Functional Annotation

The *de novo* assembled transcriptomes were functionally annotated by similarity sequence search using Trinotate (Haas et al., 2013), yielding Gene Ontology terms, PFAM annotations and Enzyme commission (EC) numbers. Gene Ontology terms were assigned by Blast+ algorithms (Altschul et al., 1990) targeting the SwissProt non-redundant database. The longest ORFs were queried at the PFAM database (Punta et al., 2012) enabling domain predictions (HMMER 3.1b1; Finn et al., 2011). Additionally, longest ORFs were queried for signal peptides (SignalP 4.1; Petersen et al., 2011) and transmembrane domains (tmHMM 2.0c; Krogh et al., 2001). Trinotate is inclusive in the sense that it allows multiple PFAM entries to be assigned to a single contig. For simplicity, only the PFAM hit with lowest E-value was retained for each transcript.

Expression Profiling

The transcript abundance was estimated by RSEM (Li and Dewey, 2011), available as a Perl script in the Trinity suite (version r20140717; Grabherr et al., 2011). For both species, the raw reads prepared in HiSeq2500 mode were mapped against the assembled contigs and both the fragment per kilobase of transcript per million reads (FPKM) and transcripts per million reads (TPM) were inferred for each transcript.

To assess differential expression, the contribution of different contigs was collapsed in single PFAMs by the sum of their FPKM. Every PFAM entry was further normalized against the abundance of the housekeeping control elongation factor 1 alpha (EF-1α; PF03143.12) from which the mean and $\log_2[FC]$ was calculated in each library. These steps were carried out in RStudio (RStudio Team, 2015). EC numbers were used to graphically visualize MetaCyc metabolic pathways (Paley and Karp, 2006; Caspi et al., 2016). Functional annotation of transcriptomes and their expression profiles are accessible at doi: 10.17045/sthlmuni.6181763.v1.

Real-Time Quantitative RT-PCR

In order to validate the RNA-Seq measurements, differential expression in roots vs. nodules was assessed by RT-qPCR. The method used was essentially as described in Zdyb et al.

(2018). Primers were designed by Primer3 at NCBI Primer-Blast server and are listed in **Supplementary Table S2** (*D. glomerata*) and **Supplementary Table S3** (*C. thyrsiflorus*). Roots and nodules were harvested in liquid nitrogen and tissues were ground using pestle and mortar. Macerated tissues were immediately used for total RNA isolation using an on-column DNase treatment (Spectrum Total RNA isolation kit, Sigma-Aldrich, Germany). Prior to cDNA synthesis, 1 μg of total RNA was treated with RNase-free DNase using the Heat&Run kit (ArcticZymes, Norway). Total RNA was reverse transcribed in a final volume of 20 μl following the instructions of the TATAA GrandScript cDNA synthesis kit (TATAA Biocenter, Sweden); cDNA preparations were 10^{-1} diluted and 2 μl were used as a template in 10 μl PCR reactions; reactions were performed with 1x Maxima SYBR green (Thermo Fisher Scientific, Lithuania) and 300 nM of each primer in an Eco Real Time PCR System (Illumina, United States); applied thermal conditions were 10 min 95°C for initial denaturation, extension at 60°C for 30 sec, 45 cycles. Melting dissociation curves were examined in order to circumvent the possibility of primer dimer. Demonstrative exponential phase Cq values of the housekeeping gene EF-1α were used to calculate normalization factors. Statistics were based in a balanced assay; relative quantities were back-transformed to Cq values from which an unpaired 2-tail Student's test was conducted; p-values were adjusted according to Benjamini—Hochberg false discovery rate FDR (Benjamini and Hochberg, 1995). Statistics were performed in RStudio (RStudio Team, 2015).

Phylogenetic Analysis

In order to assess the phylogeny of LysM-type and LysM receptors in *D. glomerata* and *C. thyrsiflorus*, NFR1, NFR5, and EPR3 protein candidates were selected from the transcriptomes by reverse Blast (Altschul et al., 1990). Blasted queries were selected based on previous reverse genetics studies involving the model legume *L. japonicus* (Radutoiu et al., 2003; Kawaharada et al., 2015). Individual candidate protein sequences were blasted against selected taxa in the RefSeq database with an *E*-value cutoff of 1e-50. Selected target taxa are listed in **Supplementary Table S4**. Unique sequences from this set (*n* = 180) plus the *D. glomerata* and *C. thyrsiflorus* candidate sequences were aligned using Clustal Omega version 1.2.4 (Sievers et al., 2014). From the alignment, truncated sequences were removed. Well-aligned positions were selected with BMGE using the BLOSUM62 substitution matrix (Criscuolo and Gribaldo, 2010). Phylogenetic trees were estimated based on maximum likelihood using RAxML version 8.2.10 (Stamatakis, 2014) using the PROTGAMMAAUTO model and rapid bootstopping (autoMRE) (Pattengale et al., 2009). The full alignment and the maximum likelihood tree are accessible at doi: 10.17045/sthlmuni.6384200.v1.

To gain insight about the phylogeny of subtilases, candidate proteins from *D. glomerata* and *C. thyrsiflorus* were placed in the comprehensive phylogeny of Taylor and Qiu (2017). Sequences were aligned to the original alignment with Clustal Omega version 1.2.4 (Sievers et al., 2014) and placed in the original phylogeny using RAxML-EPA version 8.2.10 (Stamatakis, 2014).

RESULTS AND DISCUSSION

Illumina Sequencing, *de novo* Transcriptome Assembly and Annotation

After Illumina sequencing and quality filtering stages, reads were combined in a 183.4 Mb, N50 = 1,850 with an average contig length of 1,026 nt (*D. glomerata*; **Supplementary Figure S1A**) and in a 105.0 Mb, N50 = 1,275 with an average contig length of 715, 48 nt (*C. thyrsiflorus*; **Supplementary Figure S1B**) Trinity assembly (Grabherr et al., 2011). The resulting assembly contained 95,749 Trinity "genes" from a total of 164,856 isoforms (*D. glomerata*) and 97,521 Trinity "genes" from a total of 135,576 isoforms (*C. thyrsiflorus*).

Quality and completeness evaluation of the assemblies showed that 1130 [779 singletons and 351 duplicates] (*D. glomerata*) and 1103 [838 singletons and 265 duplicates] (*C. thyrsiflorus*) complete genes were assessed from a total of 1440 groups searched. Yet, 170 genes were fragmented and 140 were missing (*D. glomerata*); 135 genes were fragmented and 202 were missing (*C. thyrsiflorus*). In terms of expected orthology 90% (*D. glomerata*) and 86% (*C. thyrsiflorus*) of the transcripts could be assigned to members of the BUSCO plant set (Simão et al., 2015).

In total, 35,815 *D. glomerata* (37%) and 49,142 *C. thyrsiflorus* (50%) unique Trinity "genes" were annotated with SwissProt identifiers. *Arabidopsis thaliana* provided the best scoring alignment for both species (**Supplementary Figures S1C,D**). Of the top 100 most abundant transcripts annotated by PFAM within the "biological process" category, 44% showed shared functions between species (*D. glomerata*, **Supplementary Figure S2**; *C. thyrsiflorus*, **Supplementary Figure S3**; for numeric detail refer to doi: 10.17045/sthlmuni.6181772.v1).

GenBank accession numbers of sequences of transcripts analyzed in this study, as well as E.C. numbers of encoded enzymes, are listed in **Supplementary Table S5**; all *p*-values for RT-qPCR analyses are given in **Supplementary Table S6**.

Genes Encoding LysM Receptors Associated With Symbiotic Signaling in Model Legumes Are Upregulated in Nodules of *D. glomerata* and *C. thyrsiflorus*

Sequences of nodulation-related genes identified based on legume mutants were used to identify putative orthologs in the nodule transcriptomes of both species. The phylogeny of LysM-type receptors and LysM receptor kinases, corresponding to NFR1 and NFR5 of the model legume *L. japonicus* (Radutoiu et al., 2003) supports the assignment of orthology (**Figure 1**). Reverse transcription – quantitative PCR (RT-qPCR) of these orthologs showed that in both *D. glomerata* and *C. thyrsiflorus,* these genes were expressed at significantly higher levels in nodules compared to roots (*p* < 0.01; **Figure 2**). While caution must be taken while interpreting these results, a role for these orthologs in an actinorhizal symbiosis with a Nod factor-producing *Frankia* strain may be suggested. Interestingly,

in *C. glauca* (Casuarinaceae, Fagales), which is nodulated by a *Frankia* strain that does not contain the canonical *nod* genes, expression levels of the putative NFR1 ortholog do not differ between roots and nodules (Hocher et al., 2011). However, the expression of genes encoding putative Nod factor receptors was recently found to be induced in nodules of *Parasponia andersonii* compared to roots (van Velzen et al., 2018). It must be pointed out that since the *P. andersonii LjNFR5/MtNFP* ortholog is considered to have been recruited from arbuscular mycorrhizal (AM) symbiosis (Op den Camp et al., 2011; Miyata et al., 2016), these receptors might have additional functions that are responsible for the induction of their expression in nodules. Furthermore, transcriptional induction in nodules compared to roots was not found for the legume Nod factor receptors; expression of *L. japonicus NFR5* and of *Medicago truncatula NFP* is root specific, and expression levels of *L. japonicus NFR1* and its ortholog *M. truncatula LYK3* are similar in roots and nodules (Amor et al., 2003; Radutoiu et al., 2003; Smit et al., 2007). Hence, *D. glomerata* and *C. thyrsiflorus* orthologs may play a slightly different role in actinorhizal nodules. Still, differences in cell-specific distribution of Nod factor receptors between legumes and actinorhizal plants would not be surprising since the sizes of the corresponding gene families and degree of redundancy might conceivably differ between plant species. Altogether, the similarities with *Parasponia* regarding the differential expression of these genes, makes it very tempting to speculate about common signaling networks between non-legume hosts and Nod factor-producing bacteria (van Velzen et al., 2018). However, studies on gene knockouts are required to provide a definite answer.

Genes Encoding Proteins Linked to Infection Thread Formation

In model legumes, several genes have been shown to encode proteins required for infection thread formation: a LysM receptor kinase, EPR3, that has been proposed to be involved in exopolysaccharide perception (Kawaharada et al., 2015), a flotillin that has been linked to infection thread growth (Haney and Long, 2010), as well as vapyrin (VPY; Murray et al., 2011) and LIN (Guan et al., 2013). These proteins are also required for AM symbioses. RT-qPCR analyses have shown that with the exception of LIN (Demina et al., 2013), the expression of the *D. glomerata* homologs of these genes was significantly enhanced in nodules compared to roots (for direct comparison with the *EPR3* and *flotillin* genes, the analysis of *VPY* and *LIN* was repeated here and included in **Figure 2A**). The corresponding homologs from *C. thyrsiflorus* were now analyzed by RT-qPCR; with the exception of VPY (*p* < 0.23), all these genes showed significant induction in nodules compared to roots (**Figure 2**). These data suggest that the corresponding proteins are required for infection thread growth in different nodulating lineages.

Expression of the putative *EPR3* ortholog genes was induced in nodules of both *D. glomerata* and *C. thyrsiflorus* (**Figures 1, 2A,B**). In legumes, the LysM receptor kinase EPR3 has been suggested to distinguish between compatible and incompatible rhizobial surfaces in order to abort infection
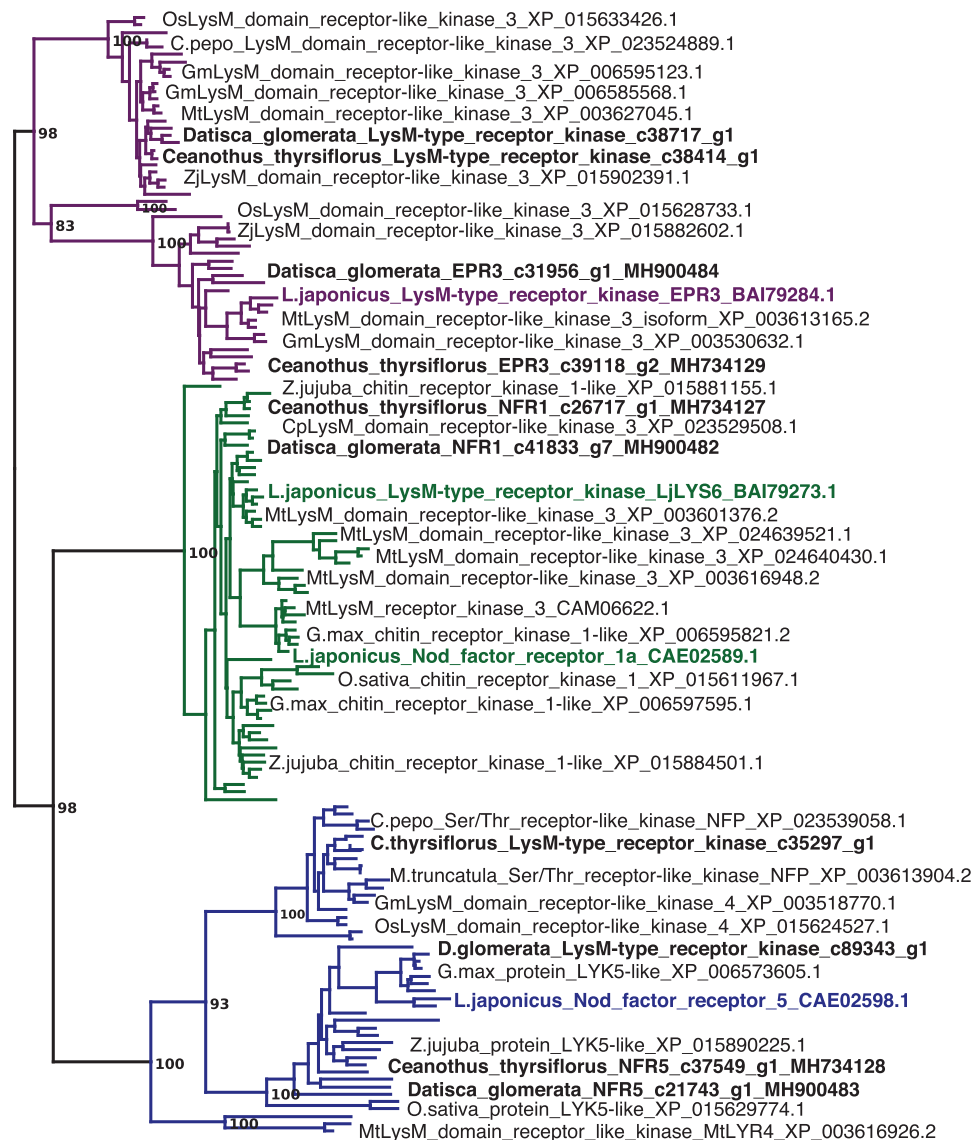
FIGURE 1 | *Datisca glomerata* and *Ceanothus thyrsiflorus* nodules express putative orthologs of Nod factor receptors from legumes. Part of a maximum-likelihood phylogenetic reconstruction of LysM-type and LysM receptor kinases is shown (the full phylogenetic tree is shown in **Supplementary Figure S10**). Proteins encoded by genes identified based on studies of legume mutants are highlighted in color. Acronyms of included species are: Cp, *Cucurbita pepo*; Gm, *Glycine max*; Lj, *Lotus japonicus*; Mt, *Medicago truncatula*; Os, *Oryza sativa*; and Zj, *Ziziphus jujuba*. Different orthogroups are distinguished by color: EPR3 in purple, NFR1 in green, and NFR5 in blue. Candidate orthologs from *D. glomerata* and *C. thyrsiflorus* are given in black bold print. GenBank accessions are given.

threads containing incompatible rhizobia (Kawaharada et al., 2015). While the *EPR3* gene is not expressed in mature determinate nodules of *L. japonicus,* expression takes place in immature, i.e., non-nitrogen fixing nodules which would correlate with expression in the infection zone of indeterminate nodules (Kawaharada et al., 2017). Since actinorhizal nodules are indeterminate structures with a developmental gradient of infected cells, including a zone of infection, in the cortex (Ribeiro et al., 1995), nodule-enhanced expression of genes encoding proteins required in the infection zone would be expected. Yet, based on the results available so far, it is impossible to conclude whether the EPR3 orthologs of *D. glomerata* and *C. thyrsiflorus*

are involved in *Frankia* surface recognition, or act as additional Nod factor receptors.

Similarly, the expression of the *flotillin* gene was induced in nodules compared to roots in both *D. glomerata* and *C. thyrsiflorus*. Flotillins have been linked to endocytosis and membrane shaping; they are targeted to membrane microdomains and have been shown to be involved in infection thread growth in legume nodules (Haney and Long, 2010). So this feature seems to be common in legume and actinorhizal nodules.

Vapyrin is required for epidermal penetration and infection thread development in AM symbioses (Pumplin et al., 2010). In legumes, vapyrin has been linked to the intracellular progression
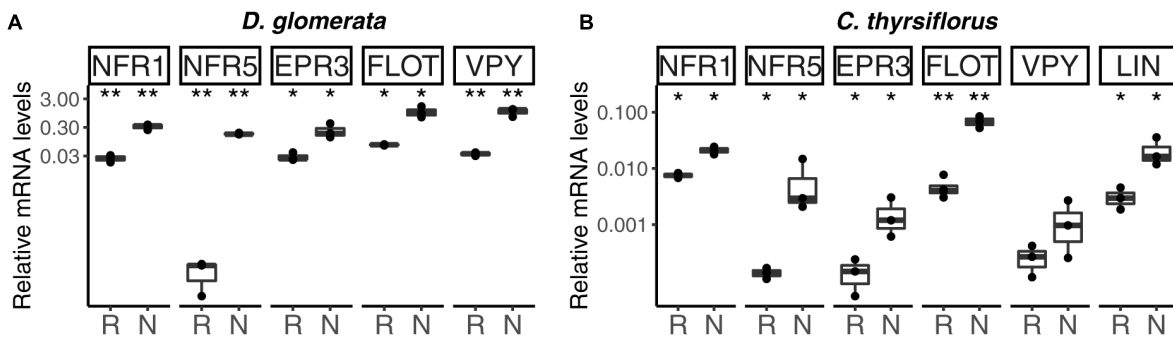
FIGURE 2 | Expression profile of genes encoding orthologs of Nod factor receptors, and of genes encoding homologs of proteins linked to infection thread growth in legumes, in roots and nodules of *D. glomerata* **(A)** and *C. thyrsiflorus* **(B)**. Gene expression levels in roots (R) and nodules (N) are given relative to those of EF-1α. The median and interquartile range (IQR) of at least three biological replicates are shown. Differences between R and N are indicated at *******p* < 0.01, **p* < 0.05 after student's *t*-test followed by FDR multi comparison correction. *Y*-axis is given in log$_{10}$ scale.

of infection threads and, consistent with this function, expression levels of *VPY* were found to be higher in nodules than in roots (Murray et al., 2011). Also in *D. glomerata*, *VPY* expression was induced in nodules (Demina et al., 2013). However, *VPY* expression was not induced in nodules of *C. thyrsiflorus*. This could be related to the fact that *C. thyrsiflorus* is infected intercellularly. *Frankia* hyphae colonize the apoplast and eventually are stably intracellularly accommodated in branching infection threads in infected cells. However, these infection threads do not show transcellular growth, in that infected cells are always infected from the apoplast, not from infection threads coming from an older infected cell (reviewed by Pawlowski and Demchenko, 2012). So it is possible that in actinorhizal symbioses, VPY is required only for transcellular growth of infection threads.

CERBERUS/LIN, an U-box protein containing WD40 repeats, is required for legume nodulation (Yano et al., 2009; Guan et al., 2013); for the *L. japonicus* ortholog (CERBERUS) also an effect on AM symbioses was shown. In *L. japonicus* mutants lacking this gene, the rhizobial infection process is aborted at a very early stage of infection thread formation, and elongation of intraradical hyphae of AM fungi is reduced (Takeda et al., 2013). While the expression of the corresponding gene is enhanced in nodules compared to roots of legumes, actinorhizal Fagales (*C. glauca*; Hocher et al., 2011) and of *C. thyrsiflorus* as representative of actinorhizal Rosales, this is not the case for *D. glomerata* (Datiscaceae, Cucurbitales; Demina et al., 2013). This result supports the hypothesis that a unique infection thread growth mechanism was established in actinorhizal Cucurbitales (Pawlowski and Demchenko, 2012).

Genes Encoding Enzymes Involved in Arginine Metabolism Show Differential Expression Profiles in Nodules and Roots of *D. glomerata* and *C. thyrsiflorus*

Nodules of actinorhizal Cucurbitales have an unusual morphology in that infected cortical cells are not interspersed

with uninfected cortical cells, but form a continuous patch, kidney-shaped in cross section, on one side of the acentric stele (Pawlowski and Demchenko, 2012). *D. glomerata* nodules have an unusual nitrogen assimilatory metabolism (Berry et al., 2011). Generally, in root nodule symbioses examined to date, the microsymbionts, whether rhizobia or Frankia, export the product of nitrogen fixation, ammonia, which is directly assimilated in the cytosol of the host infected cells via the glutamine synthetase (GS)/glutamate synthase (GOGAT) cycle (**Supplementary Figure S4**; Patriarca et al., 2002). In *D. glomerata,* however, cytosolic GS is expressed at high levels in the uninfected cells surrounding the patch of infected cells, but not in the infected tissue (Berry et al., 2004). Based on bacterial transcriptomics (Persson et al., 2015) and on the N metabolome of nodules (Berry et al., 2004), it was postulated that in *D. glomerata* nodules, symbiotic *Frankia* assimilates the ammonium produced by nitrogenase via bacterial GS/GOGAT, and exports an intermediate N storage product of the arginine cycle, presumably arginine, which is then broken down in the uninfected cells, leading to the re-assimilation of ammonium via the GS/GOGAT cycle, and, as demonstrated (Berry et al., 2004), the export of glutamine and glutamate from the nodule to the xylem stream.

In our analysis, the genes encoding GS and GOGAT were upregulated in nodules vs. roots of *D. glomerata* (**Supplementary Figure S4C**), although the nodule/root ratio for GS expression was not significant due to the wide variation displayed between samples (*p* < 0.07). Expression levels of the gene encoding asparagine synthase (ASN1) showed a trend toward lower expression in nodules compared to roots (*p* < 0.11; **Figure 3A**). Pathways for arginine biosynthesis and degradation (Slocum, 2005; Winter et al., 2015) are depicted in **Figure 3** and **Supplementary Figure S5**. Transcripts encoding several enzymes involved in these pathways were identified in both species and their expression was analyzed using RT-qPCR. Strikingly, a gene encoding an arginase homolog (ARGH1) which catalyzes the breakdown of arginine to urea and ornithine, was upregulated to very high levels in nodules of *D. glomerata* compared to roots (*p* < 0.027; **Figure 3A**). Furthermore, a gene encoding the
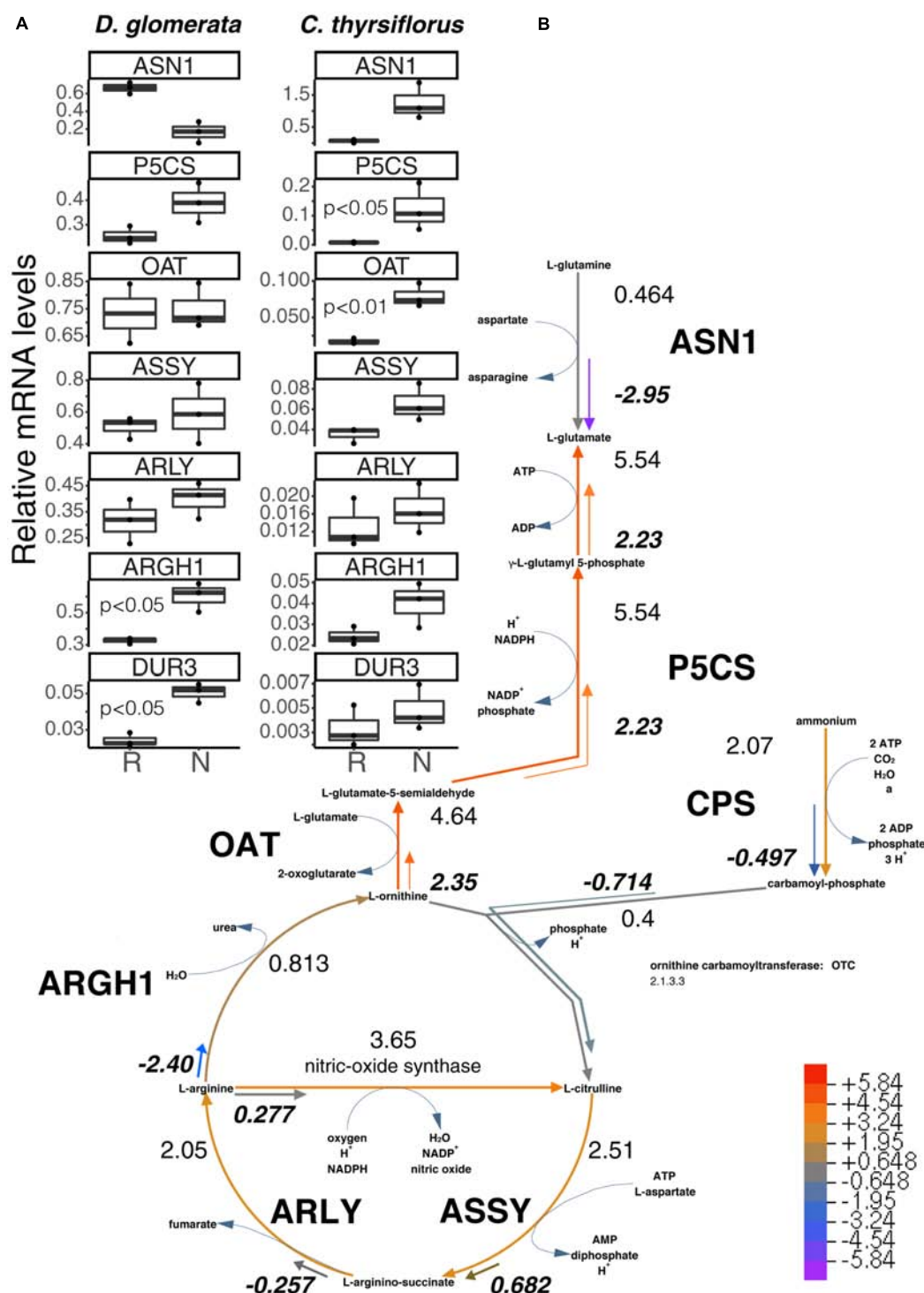
FIGURE 3 | Superpathway for citrulline biosynthesis and link to the urea cycle. **(A)** Gene expression levels in roots (R) and nodules (N) were quantified by RT-qPCR for *D. glomerata* and *C. thyrsiflorus* and are relative to those of the housekeeping gene EF-1α. Median and IQR of three biological replicates are shown. Significant differences between R and N are indicated by *p*-values based on student's *t*-test with FDR multi comparisons correction. *Y*-axis is given in log10 scale. **(B)** Gene expression levels are given in the context of the pathway, calculated as the log$_2$[FC] relative to those of EF-1α estimated from the mean TPM of five (*D. glomerata*, long arrows) and three (*C. thyrsiflorus,* short arrows with values written in bold italics) independent RNA preparations. Explanatory heatmap is provided. Enzymes catalyzing these steps include: ASN1, asparagine synthase; P5CS, glutamate-5-semialdehyde dehydrogenase and glutamate-5-kinase (double function); CPS, ammonia-dependent carbamoyl phosphate synthetase; ASSY, argininosuccinate synthase; ARLY, argininosuccinate lyase; ARGH1, arginase; and OAT, ornithine aminotransferase. DUR3, high affinity plasma membrane urea transporter.

homolog of the high affinity plasma membrane urea transporter DUR3 (Kojima et al., 2007) was identified and found to be induced in nodules compared to roots of *D. glomerata* ($p < 0.05$). Numbers of reads of urease transcripts were low, indicating that this function was not controlled on the transcriptional level (data not shown).

Genes encoding enzymes involved in the degradation of ornithine (**Figure 3** and **Supplementary Figure S5**) were examined as well. A candidate for the mitochondrial isoform of *N*-acetylornithine aminotransferase (NAOAT), a close homolog of the aminotransferase class-III participating in the biosynthesis of citrulline in the actinorhizal tree *A. glutinosa* (Guan et al., 1997), was downregulated in *D. glomerata* nodules compared to roots ($p < 0.05$). However, no significant differences were found between the expression levels of the gene encoding the enzyme that catalyzes the transformation of L-ornithine to L-glutamate-5semialdehyde (ornithine aminotransferase, OAT; $p < 0.55$; **Figure 3A**) or for the gene encoding glutamate-5-semialdehyde dehydrogenase (P5CS; $p < 0.074$; **Figure 3A**).

Thus, the transcriptome data showed evidence for upregulation of arginine degradation in *D. glomerata* nodules. On the other hand, no evidence was found for the upregulation of arginine biosynthesis: the expression levels of the genes for argininosuccinate synthase (ASSY) and argininosuccinate lyase (ARLY) were similar in roots and nodules (**Figure 3A**).

In summary, the transcriptome data strongly support the role of arginine as intermediary nitrogen storage compound in root nodules of *D. glomerata* that is exported from *Frankia* and broken down by the plant in the uninfected cells. One breakdown product, urea, would be degraded to ammonium and CO_2, with ammonium being reassimilated via the GS/GOGAT pathway, while the breakdown product ornithine could be used to form 2-oxoglutarate or another dicarboxylate, since carbon skeletons for ammonium assimilation have to be provided to *Frankia*. The upregulation of the glutamine-dependent carbamoyl phosphate synthetase (CPS; **Supplementary Figure S5**) in nodules compared to roots indicates that some of the reassimilated ammonium likely would go into plastidic arginine biosynthesis, either to replenish the arginine pool for protein biosynthesis, or as xylem transport form. The latter would be consistent with the observation that, in addition to the major xylem exports (glutamine and glutamate), low levels of arginine were detected in roots of nodulated plants of *D. glomerata* (Persson et al., 2016).

In *C. thyrsiflorus,* a different nitrogen assimilatory pattern emerges: expression of plant genes encoding GS and GOGAT was significantly induced in nodules vs. roots (**Supplementary Figure S4C**). Expression levels of the gene encoding asparagine synthase (ASN1) showed a trend toward elevated expression in nodules ($p < 0.07$), consistent with the fact that asparagine was reported as the major nodule amino acid in *Ceanothus* sp. (Schubert, 1986). In contrast with *D. glomerata,* expression of ARGH1 encoding arginase was not induced in nodules compared to roots ($p < 0.11$; **Figure 3A**), nor was expression of the gene encoding the high affinity plasma membrane urea transporter DUR3 ($p < 0.36$; **Figure 3A**). With regard to arginine biosynthesis, the gene encoding the glutamine-dependent CPS

showed similar expression levels in nodules compared to roots ($p < 0.93$; **Supplementary Figure S5**), ASSY transcription was slightly elevated in nodules ($p < 0.07$), while ARLY showed similar expression levels in both organs ($p < 0.45$; **Figure 3A**). NAOAT showed a slight induction in nodules compared to roots ($p < 0.056$; **Supplementary Figure S5C**). By contrast, expression levels of OAT ($p < 0.01$; **Figure 3A**) and P5CS ($p < 0.05$; **Figure 3A**) were highly induced in *C. thyrsiflorus* nodules compared to roots (**Figure 3A**), suggesting the possibility of proline biosynthesis.

In summary, transcriptional analysis shows that plant N metabolism in nodules of *C. thyrsiflorus* (Rosales) is different from that in nodules of *D. glomerata*, even though both hosts are nodulated by closely related Cluster II *Frankia* strains. Thus, the specialization of N metabolism that occurs in *D. glomerata* is not necessitated by the capabilities of the microsymbiont, but shows the flexibility of metabolic adaptations possible in root nodules, in both the host and the microsymbiont.

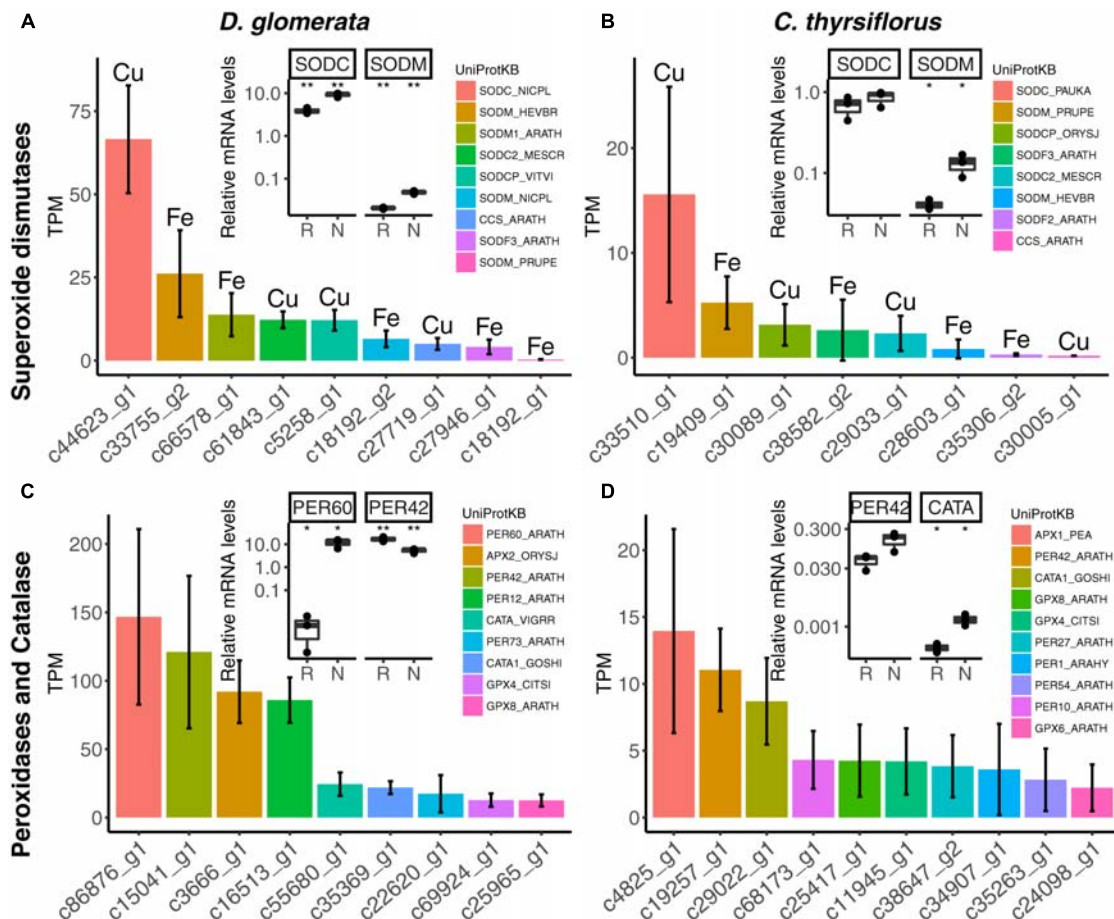## Hemoglobins in Nodules of *D. glomerata* and *C. thyrsiflorus*

No class II hemoglobin transcripts were identified in either transcriptome, suggesting that in neither system the buffering of free oxygen was required in the plant cytosol. Yet, in *Parasponia*, this function is performed by a class I hemoglobin (Sturms et al., 2010). In all cases, genes encoding the oxygen-buffering hemoglobin showed more than 1000-fold induction in nodules compared to roots.

Altogether, three hemoglobins were identified in the transcriptome of *D. glomerata*. Induction of the previously published (Pawlowski et al., 2007) truncated globin gene *trHB1* in nodules compared to roots was confirmed ($p < 0.01$; **Figure 4A**). Expression levels of a class I hemoglobin gene (Hb1-1) were also analyzed by RT-qPCR and shown to be similar in roots and nodules (**Figure 4A**). However, expression levels of a second class I hemoglobin gene (Hb1-2) were enhanced in nodules compared to roots ($p < 0.05$;
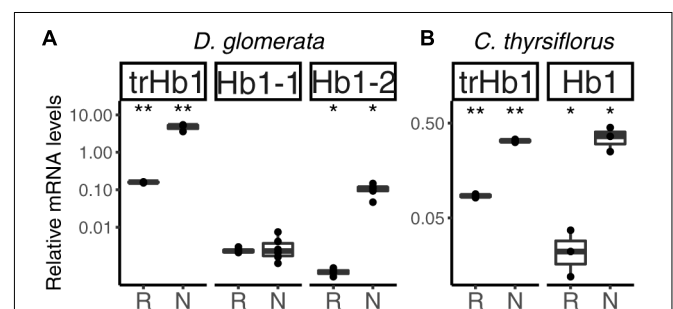
FIGURE 4 | Hemoglobins in actinorhizal nodules of *D. glomerata* **(A)** and *C. thyrsiflorus* **(B)**. Transcript abundance was quantified by RT-qPCR. Gene expression levels in roots (R) and nodules (N) are relative to those of EF-1α. The median and IQR of at least three biological replicates are shown. Asterisks denote differences between R and N at $^{**}p < 0.01$, $^{*}p < 0.05$ by student's t-test followed by FDR multi comparison correction. Y-axis is given in $\log_{10}$ scale.

FIGURE 5 | Expression of genes encoding superoxide dismutases, peroxidases and catalases in nodules of *D. glomerata* (left) and *C. thyrsiflorus* (right). Expression levels of superoxide dismutases **(A,B)**, peroxidases and catalases **(C,D)** are shown. Expression levels were estimated based on the mean TPM of five (*D. glomerata*) and three (*C. thyrsiflorus*) independent libraries. Bars were colored according to their annotated UniProt identifiers (see legend) and the cofactor used by each enzyme is given for **(A,B)**. Error bars show standard deviations. The relative abundance of mRNA in roots (R) and nodules (N) was quantified by RT-qPCR and is relative to that of the housekeeping gene EF-1α. The median and IQR of three biological replicates are shown. Asterisks highlight differences between organs at ******$p < 0.01$, *****$p < 0.05$ by student's *t*-test followed by FDR multi comparison correction. *Y*-axis is given in $\log_{10}$ scale. UniProtKB homologs are indicated. After doi: 10.17045/sthlmuni.6181766.v1.

Figure 4A). In a similar manner, the nodule transcriptome of *C. thyrsiflorus* contained transcripts of a truncated globin gene *trHB1*, whose expression levels were also markedly enhanced in nodules compared to roots ($p < 0.01$; **Figure 4B**). Expression of a class I hemoglobin gene was induced in nodules compared to roots ($p < 0.05$; *Hb1*; **Figure 4B**). In both species, the fold change of the class I hemoglobin gene in nodules was far below those found for oxygen-buffering hemoglobins.

In summary, among the set of genes encoding globins, only those associated with NO detoxification (one *trHb1*, one *Hb1* per species) were induced in nodules compared to roots in both systems. These results indicate that in both *D. glomerata* and *C. thyrsiflorus*, consistent with observations in actinorhizal Fagales (Heckmann et al., 2006; Sasakura et al., 2006), high activity toward NO detoxification is expected. However, despite the fact that hypoxia has also been previously associated

with NO production in the model legume *M. truncatula* (Baudouin et al., 2006), and hypoxic conditions may prevail in the inner parts of infected cells of *D. glomerata* (Silvester et al., 1999), there is no evidence for differences in levels of nitrosative stress in nodules of *D. glomerata* vs. *C. thyrsiflorus* so far.

Under normoxic conditions, NO, an inhibitor of cytochrome c oxidase, can be oxidized to nitrite (Gupta and Igamberdiev, 2011). Nitrite is toxic and therefore needs to be rapidly catalyzed to ammonium by nitrite reductase NIR (**Supplementary Figure S4B**). Expression levels of *NIR* were repressed in nodules of *D. glomerata* ($p < 0.02$), but induced ($p < 0.04$) in those of *C. thyrsiflorus* (**Supplementary Figure S4C**). This difference could be explained by the hypothesis that oxidation of NO does not take place in infected cells of *D. glomerata* nodules since the high amounts of mitochondria (Silvester et al., 1999)
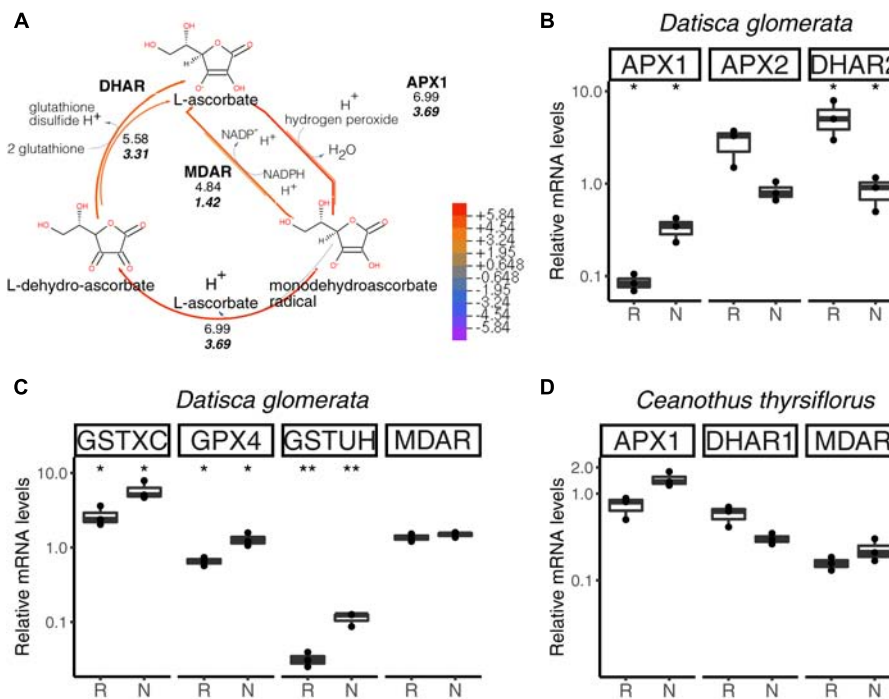
FIGURE 6 | Hydrogen peroxide (H_2O_2) detoxification *via* the ascorbate/glutathione cycle. H_2O_2 is generated from the superoxide radical by superoxide dismutase (SOD; not shown). L-ascorbate serves as electron donor *via* ascorbate peroxidase (APX1, APX2). GPX4, glutathione peroxidase; GSTXC and GSTUH, glutathione *S*-transferases; MDAR, monodehydroascorbate reductase; DHAR, dehydroascorbate reductase. **(A)** Expression levels represented in the cycle were calculated as the $\log_2$[FC] relative to those of EF-1α estimated from the mean TPM of five (*D. glomerata*, long arrows) and three (*C. thyrsiflorus,* short arrows with values written in bold italics) independent libraries. Explanatory heatmap is provided. **(B–D)** Transcript abundance quantified by RT-qPCR for candidate genes in nodules of *D. glomerata* **(B,C)** and *C. thyrsiflorus* **(D)** is shown. Gene expression levels in roots (R) and nodules (N) are relative to those of the housekeeping control EF-1α. The median and IQR of three biological replicates are shown. Differences between R and N are highlighted at **$p < 0.01$, *$p < 0.05$ after student's *t*-test with FDR multi comparison correction. *Y*-axis is given in $\log_{10}$ scale. Species names and UniProtKB homologs are indicated. After doi: 10.17045/sthlmuni.6181766.v1.

lead to a reduction of the oxygen tension below normoxic conditions.

Defense Against Reactive Oxygen Species (ROS) in Nodules of *D. glomerata* and *C. thyrsiflorus*

In many types of nitrogen-fixing root nodules, locally microaerobic conditions are established to protect the oxygen-sensitive nitrogenase enzyme complex (Jacobsen-Lyon et al., 1995). These conditions contribute to the production of reactive oxygen species (ROS; Fukao and Bailey-Serres, 2004). ROS can cause oxidative damage and therefore need to be quickly detoxified, a process that involves superoxide dismutase (SOD), peroxidases (PER) including ascorbate peroxidase (APX), and catalase (CATA).

The nodule transcriptomes contained transcripts representing a number of *SOD* (**Figures 5A,B**) and *PER/CATA* genes (**Figures 5C,D**; listed in doi: 10.17045/sthlmuni.6181766.v1). Members displaying the highest TPM in each class were assayed by RT-qPCR. In *D. glomerata*, expression levels of two out of three *SOD* genes – encoding a copper-dependent SOD (SODC) and a manganese-dependent

SOD (SODM) – were enhanced in nodules compared to roots ($p < 0.01$; **Figure 5A**). *PERs* displaying the highest transcriptome abundance (*PER60* and *PER42)* showed alternative regulation patterns based on RT-qPCR analysis: expression of *PER60* was strongly enhanced in nodules compared to roots, while *PER42* was repressed ($p < 0.05$ for both; **Figure 5C**).

For *C. thyrsiflorus*, the sub-classes of *SOD* genes with the highest transcriptome abundance were also encoding an SODC and an SODM (**Figure 5B**); only the expression of *SODM* was significantly induced in nodules compared to roots ($p < 0.01$; **Figure 5B**). Like in *D. glomerata,* *PER42* represented the most abundant *PER*; however, in *C. thyrsiflorus* its expression levels were not significantly increased in nodules compared to roots ($p < 0.076$; **Figure 5D**). On the other hand, albeit being expressed at relative extreme low levels, *CATA* was induced in nodules compared to roots ($p < 0.05$; **Figure 5D**).

The main pathway for ROS scavenging in plants is the ascorbate-glutathione cycle (**Figure 6A**). In *D. glomerata*, expression assessment performed by RT-qPCR for members of this pathway showed that an L-ascorbate peroxidase 1 gene (*APX1*; **Figure 6B**), two glutathione *S*-transferase genes (*GSTXC* and *GSTUH*; **Figure 6C**), and a glutathione peroxidase gene

(*GPX4*; **Figure 6C**) were induced in nodules compared to roots (*p* < 0.05). Expression of an *APX2* gene, the L-ascorbate peroxidase gene displaying the highest TPM, was slightly enhanced in roots compared to nodules (*p* = 0.054; **Figure 6B**); while the monodehydroascorbate reductase gene *MDAR* was expressed at similar levels in roots and nodules (**Figure 6C**).

For *C. thyrsiflorus*, expression levels of *APX1*, *DHAR1*, and *MDAR* were analyzed in roots vs. nodules. Although no significant differences were found, *APX1* expression was slightly enhanced in nodules (*p* = 0.060), while *DHAR1* expression was slightly enhanced in roots (*p* = 0.066) (**Figure 6D**). Additionally, the levels of three glutathione peroxidases genes (encoding the strongest homologs of GPX4, GPX6, and GPX8) were analyzed. Unlike in *D. glomerata*, all these genes were expressed at similar levels in roots and nodules (data not shown).

In summary, when comparing gene expression levels for components of the ascorbate/glutathione cycle, no significant differences were found between roots and nodules of *C. thyrsiflorus*; whilst, in *D. glomerata*, some genes encoding enzymes of these pathways were induced in nodules compared to roots. Although caution must be exercised while interpreting transcriptome data, these results suggest that ROS stress might be higher in nodules of *D. glomerata* than in those of *C. thyrsiflorus*. This is consistent with the assumption that in *D. glomerata* the plant participates in the protection of nitrogenase from oxygen, creating microaerobic conditions *via* a blanket of mitochondria (Silvester et al., 1999), and these conditions lead to enhanced ROS production (Fukao and Bailey-Serres, 2004). In *C. thyrsiflorus,* on the other hand, nitrogenase seems to be protected from oxygen due to the thick envelope surrounding spherical *Frankia* vesicles (Strand and Laetsch, 1977), similar to the situation in *Alnus* nodules (Kleemann et al., 1994), while normoxic conditions prevail in the plant cell. Thus, the induction of *NIR* expression in *C. thyrsiflorus* nodules must be due to nitrosative, and not oxidative stress.

Key Genes Involved in the Biosynthesis of the Vitamins Thiamine and Folic Acid Are Differentially Expressed in Roots and Nodules of *D. glomerata* and *C. thyrsiflorus*

Transcripts of genes encoding members of the thiamine biosynthetic pathway were identified in nodules of *D. glomerata* (**Supplementary Figure S6**, long arrows) and in those of *C. thyrsiflorus* (**Supplementary Figure S6**, short arrows). Expression levels of the key genes *THIC* and *THI4* were quantified by RT-qPCR in roots and nodules of both species (**Figure 7A**). *THIC* showed strong induction in nodules compared to roots in both species (*p* < 0.01). Expression levels of THI4, the homolog of which is upregulated in nodules of *A. glutinosa* compared to roots (Ribeiro et al., 1996), showed slight upregulation in nodules of *D. glomerata* (*p* < 0.053), but not in those of *C. thyrsiflorus* (*p* < 0.1; **Figure 7A**).

Thus, like nodules of legumes and actinorhizal Fagales, nodules of actinorhizal Cucurbitales and Rosales require high

amounts of thiamine, possibly due to its function in the oxidative stress response (Tunc-Ozdemir et al., 2009). Recently, it was shown that the AM fungus *Rhizophagus irregularis* lacks the toolkit for thiamine biosynthesis (Tisserant et al., 2013); spores of *R. irregularis* accumulated higher levels of thiamine than roots of *L. japonicus* during AM symbiosis (Nagae et al., 2016b). However, it is unlikely that symbiotic auxotrophy of symbiotic nitrogen-fixers, analogous for the symbiotic auxotrophy of rhizobia for branched chain amino acids (Prell et al., 2009), is the reason behind the upregulation
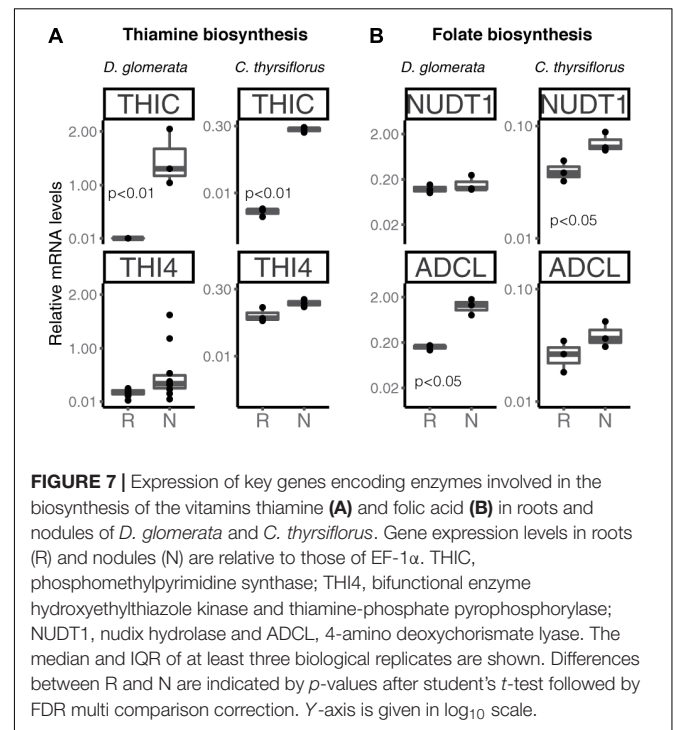

FIGURE 7 | Expression of key genes encoding enzymes involved in the biosynthesis of the vitamins thiamine **(A)** and folic acid **(B)** in roots and nodules of *D. glomerata* and *C. thyrsiflorus*. Gene expression levels in roots (R) and nodules (N) are relative to those of EF-1α. THIC, phosphomethylpyrimidine synthase; THI4, bifunctional enzyme hydroxyethylthiazole kinase and thiamine-phosphate pyrophosphorylase; NUDT1, nudix hydrolase and ADCL, 4-amino deoxychorismate lyase. The median and IQR of at least three biological replicates are shown. Differences between R and N are indicated by *p*-values after student's *t*-test followed by FDR multi comparison correction. *Y*-axis is given in $\log_{10}$ scale.
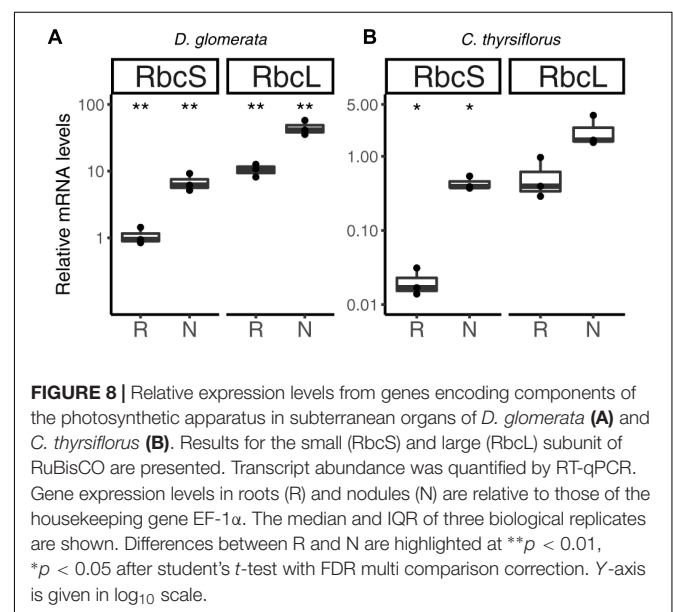
FIGURE 8 | Relative expression levels from genes encoding components of the photosynthetic apparatus in subterranean organs of *D. glomerata* **(A)** and *C. thyrsiflorus* **(B)**. Results for the small (RbcS) and large (RbcL) subunit of RuBisCO are presented. Transcript abundance was quantified by RT-qPCR. Gene expression levels in roots (R) and nodules (N) are relative to those of the housekeeping gene EF-1α. The median and IQR of three biological replicates are shown. Differences between R and N are highlighted at *******p* < 0.01, *p < 0.05 after student's *t*-test with FDR multi comparison correction. *Y*-axis is given in $\log_{10}$ scale.
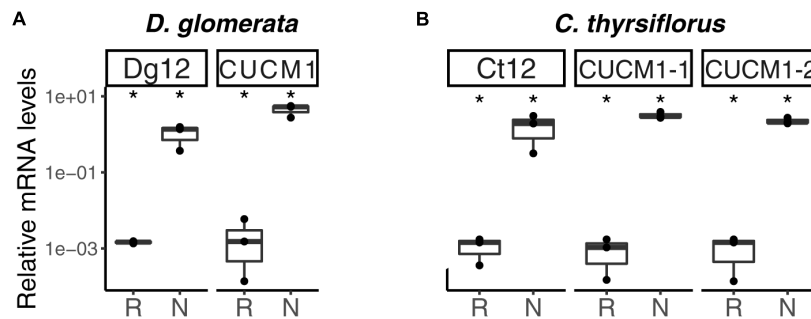
FIGURE 9 | Subtilisin-type peptidases belonging to the S8 family are expressed in nodules of *D. glomerata* **(A)** and *C. thyrsiflorus* **(B)**. The transcript abundance in roots (R) and nodules (N) was quantified by RT-qPCR relative to that of the housekeeping gene EF-1α. The median and IQR of three biological replicates are shown. Differences between R and N are highlighted at *$p < 0.05$ after student's *t*-test with FDR multi comparison correction. Y-axis is given in $\log_{10}$ scale.

of thiamine biosynthesis-related genes in nodules. No changes in gene expression levels were observed for *THIC in planta* vs. in N-replete cultures for the *A. glutinosa*-infective *Frankia* strain ACN14a (Alloisio et al., 2010); so there is no reason to assume that *Frankia* thiamine biosynthesis is shut down in symbiosis.

Transcripts of all members of the folic acid (vitamin B9) biosynthesis pathway were identified in nodules of both plant species (Hanson and Gregory, 2011; **Supplementary Figure S7**); RT-qPCR was performed for two key genes encoding nudix hydrolase (NUDT1) and 4-amino deoxychorismate lyase (ADCL) (**Figure 7B**). These enzymes map to different branches of the pathway; it should be pointed out that NUDT1 may have another function as well (Ogawa et al., 2005). For *D. glomerata*, *ADCL* was significantly induced in nodules compared to roots (**Figure 7B**); for *C. thyrsiflorus*, *NUDT1* was induced in nodules compared to roots, while *ADCL* was not ($p = 0.2$) (**Figure 7B**). This is, to our knowledge, the first evidence for a possible role of folic acid in actinorhizal nodules; further studies are required to elucidate its function.

Photosynthesis-Related Genes Are Expressed in Nodules of *D. glomerata* and *C. thyrsiflorus*

Previous studies have shown that nitrogen-fixing nodules of *D. glomerata* express the gene encoding ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) activase, an enzyme commonly associated with photosynthesis, and also the gene encoding the large subunit of RuBisCO (Okubara et al., 1999). The latter was confirmed in this study using RT-qPCR; transcription of both the nuclear gene for the small (RbcS) and the plastidic gene for the large (RbcL) subunit was highly induced in nodules compared to roots ($p < 0.01$; **Figure 8A**).

Similarly, RuBisCO activase transcripts were found in the nodule transcriptome of *C. thyrsiflorus* (data not shown). RT-qPCR analysis showed that also here, *RbcS* transcription was induced in nodules compared to roots ($p < 0.05$); instead, expression levels of *RbcL* were not significantly elevated in nodules compared to roots ($p < 0.06$) (**Figure 8B**).

Plants express their photoreceptors – phytochromes, cryptochromes and phototropins – not only in their aerial parts, but also in roots (Yokawa and Baluška, 2015; van Gelderen et al., 2018). Roots show negative phototropism in response to direct light, and can also react to stem-piped light. Consequently, light responses in subterranean organs might not be surprising, but the expression of genes encoding RuBisCO subunits is. Nevertheless, a non–Calvin cycle, i.e., a CO_2-scavenging role for RuBisCO has been demonstrated for developing *Brassica napus* embryos (Schwender et al., 2004); so it is possible that RuBisCO performs a similar function in nodules.

Expression of Several Serine Proteases Is Strongly Induced in Nodules Compared to Roots in Both Species

Analysis of the *D. glomerata* nodule transcriptome revealed transcripts encoding 178 peptidases from 27 families. Subtilases represented 23.6% of this total, and from this group, transcripts encoding for members of the S8 family (subtilisin-like proteases) were the most abundant (42 contigs; **Supplementary Figure S8A**; Taylor and Qiu, 2017). Based on TPM data, the two S8 subtilases displaying the highest expression levels were selected for RT-qPCR analysis. Results showed high levels of induction in nodules compared to roots of the ortholog of *Ag12/Cg12/Dt12* (Ribeiro et al., 1995; Laplaze et al., 2000; Fournier et al., 2018), named *Dg12*, and of a gene encoding a cucumisin homolog (*CUCM1*; **Figure 9A**).
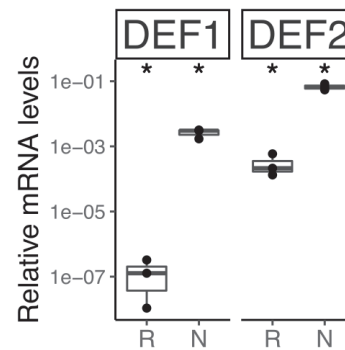
Global expression analysis of peptidases in *C. thyrsiflorus* nodules revealed major commonalities with that of *D. glomerata* nodules. Despite that, subtle differences could also be noticed. For instance, although fewer transcripts were annotated as peptidases (total = 147), they were actually more diverse occurring in 37 distinct families. In a similar manner, genes encoding serine proteases of the subtilisin S8 family were expressed at the highest levels (22 contigs, 15%; **Supplementary Figure S8B**). Those displaying the highest TPM values were selected for roots vs. nodules RT-qPCR analysis. In line with the results described for *D. glomerata,* expression of the
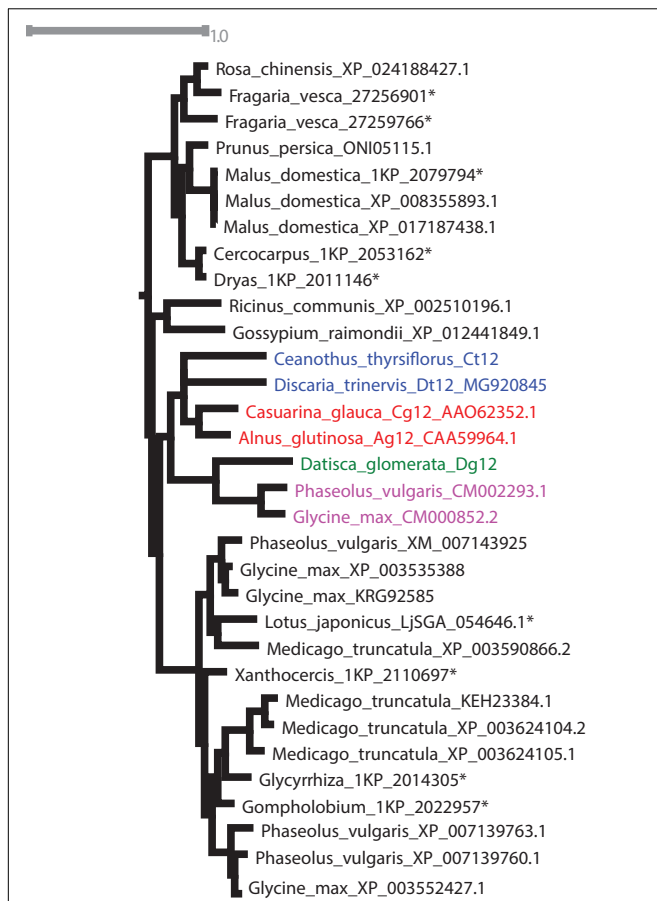
FIGURE 10 | Maximum-likelihood phylogenetic tree of subtilases. Candidates from *D. glomerata* and *C. thyrsiflorus* (in bold) were analyzed for phylogenetic placement using the comprehensive dataset of Taylor and Qiu, 2017 (see section "Materials and Methods"). When possible, GenBank accession numbers are provided after the species name. Otherwise, an asterisk (*) refers to the nomenclature used by Taylor and Qiu (2017). The analyzed candidates form a clade with proteins from Rosales (in blue), Dt12, *D. trinervis* (Fournier et al., 2018); Fagales (in red), Cg12, *Casuarina glauca* (Laplaze et al., 2000; Svistoonoff et al., 2003, 2004); Ag12, *A. glutinosa* (Ribeiro et al., 1995); and Fabales (in pink), CM002293.1 from *Phaseolus vulgaris* and CM000852.2 from *Glycine max*.

FIGURE 11 | Relative transcript abundance of genes encoding defensin-like peptides in belowground organs of *C. thyrsiflorus*. Expression levels in roots (R) and nodules (N) are relative to those of the housekeeping gene EF-1α. The median and IQR of three biological replicates areshown. Differences between R and N are indicated for *$p < 0.05$ after student's *t*-test followed by FDR multi comparison correction. Y-axis is given in $\log_{10}$ scale.

(Ribeiro et al., 1995; Svistoonoff et al., 2003, 2004), and with intercellular infection foci in intercellularly infected actinorhizal species (Fournier et al., 2018) seem to have a common origin. Further studies will be required to determine whether the nodule-specific cucumisins are a feature of actinorhizal Cucurbitales, or of actinorhizal symbioses in general.

Nodule-Specific Defensins: The Actinorhizal Equivalents of Legume Nodule Cysteine-Rich Peptides (NCR)

The transcriptome of *C. thyrsiflorus* was examined for genes encoding small cysteine-rich peptides. Two were identified and they represent class I defensins (Parisi et al., 2018). RT-qPCR analysis showed that both were strongly induced in nodules compared to roots (**Figure 11**).

While in legumes NCRs only occur in the IRLC clade and in the Dalbergioid lineage (Van de Velde et al., 2010; Czernic et al., 2015), in actinorhizal plants nodule-specific defensins have now been found across the three orders: Cucurbitales (Demina et al., 2013), Fagales (Carro et al., 2015), and Rosales (this study). Legume NCRs have been shown to affect bacteroid membrane permeability and induce polyploidy, turning the bacteroid state in a terminal differentiation (Alunni and Gourion, 2016). While the latter function has not yet been shown for actinorhizal defensins since analysis of polyploidy in a filamentous bacterium is a technical problem, *A. glutinosa* nodule-specific defensins have been demonstrated to affect membrane permeability of *Frankia alni* ACN14a for amino acids and to negatively affect viability of ACN14a *in vivo* (Carro et al., 2015). This might be interpreted to mean that in perennial nodulating plants (*D. glomerata* is biennial, and all other actinorhizal species are perennial) the defense against microbial "cheaters" is more urgent than in annual plants (most legumes analyzed in detail are annuals). Further studies are required to understand the distribution of antimicrobial peptides in actinorhizal vs. legume nodules.

homolog of *Ag12/Cg12/Dt12*, named *Ct12*, and of two cucumisin genes (*CUCM1-1*, transcript c29196_g1; *CUCM1-2*, transcript c34419_g1) was highly induced in nodules compared to roots (*p* < 0.05; **Figure 9B**). In addition, the expression levels of the genes encoding the homologs of the *Arabidopsis* subtilases AIR3 and SUBL were also analyzed, but they did not differ between roots and nodules (data not shown). These results are summarized for both species in doi: 10.17045/sthlmuni.6181760. v1.

Phylogenetic analysis confirmed the orthology of Ag12, Cg12, Dt12, Dg12, and Ct12 (**Figure 10**) and of the nodule-enhanced cucumisins of *D. glomerata* and *C. thyrsiflorus,* respectively (**Supplementary Figure S9**). In summary, the extracellular subtilases whose expression is correlated with infection thread formation in intracellularly infected actinorhizal species

CONCLUSION

All host plants examined thus far of *Frankia* strains containing the canonical *nod* genes, contain orthologs of legume Nod factor receptors the expression of which is induced in nodules compared to roots.

Analysis of transcript levels in roots vs. nodules of genes encoding enzymes from arginine metabolism indicates that while arginine is the likely form of nitrogen exported by *Frankia* in nodules of *D. glomerata*, it does not seem to play this role in nodules of *C. thyrsiflorus*.

The oxygen protection system for nitrogenase realized in nodules of *D. glomerata* seems to lead to greater oxidative stress than the system realized in nodules of *C. thyrsiflorus*. However, the production of high levels of nitric oxide under normoxic conditions seems to lead to nitrite production in *C. thyrsiflorus* nodules, as indicated by the induction of nitrite reductase expression.

Thiamine biosynthesis is induced in actinorhizal nodules of all three different orders. Folic acid biosynthesis so far was only found to be induced in nodules of *D. glomerata*.

Nodule-specific subtilisin-like proteases that are involved in infection in actinorhizal nodules, seem to have a common evolutionary origin.

Nodule-specific defensins are found in actinorhizal species from all three different orders.

DATA AVAILABILITY

The transcriptome raw data were submitted to GenBank as BioProjects (PRJNA454375 and PRJNA454377). Transcriptome Shotgun Assemblies were deposited at DDBJ/EMBL/GenBank under the accessions GGXR01000000 and GGXO01000000. All cDNA sequences analyzed by RT-qPCR were submitted to GenBank; a list of names and accession numbers is presented in **Supplementary Table S5**. The annotation lists are available on FigShare (references in the text).

AUTHOR CONTRIBUTIONS

MS and KP: conceptualization. MS, DL, TVN, RvV, and KP: methodology. MS, DL, RvV, and KB: investigation. MS, RvV, and DL: formal analysis. MS and DL: visualization. MS writing – original draft. MS, RvV, DL, KB, AB, and KP: writing – review and editing. KP: funding acquisition and supervision.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2018.01629/full#supplementary-material

FIGURE S1 | Statistics of Trinity assembly (upper panels) and Trinotate annotation (lower panels). Length distribution of assembled transcripts from nodules of *D. glomerata* **(A)** and *C. thyrsiflorus* **(B)** is shown. For *D. glomerata*, five independent libraries prepared on Illumina HiSeq 2500 technology were combined and processed on Trinity to generate an assembly with a median (N50) of 1,850 nt. *C. thyrsiflorus* assembly was generated from an Illumina MiSeq library (see section "Materials and Methods"); when combined, the processed reads yielded an assembly with a median (N50) of 1,275 nt. Axes are given in $\log_{10}$ scale. The frequency of hierarchical species assigned by Trinotate for *D. glomerata* **(C)** and *C. thyrsiflorus* **(D)** is based on the total number of BlastX hits retrieved. Legend: ARATH, *Arabidopsis thaliana*; TOBAC, *Nicotiana tabacum* (common tobacco)*;* ORYSJ, *Oryza sativa* ssp. *japonica* (rice)*;* HUMAN, *Homo sapiens* (human)*;* MOUSE, *Mus musculus* (mouse)*;* SCHPO, *Schizosaccharomyces pombe* (fission yeast)*;* DICDI, *Dictyostelium discoideum* (slime mold)*;* DROME, *Drosophila melanogaster* (Fruit fly)*;* YEAST, *Saccharomyces cerevisiae* (baker's yeast). Species frequency is represented by $\log_{10}$ scale with actual numbers at the top of the bars.

FIGURE S2 | Averaged Transcript per Million reads (TPM) covering the Top 100 contigs assigned within the "biological process" category of *D. glomerata*. Top hits were selected based on the mean TPM of five libraries. Plotted is the median and IQR across these libraries (after doi: 10.17045/sthlmuni.6181772.v1).

FIGURE S3 | Averaged Transcript per Million reads (TPM) covering the top 100 contigs assigned within the "biological process" category of *C. thyrsiflorus*. Top hits were selected based on the mean TPM of three libraries. Plotted is the median and IQR across these libraries (after doi: 10.17045/sthlmuni.6181772.v1).

FIGURE S4 | Routes for primary nitrogen fixation in plants include those of ammonia **(A)** and nitrate assimilation **(B)**. The assimilation of ammonia is catalyzed in two steps in the GS/GOGAT cycle. Ammonia is first bound to L-glutamate, yielding L-glutamine, by glutamine synthetase (GS). L-glutamine and 2-oxoglutarate serve as substrates for glutamate synthase (GOGAT) to yield two molecules of L-glutamate **(A)**. Alternatively, nitrate can be uptaken from the rhizosphere. In the cytosol, nitrate can be reduced to nitrite by nitrate reductase (NR) and converted to ammonia by nitrite reductase (NIR). The final step involves the incorporation of ammonia into glutamine by GS **(B)**. Expression levels in graphic pathways **(A,B)** were calculated as the $\log_2[FC]$ relative to those of the housekeeper EF-1α estimated from the mean TPM of five (*D. glomerata*, long arrows with values written in normal font) and three (*C. thyrsiflorus*, short arrows with values written in bold italics) independent RNA preparations. Enzyme commission (EC) numbers and explanatory heatmap are provided. **(C)** Displays the transcript abundance quantified by RT-qPCR for key genes involved in the above mentioned pathways. Plotted results comparing the expression levels in roots (R) and nodules (N) are relative to those of the housekeeping gene EF-1α; the median and IQR of at least three biological replicates are given. Asterisks mark differences between roots and nodules for *$p < 0.05$ calculated by student's t-test followed by Benjamini–Hochberg FDR correction. The Y-axis is given in $\log_{10}$ scale.

FIGURE S5 | Arginine biosynthesis in chloroplasts *via* L-ornithine and its link to ammonium assimilation in the GS/GOGAT pathway. Ornithine synthesis starts from glutamate in a series of committed acetylation reactions (steps 1–5). Enzymes catalyzing these first steps include: (1) NAGS: *N*-acetylglutamate synthase; (2) NAGK: *N*-acetylglutamate kinase; (3) NAGPR: *N*-acetyl-glutamyl-*P*-reductase; (4) NAOAT: *N*-acetylornithine aminotransferase; and (5) NAOD: *N*-acetylornithine deacetylase. L-ornithine is combined with carbamoyl-phosphate, formed by by glutamine-dependent carbamoyl phosphate synthetase (CPS), to form L-citrulline. Ultimately, argininosuccinate synthase (ASSY) catalyzes the interconversion of L-citrulline and L-aspartate to L-arginino-succinate. Release of fumarate by argininosuccinate lyase (ARLY) yields L-arginine. Enzyme commission (EC) numbers are provided **(A)**. Expression levels represented in the metabolic map were calculated as the $\log_2$[FC] relative to those of the housekeeper EF-1α; values for *D. glomerata* appear in normal font over long arrows; for *C. thyrsiflorus* values are given in bold italic over short arrows. A metabolic heatmap is provided. Contig mean TPM values are plotted in **(B)**. **(C)** Displays the transcript abundance quantified by RT-qPCR for NAOAT and CPS. Gene expression levels in roots (R) and nodules (N) are relative to those of EF-1α. The median and IQR of three biological replicates is shown. Species name is given. Gene nomenclature used is based on Winter et al. (2015).

FIGURE S6 | Biosynthesis of thiamine. Thiamine (vitamin B1) is assembled by the condensation of pyrimidine and thiazole moieties. In *Arabidopsis thaliana*, THIC and TH1 catalyze the biosynthesis of the pyrimidine moiety, whereas thiazole is synthesized by the bifunctional THI1 enzyme (designated THI4 in other organisms; Godoi et al., 2006). After the coupling of the two precursors by TH1, thiamine phosphate is dephosphorylated to thiamine, which is subsequently dephosphorylated by TPK1 to its active form, thiamine diphosphate. Expression levels given in the metabolic map were calculated as the $\log_2$[FC] relative to those of EF-1α estimated from the mean TPM of five (*D. glomerata*, long arrows with values written in normal font) and three (*C. thyrsiflorus*, short arrows with values written in bold italic) independent RNA preparations. EC numbers and Arabidopsis thaliana homologs are given. Explanatory heatmap is provided.

FIGURE S7 | Tetrahydrofolate (THF) biosynthetic pathway. Generically known as folic acid, "folate" or vitamin B9, THF is assembled by three moieties: a pterin ring, p-aminobenzoate (pABA), and a glutamate tail. Enzymes involved in the synthesis of the pterin moity are: GTP cyclohydrolase I (GTPCHI), Nudix hydrolase (NUDT1), and dihydroneopterin aldolase (DHNA). In parallel, the pABA moiety is synthesized from chorismate. Enzymes operating in this branch are amino deoxychorismate synthase (ADCS) and 4-amino deoxychorismate lyase (ADCL). The coupling of these two branches is mediated by dihydropteroate synthase (DHPS), whose product is converted by dihydrofolate synthase (DHFS), followed by dihydrofolate reductase (DHFR). Expression levels given in the metabolic map were calculated as the $\log_2$[FC] relative to those of EF-1α based on the mean TPM of five (*D. glomerata*, long arrows with values written in normal font) and three

(*C. thyrsiflorus*, short arrows with values written in bold italics) independent RNA preparations. Explanatory heatmap is provided. EC numbers and Arabidopsis thaliana gene orthologs are given. Nomenclature is based on Gorelova et al. (2017).

FIGURE S8 | Hierarchical screening of transcripts encoding for members of the S8 subtilisin-like peptidase family from nodules of *D. glomerata* **(A)** and *C. thyrsiflorus* **(B)**. Heatmap *omic* colors translate the TPM in each library as low (white), medium low (green), medium (dark green), medium high (violet), and high (purple). Species names are provided. After doi: 10.17045/sthlmuni.6181760.v1.

FIGURE S9 | Maximum-likelihood phylogenetic tree of cucumisins. *Datisca glomerata* and *Ceanothus thyrsiflorus* candidates (in bold) were queried for phylogenetic placement using the comprehensive dataset of Taylor and Qiu (2017) (see section "Materials and Methods"). When possible, GenBank accessions are provided after species name. Otherwise, an asterisk (*) indicates the nomenclature used by Taylor and Qiu (2017).

FIGURE S10 | *Datisca glomerata* and *Ceanothus thyrsiflorus* nodules express putative orthologs of Nod factor receptors from legumes. Maximum-likelihood phylogenetic reconstruction of LysM-type and LysM receptor kinases is shown. Proteins encoded by genes identified based on studies of legume mutants are highlighted in color. Acronyms of included species are: At, *Arabidopsis thaliana*; Ca, *Cicer arietinum*; Cp, *Cucurbita pepo*; Fv, *Fragaria vesca*; Gm, *Glycine max*; Lj, *Lotus japonicus*; Me, *Manihot esculenta*; Mt, *Medicago truncatula*; Mn, *Morus notabilis*; Os, *Oryza sativa*; Ps, *Pisum sativum*; Pp, *Prunus persica*; Rc, *Ricinus communis*; Sl, *Solanum lycopersicum*; Sb, *Sorghum bicolor*; Tc, *Theobroma cacao*; Vr, *Vigna radiata*; and Zj, *Ziziphus jujuba*. Different orthogroups are distinguished by color: EPR3 in purple, NFR1 in green, and NFR5 in blue. Candidate orthologs from *D. glomerata* and *C. thyrsiflorus* are given in black in bold print. GenBank accessions are given.

TABLE S1 | *Datisca glomerata* pre-assembly summary: raw reads and filtering stages evaluation across five libraries. The last row displays the representation of each library into the Trinity assembly.

TABLE S2 | Primers used in this study for *Datisca glomerata*.

TABLE S3 | Primers used in this study for *Ceanothus thyrsiflorus*.

TABLE S4 | Target taxa used for phylogenetic reconstruction of LysM and LysM-type receptors kinases.

TABLE S5 | List of genes analysed by RT-qPCR. Gene names, protein functions, number of contig in assembly, GenBank accession number and, if available, E.C. number are given.

TABLE S6 | *p* values for RT-qPCR analysis. For all RT-qPCT analysis, *p*-values and adjusted (FDR) *p*-values are listed.

REFERENCES

Alloisio, N., Queiroux, C., Fournier, P., Pujic, P., Normand, P., Vallenet, D., et al. (2010). The *Frankia alni* symbiotic transcriptome. *Mol. Plant Microbe Interact.* 23, 593–607. doi: 10.1094/MPMI-23-5-0593

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Alunni, B., and Gourion, B. (2016). Terminal bacteroid differentiation in the legume-rhizobium symbiosis: nodule-specific cysteine-rich peptides and beyond. *New Phytol.* 211, 411–417. doi: 10.1111/nph.14025

Amor, B. B., Shaw, S. L., Oldroyd, G. E., Maillet, F., Penmetsa, R. V., Cook, D., et al. (2003). The NFP locus of *Medicago truncatula* controls an early step of Nod factor signal transduction upstream of a rapid calcium flux and root hair deformation. *Plant J.* 34, 495–506. doi: 10.1046/j.1365-313X.2003.01743.x

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Angiosperm Phylogeny Group. (2016). An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linnean Soc.* 181, 1–20. doi: 10.1111/boj.12385

Aronesty, E. (2013). Comparison of sequencing utility programs. *Open Bioinformatics J.* 7, 1–8. doi: 10.2174/1875036201307010001

Ayala-Rodríguez, J. Á., Barrera-Ortiz, S., Ruiz-Herrera, L. F., and López-Bucio, J. (2017). Folic acid orchestrates root development linking cell elongation with auxin response and acts independently of the TARGET OF RAPAMYCIN signaling in *Arabidopsis thaliana*. *Plant Sci.* 264, 168–178. doi: 10.1016/j.plantsci.2017.09.011

Baudouin, E., Pieuchot, L., Engler, G., Pauly, N., and Puppo, A. (2006). Nitric oxide is formed in *Medicago truncatula-Sinorhizobium meliloti* functional nodules. *Mol. Plant-Microbe Interact.* 19, 970–975. doi: 10.1094/MPMI-19-0970

Becana, M., Dalton, D. A., Moran, J. F., Iturbe-Ormaetxe, I., Matamoros, M. A., and Rubio, M. C. (2000). Reactive oxygen species and antioxidants in legume nodules. *Physiol. Plant.* 109, 372–381. doi: 10.1034/j.1399-3054.2000.100402.x

Bell, C. D., Soltis, D. E., and Soltis, P. S. (2010). The age and diversification of the angiosperms re-revisited. *Am. J. Bot.* 97, 1296–1303.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B* 57, 289–300. doi: 10.3732/ajb.0900346

Berg, R. H., Langenstein, B., and Silvester, W. B. (1999). Development in the *Datisca-Coriaria* nodule type. *Can. J. Bot.* 77, 1334–1350. doi: 10.1139/b99-076

Berry, A. M., Harriott, O. T., Moreau, R. A., Osman, S. F., Benson, D. R., and Jones, A. D. (1993). Hopanoid lipids compose the *Frankia* vesicle envelope, presumptive barrier of oxygen diffusion to nitrogenase. *Proc. Natl. Acad. Sci. U.S.A.* 90, 6091–6094. doi: 10.1073/pnas.90.13.6091

Berry, A. M., Mendoza-Herrera, A., Guo, Y.-Y., Hayashi, J., Persson, T., Barabote, R., et al. (2011). New perspectives on nodule nitrogen assimilation in actinorhizal symbioses. *Funct. Plant Biol.* 38, 645–652. doi: 10.1071/FP11095

Berry, A. M., Murphy, T. M., Okubara, P. A., Jacobsen, K. R., Swensen, S. M., and Pawlowski, K. (2004). Novel expression pattern of cytosolic Gln synthetase in nitrogen-fixing root nodules of the actinorhizal host, *Datisca glomerata*. *Plant Physiol.* 135, 1849–1862. doi: 10.1104/pp.103.031534

Bethke, P. C., Badger, M. R., and Jones, R. L. (2004). Apoplastic synthesis of nitric oxide by plant tissues. *Plant Cell* 16, 332–341. doi: 10.1105/tpc.017822

Carro, L., Pujic, P., Alloisio, N., Fournier, P., Boubakri, H., Hay, A. E., et al. (2015). Alnus peptides modify membrane porosity and induce the release of nitrogen-rich metabolites from nitrogen-fixing *Frankia*. *ISME J.* 9, 1723–1733. doi: 10.1038/ismej.2014.257

Casal, J. J., and Qüesta, J. I. (2018). Light and temperature cues: multitasking receptors and transcriptional integrators. *New Phytol.* 217, 1029–1034. doi: 10.1111/nph.14890

Caspi, R., Billington, R., Ferrer, L., Foerster, H., Fulcher, C. A., Keseler, I. M., et al. (2016). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucl Acids Res.* 44, D471–D480. doi: 10.1093/nar/gkv1164

Criscuolo, A., and Gribaldo, S. (2010). BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* 10:210. doi: 10.1186/1471-2148-10-210

Czernic, P., Gully, D., Cartieaux, F., Moulin, L., Guefrachi, I., Patrel, D., et al. (2015). Convergent evolution of endosymbiont differentiation in dalbergioid and inverted repeat-lacking clade legumes mediated by nodule-specific cysteine-rich peptides. *Plant Physiol.* 169, 1254–1265. doi: 10.1104/pp.15.00584

Demina, I. V., Persson, P., Santos, P., Płaszczyca, M., and Pawlowski, K. (2013). Comparison of the nodule vs. root transcriptome of the actinorhizal plant *Datisca glomerata*: actinorhizal nodules contain a specific class of defensins. *PLoS One* 8:8. doi: 10.1371/journal.pone.0072442

Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucl. Acids Res.* 39, W29–W37. doi: 10.1093/nar/gkr367

Forde, B. G., Day, H. M., Turton, J. F., Shen, W. J., Cullimore, J. V., and Oliver, J. E. (1989). Two glutamine synthetase genes from *Phaseolus vulgaris* L. display contrasting developmental and spatial patterns of expression in transgenic *Lotus corniculatus* plants. *Plant Cell* 1, 391–401. doi: 10.1105/tpc.1.4.391

Fournier, J., Imanishi, L., Chabaud, M., Abdou-Pavy, I., Genre, A., Brichet, L., et al. (2018). Cell remodeling and subtilase gene expression in the actinorhizal plant *Discaria trinervis* highlight host orchestration of intercellular *Frankia* colonization. *New Phytol.* 219, 1018–1030. doi: 10.1111/nph.15216

Fukao, T., and Bailey-Serres, J. (2004). Plant responses to hypoxia – Is survival a balancing act? *Trends Plant Sci.* 9, 449–456. doi: 10.1016/j.tplants.2004.07.005

Garrocho-Villegas, V., Gopalasubramaniam, S. K., and Arredondo-Peter, R. (2007). Plant hemoglobins: what we know six decades after their discovery. *Gene* 398, 78–85. doi: 10.1016/j.gene.2007.01.035

Godoi, P. H., Galhardo, R. S., Luche, D. D., Van Sluys, M. A., Menck, C. F., and Oliva, G. (2006). Structure of the thiazole biosynthetic enzyme THI1 from *Arabidopsis thaliana*. *J. Biol. Chem.* 281, 30957–30966. doi: 10.1074/jbc.M604469200

Gorelova, V., Ambach, L., Rébeillé, F., Stove, C., and Van Der Straeten, D. (2017). Folates in plants: research advances and progress in crop biofortification. *Front. Chem.* 5:21. doi: 10.3389/fchem.2017.00021

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883

Gtari, M., Ghodhbane-Gtari, F., Nouioui, I., Ktari, A., Hezbri, K., Mimouni, W., et al. (2015). Cultivating the uncultured: growing the recalcitrant cluster-2 *Frankia* strains. *Sci. Rep.* 5:13112. doi: 10.1038/srep13112

Guan, C., Ribeiro, A., Akkermans, A. D. L., Jing, Y., van Kammen, A., Bisseling, T., et al. (1997). Nitrogen metabolism in actinorhizal nodules of *Alnus glutinosa*:

expression of glutamine synthetase and acetylornithine transaminase. *Plant Mol. Biol.* 32, 1177–1184. doi: 10.1007/BF00041403

Guan, D., Stacey, N., Liu, C., Wen, J., Mysore, K. S., Torres-Jerez, I., et al. (2013). Rhizobial infection is associated with the development of peripheral vasculature in nodules of *Medicago truncatula*. *Plant Physiol.* 162, 107–115. doi: 10.1104/pp.113.215111

Gueddou, A., Swanson, E., Hezbri, K., Nouioui, I., Ktari, A., Simpson, S., et al. (2018). Draft genome sequence of the symbiotic *Frankia* sp. *strain* BMG5.30 isolated from root nodules of *Coriaria myrtifolia* in Tunisia. *Antonie Van Leeuwenhoek* [Epub ahead of print]. doi: 10.1007/s10482-018-1138-1

Gundel, P. E., Pierik, R., Mommer, L., and Ballaré, C. L. (2014). Competing neighbors: light perception and root function. *Oecologia* 176, 1–10. doi: 10.1007/s00442-014-2983-x

Günther, C., Schlereth, A., Udvardi, M., and Ott, T. (2007). Metabolism of reactive oxygen species is attenuated in leghemoglobin-deficient nodules of *Lotus japonicus*. *Mol. Plant-Microbe Interact.* 20, 1596–1603. doi: 10.1094/MPMI-20-12-1596

Gupta, K. J., and Igamberdiev, A. U. (2011). The anoxic plant mitochondrion as a nitrite:NO reductase. *Mitochondrion* 11, 537–543. doi: 10.1016/j.mito.2011.03.005

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/nprot.2013.084

Haney, C. H., and Long, S. R. (2010). Plant flotillins are required for infection by nitrogen-fixing bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 107, 478–483. doi: 10.1073/pnas.0910081107

Hanson, A. D., and Gregory, J. F. III (2011). Folate biosynthesis, turnover, and transport in plants. *Annu. Rev. Plant Biol.* 62, 105–125. doi: 10.1146/annurev-arplant-042110-103819

Heckmann, A. B., Hebelstrup, K. H., Larsen, K., Micaelo, N. M., and Jensen E. Ø. (2006). A single hemoglobin gene in *Myrica gale* retains both symbiotic and non-symbiotic specificity. *Plant Mol. Biol.* 61, 769–779. doi: 10.1007/s11103-006-0048-1

Hichri, I., Boscari, A., Castella, C., Rovere, M., Puppo, A., and Brouquisse, R. (2015). Nitric oxide: a multifaceted regulator of the nitrogen-fixing symbiosis. *J. Exp. Bot.* 66, 2877–2887. doi: 10.1093/jxb/erv051

Hirel, B., Perrot-Rechenmann, C., Maudinas, B., and Gadal, P. (1982). Glutamine synthetase in alder (*Alnus glutinosa*) root nodules, purification, properties and cytoimmunochemical localization. *Physiol. Plant.* 55, 197–203. doi: 10.1111/j.1399-3054.1982.tb02287.x

Hoagland, D. R., and Arnon, D. T. (1938). *The Water-Culture Method for Growing Plants Without Soil: California Agriculture Experiment Station Circular*. Berkeley, CA: University of California, 347.

Hocher, V., Alloisio, N., Auguy, F., Fournier, P., Doumas, P., Pujic, P., et al. (2011). Transcriptomics of actinorhizal symbioses reveals homologs of the whole common symbiotic signaling cascade. *Plant Physiol.* 156, 700–711. doi: 10.1104/pp.111.174151

Igamberdiev, A. U., Bykova, N. V., and Hill, R. D. (2006). Nitric oxide scavenging by barley hemoglobin is facilitated by a monodehydroascorbate reductase-mediated ascorbate reduction of methemoglobin. *Planta* 223, 1033–1040. doi: 10.1007/s00425-005-0146-3

Jacobsen-Lyon, K., Jensen, E. O., Jørgensen, J. E., Marcker, K. A., Peacock, W. J., and Dennis, E. S. (1995). Symbiotic and nonsymbiotic hemoglobin genes of *Casuarina glauca*. *Plant Cell* 7, 213–223. doi: 10.1105/tpc.7.2.213

Kawaharada, Y., Kelly, S., Nielsen, M. W., Hjuler, C. T., Gysel, K., Muszyński, A., et al. (2015). Receptor-mediated exopolysaccharide perception controls bacterial infection. *Nature* 523, 308–312. doi: 10.1038/nature14611

Kawaharada, Y., Nielsen, M. W., Kelly, S., James, E. K., Andersen, K. R., Rasmussen, S. R., et al. (2017). Differential regulation of the EPR3 receptor coordinates membrane-restricted rhizobial colonization of root nodule primordia. *Nat. Commun.* 8:14534. doi: 10.1038/ncomms14534

Kleemann, G., Alskog, G., Berry, A. M., and Huss-Danell, K. (1994). Lipid composition and nitrogenase activity of symbiotic *Frankia* (*Alnus incana*) in response to different oxygen concentrations. *Protoplasma* 183, 107–115. doi: 10.1007/BF01276818

Kojima, S., Bohner, A., Gassert, B., Yuan, L., and von Wirén, N. (2007). AtDUR3 represents the major transporter for high-affinity urea transport across the plasma membrane of nitrogen-deficient *Arabidopsis* roots. *Plant J.* 52, 30–40. doi: 10.1111/j.1365-313X.2007.03223.x

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580. doi: 10.1006/jmbi.2000.4315

Laplaze, L., Ribeiro, A., Franche, C., Duhoux, E., Auguy, F., Bogusz, D., et al. (2000). Characterization of a *Casuarina glauca* nodule-specific subtilisin-like protease gene, a homolog of *Alnus glutinosa* ag12. *Mol. Plant-Microbe Interact.* 13, 113–117. doi: 10.1094/MPMI.2000.13.1.113

Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323. doi: 10.1186/1471-2105-12-323

Miao, G. H., Hirel, B., Marsolier, M. C., Ridge, R. W., and Verma, D. P. (1991). Ammonia-regulated expression of a soybean gene encoding cytosolic glutamine synthetase in transgenic *Lotus corniculatus*. *Plant Cell* 3, 11–22. doi: 10.1105/tpc.3.1.11

Minchin, F. R. (1997). Regulation of oxygen diffusion in legume nodules. *Soil Biol. Biochem.* 29, 881–888. doi: 10.1016/S0038-0717(96)00204-0

Miyata, K., Hayafune, M., Kobae, Y., Kaku, H., Nishizawa, Y., Masuda, Y., et al. (2016). Evaluation of the role of the LysM receptor-like kinase, OsNFR5/OsRLK2 for AM symbiosis in rice. *Plant Cell Physiol.* 57, 2283–2290. doi: 10.1093/pcp/pcw144

Murray, J. D., Muni, R. R., Torres-Jerez, I., Tang, Y., Allen, S., Andriankaja, M., et al. (2011). Vapyrin, a gene essential for intracellular progression of arbuscular mycorrhizal symbiosis, is also essential for infection by rhizobia in the nodule symbiosis of *Medicago truncatula*. *Plant J.* 65, 244–252. doi: 10.1111/j.1365-313X.2010.04415.x

Mylona, P., Pawlowski, K., and Bisseling, T. (1995). Symbiotic nitrogen fixation. *Plant Cell* 7, 869–885. doi: 10.1105/tpc.7.7.869

Nagae, M., Parniske, M., Kawaguchi, M., and Takeda, N. (2016a). The thiamine biosynthesis gene THI1 promotes nodule growth and seed maturation. *Plant Physiol.* 172, 2033–2043. doi: 10.1104/pp.16.01254

Nagae, M., Parniske, M., Kawaguchi, M., and Takeda, N. (2016b). The relationship between thiamine and two symbioses: root nodule symbiosis and arbuscular mycorrhiza. *Plant Signal. Behav.* 11:e1265723. doi: 10.1080/15592324.2016.1265723

Nguyen, T. V., Wibberg, D., Battenberg, K., Vanden Heuvel, B., Blom, J., Berry, A. M., et al. (2016). An assemblage of Californian cluster II *Frankia* strains contains the canonical nod genes and also the sulfotransferase gene nodH. *BMC Genomics* 17:796. doi: 10.1186/s12864-016-3140-1

Normand, P., Lapierre, P., Tisa, L. S., Gogarten, J. P., Alloisio, N., Bagnarol, E., et al. (2007). Genome characteristics of facultatively symbiotic *Frankia* sp. strains reflect host range and host plant biogeography. *Genome Res.* 17, 7–15. doi: 10.1101/gr.5798407

Normand, P., Orso, S., Cournoyer, B., Jeannin, P., Chapelon, C., Dawson, J., et al. (1996). Molecular phylogeny of the genus *Frankia* and related genera and emendation of the family *Frankiaceae*. *Int. J. Syst. Bacteriol.* 46, 1–9. doi: 10.1099/00207713-46-1-1

Ogawa, T., Ueda, Y., Yoshimura, K., and Shigeoka, S. (2005). Comprehensive analysis of cytosolic Nudix hydrolases in *Arabidopsis thaliana*. *J. Biol. Chem.* 280, 25277–25283. doi: 10.1074/jbc.M503536200

Okubara, P. A., Pawlowski, K., Murphy, T. M., and Berry, A. M. (1999). Symbiotic root nodules of the actinorhizal plant *Datisca glomerata* express rubisco activase mRNA. *Plant Physiol.* 120, 411–420. doi: 10.1104/pp.120.2.411

Oldroyd, G. E. (2013). Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nat. Rev. Microbiol.* 11, 252–263. doi: 10.1038/nrmicro2990

Oono, R., Denison, R. F., and Kiers, E. T. (2009). Controlling the reproductive fate of rhizobia: how universal are legume sanctions? *New Phytol.* 183, 967–979. doi: 10.1111/j.1469-8137.2009.02941.x

Op den Camp, R., Streng, A., De Mita, S., Cao, Q., Polone, E., et al. (2011). LysM-type mycorrhizal receptor recruited for rhizobium symbiosis in nonlegume *Parasponia*. *Science* 331, 909–912. doi: 10.1126/science.1198181

Ott, T., van Dongen, J. T., Gunther, C., Krusell, L., Desbrosses, G., Vigeolas, H., et al. (2005). Symbiotic leghemoglobins are crucial for nitrogen fixation in legume root nodules but not for general plant growth and development. *Curr. Biol.* 15, 531–535. doi: 10.1016/j.cub.2005.01.042

Paley, S. M., and Karp, P. D. (2006). The pathway tools cellular overview diagram and omics viewer. *Nucl. Acids Res.* 34, 3771–3778. doi: 10.1093/nar/gkl334

Pan, H., and Wang, D. (2017). Nodule cysteine-rich peptides maintain a working balance during nitrogen-fixing symbiosis. *Nat. Plants* 3:17048. doi: 10.1038/nplants.2017.48

Parisi, K., Shafee, T. M. A., Quimbar, P., van der Weerden, N. L., Bleackley, M. R., and Anderson, M. A. (2018). The evolution, function and mechanisms of action for plant defensins. *Semin. Cell Dev. Biol.* [Epub ahead of print]. doi: 10.1016/j.semcdb.2018.02.004

Parsons, R., Silvester, W. B., Harris, S., Gruiters, W. T. M., and Bullivant, S. (1987). *Frankia* vesicles provide inducible and absolute oxygen protection for nitrogenase. *Plant Physiol.* 83, 728–731. doi: 10.1104/pp.83.4.728

Patriarca, E. J., Tate, R., and Iaccarino, M. (2002). Key role of bacterial NH4 + metabolism in *Rhizobium*-plant symbiosis. *Microbiol. Mol. Biol. Rev.* 66, 203–222. doi: 10.1128/MMBR.66.2.203-222.2002

Pattengale, N. D., Alipour, M., Bininda-Emonds, O. R. P., Moret, B. M. E., and Stamatakis, A. (2009). How many bootstrap replicates are necessary? *Lecture Notes Comp. Sci.* 17, 337–354. doi: 10.1007/978-3-642-02008-7_13

Pawlowski, K., and Demchenko, K. N. (2012). The diversity of actinorhizal symbiosis. *Protoplasma* 249, 967–979. doi: 10.1007/s00709-012-0388-4

Pawlowski, K., Jacobsen, K. R., Alloisio, N., Denison, R. F., Klein, M., Tjepkema, J. D., et al. (2007). Truncated hemoglobins in sctinorhizal nodules of *Datisca glomerata*. *Plant Biol.* 9, 776–785. doi: 10.1055/s-2007-965258

Perazzolli, M., Romero-Puertas, M. C., and Delledonne, M. (2006). Modulation of nitric oxide bioactivity by plant haemoglobins. *J. Exp. Bot.* 57, 479–488. doi: 10.1093/jxb/erj051

Persson, T., Battenberg, K., Demina, I. V., Vigil-Stenman, T., Vanden Heuvel, B., Pujic, P., et al. (2015). *Candidatus* Frankia datiscae Dg1, the actinobacterial microsymbiont of *Datisca glomerata*, expresses the canonical *nod* genes nodABC in symbiosis with its host plant. *PLoS One* 10:e0127630. doi: 10.1371/journal.pone.0127630

Persson, T., Van Nguyen, T., Alloisio, N., Pujic, P., Berry, A. M., Normand, P., et al. (2016). The N-metabolites of roots and actinorhizal nodules from *Alnus glutinosa* and *Datisca glomerata*: can *D. glomerata* change N-transport forms when nodulated? *Symbiosis* 70, 149–157. doi: 10.1007/s13199-016-0407-x

Petersen, R. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786. doi: 10.1038/nmeth.1701

Prell, J., White, J. P., Bourdes, A., Bunnewell, S., Bongaerts, R. J., and Poole, P. S. (2009). Legumes regulate rhizobium bacteroid development and persistence by the supply of branched-chain amino acids. *Proc. Natl. Acad. Sci. U.S.A.* 106, 12477–12482. doi: 10.1073/pnas.0903653106

Pumplin, N., Mondo, S. J., Topp, S., Starker, C. G., Gantt, J. S., and Harrison, M. J. (2010). *Medicago truncatula* vapyrin is a novel protein required for arbuscular mycorrhizal symbiosis. *Plant J.* 61, 482–494. doi: 10.1111/j.1365-313X.2009.04072.x

Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., et al. (2012). The Pfam protein families database. *Nucl. Acids Res.* 40, D290–D301. doi: 10.1093/nar/gkr1065

Radutoiu, S., Madsen, L. H., Madsen, E. B., Felle, H. H., Umehara, Y., Grønlund, M., et al. (2003). Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *Nature* 425, 585–592. doi: 10.1038/nature02039

Rapala-Kozik, M., Wolak, N., Kujda, M., and Banas, A. K. (2012). The upregulation of thiamine (vitamin B1) biosynthesis in *Arabidopsis thaliana* seedlings under salt and osmotic stress conditions is mediated by abscisic acid at the early stages of this stress response. *BMC Plant Biol.* 12:2. doi: 10.1186/1471-2229-12-2

Ribeiro, A., Akkermans, A. D. L., van Kammen, A., Bisseling, T., and Pawlowski, K. (1995). A nodule-specific gene encoding a subtilisin-like protease is expressed in early stages of actinorhizal nodule development. *Plant Cell* 7, 785–794. doi: 10.1105/tpc.7.6.785

Ribeiro, A., Praekelt, U., Akkermans, A. D. L., Meacock, P. A., van Kammen, A., Bisseling, T., et al. (1996). Identification of agthi1, whose product is involved in biosynthesis of the thiamine precursor thiazole, in actinorhizal nodules of *Alnus glutinosa*. *Plant J.* 10, 361–368. doi: 10.1046/j.1365-313X.1996.10020361.x

Ribeiro, D. T., Farias, L. P., de Almeida, J. D., Kashiwabara, P. M., Ribeiro, A. F., Silva Filho, M. C., et al. (2005). Functional characterization of the thi1 promoter region from *Arabidopsis thaliana*. *J. Exp. Bot.* 56, 1797–1804. doi: 10.1093/jxb/eri168

RStudio Team (2015) *RStudio: Integrated Development.* Boston, MA: R. RStudio, Inc. Available at: http://www.rstudio.com/

Sanz-Luque, E., Ocaña-Calahorro, F., de Montaigu, A., Chamizo-Ampudia, A., Llamas, Á., Galván, A., et al. (2015). THB1, a truncated hemoglobin, modulates nitric oxide levels and nitrate reductase activity. *Plant J.* 81, 467–479. doi: 10.1111/tpj.12744

Sasakura, F., Uchiumi, T., Shimoda, Y., Suzuki, A., Takenouchi, K., Higashi, S., et al. (2006). A class 1 hemoglobin gene from Alnus firma functions in symbiotic and nonsymbiotic tissues to detoxify nitric oxide. *Mol. Plant-Microbe Interact.* 19, 441–450. doi: 10.1094/MPMI-19-0441

Schubert, K. R. (1986). Products of biological nitrogen fixation in higher plants: synthesis, transport and metabolism. *Annu. Rev. Plant Physiol.* 37, 539–574. doi: 10.1146/annurev.pp.37.060186.002543

Schwender, J., Goffman, F., Ohlrogge, J. B., and Shachar-Hill, Y. (2004). Rubisco without the Calvin cycle improves the carbon efficiency of developing green seeds. *Nature* 432, 779–782. doi: 10.1038/nature03145

Sen, A., Daubin, V., Abrouk, D., Gifford, I., Berry, A. M., and Normand, P. (2014). Phylogeny of the class Actinobacteria revisited in the light of complete genomes. The orders 'Frankiales' and Micrococcales should be split into coherent entities: proposal of Frankiales ord. nov., Geodermatophilales ord. nov., Acidothermales ord. nov. and Nakamurellales ord. nov. *Int. J. Syst. Evol. Microbiol.* 64, 3821–3832. doi: 10.1099/ijs.0.063966-0

Shah, V. K., and Brill, W. J. (1973). Nitrogenase, IV. Simple method of purification to homogeneity of nitrogenase components from *Azotobacter vinelandii*. *Biochim. Biophys. Acta* 305, 445–454. doi: 10.1016/0005-2728(73)90190-4

Shimomura, A., Naka, A., Miyazaki, N., Moriuchi, S., Arima, S., Sato, S., et al. (2016). Blue light perception by both roots and rhizobia inhibits nodule formation in *Lotus japonicus*. *Mol. Plant-Microbe Interact.* 29, 786–796. doi: 10.1094/MPMI-03-16-0048-R

Sievers, F., Hughes, G. M., and Higgins, D. G. (2014). Systematic exploration of guide-tree topology effects for small protein alignments. *BMC Bioinformatics* 15:338. doi: 10.1186/1471-2105-15-338

Silvester, W. B., Harris, S. L., and Tjepkema, J. D. (1990). "Oxygen regulation and hemoglobin," in *The Biology of Frankia and Actinorhizal Plants*, C. R. Schwintzer and J. D. Tjepkema (San Diego, CA: Academic Press Inc), 157–193.

Silvester, W. B., Langenstein, B., and Berg, R. H. (1999). Do mitochondria provide the oxygen diffusion barrier in root nodules of *Coriaria* and *Datisca*? *Can. J. Bot.* 77, 1358–1366. doi: 10.1139/cjb-77-9-1358

Silvester, W. B., Silvester, J. K., and Torrey, J. G. (1988). Adaptation of nitrogenase to varying oxygen tension and the role of the vesicle in root nodules of *Alnus incana* ssp. *rugosa*. *Can. J. Bot.* 66, 1772–1779. doi: 10.1139/b88-243

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351

Slocum, R. D. (2005). Genes, enzymes and regulation of arginine biosynthesis in plants. *Plant Physiol. Biochem.* 43, 729–745. doi: 10.1016/j.plaphy.2005.06.007

Smit, P., Limpens, E., Geurts, R., Fedorova, E., Dolgikh, E., Gough, C., et al. (2007). *Medicago* LYK3, an entry receptor in rhizobial nodulation factor signaling. *Plant Physiol.* 145, 183–191. doi: 10.1104/pp.107.100495

Soltis, D. E., Soltis, P. S., Morgan, D. R., Swensen, S., Mullin, B. C., Dowd, J. M., et al. (1995). Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. *Proc. Natl. Acad. Sci. U.S.A.* 92, 2647–2651. doi: 10.1073/pnas.92.7.2647

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Strand, R., and Laetsch, W. M. (1977). Cell and endophyte structure of the nitrogen-fixing root nodules of Ceanothus integerrimus H. and A. I. Fine structure of the nodule and its endosymbiont. *Protoplasma* 93, 165–178. doi: 10.1007/BF01275651

Sturms, R., Kakar, S., Trent, J. III, and Hargrove, M. S. (2010). *Trema* and *Parasponia* hemoglobins reveal convergent evolution of oxygen transport in plants. *Biochemistry* 49, 4085–4093. doi: 10.1021/bi1002844

Suzaki, T., Yoro, E., and Kawaguchi, M. (2015). Leguminous plants: inventors of root nodules to accommodate symbiotic bacteria. *Int. Rev. Cell Mol. Biol.* 316, 111–158. doi: 10.1016/bs.ircmb.2015.01.004

Svistoonoff, S., Laplaze, L., Auguy, F., Runions, J., Duponnois, R., Haseloff, J., et al. (2003). Cg12 expression is specifically linked to infection of root hairs and cortical cells during *Casuarina glauca* and *Allocasuarina verticillata* actinorhizal nodule development. *Mol. Plant-Microbe Interact.* 16, 600–607. doi: 10.1094/MPMI.2003.16.7.600

Svistoonoff, S., Laplaze, L., Liang, J., Ribeiro, A., Gouveia, M. C., Auguy, F., et al. (2004). Infection-related activation of the *cg12* promoter is conserved between actinorhizal and legume-rhizobia root nodule symbiosis. *Plant Physiol.* 136, 3191–3197. doi: 10.1104/pp.104.048967

Swensen, S. M. (1996). The evolution of actinorhizal symbioses: evidence for multiple origins of the symbiotic association. *Am. J. Bot.* 83, 1503–1512. doi: 10.1002/j.1537-2197.1996.tb13943.x

Takeda, N., Sato, S., Asamizu, E., Tabata, S., and Parniske, M. (2009). Apoplastic plant subtilases support arbuscular mycorrhiza development in *Lotus japonicus*. *Plant J.* 58, 766–777. doi: 10.1111/j.1365-313X.2009.03824.x

Takeda, N., Tsuzuki, S., Suzaki, T., Parniske, M., and Kawaguchi, M. (2013). CERBERUS and NSP1 of *Lotus japonicus* are common symbiosis genes that modulate arbuscular mycorrhiza development. *Plant Cell Physiol.* 54, 1711–1723. doi: 10.1093/pcp/pct114

Taylor, A., and Qiu, Y.-L. (2017). Evolutionary history of subtilases in land plants and their involvement in symbiotic interactions. *Mol. Plant-Microbe Interact.* 30, 489–501. doi: 10.1094/MPMI-10-16-0218-R

Temple, S. J., Heard, J., Ganter, G., Dunn, K., and Sengupta-Gopalan, C. (1995). Characterization of a nodule-enhanced glutamine synthetase from alfalfa: nucleotide sequence, in situ localization, and transcript analysis. *Mol. Plant-Microbe Interact.* 8, 218–227. doi: 10.1094/MPMI-8-0218

Tisserant, E., Malbreil, M., Kuo, A., Kohler, A., Symeonidi, A., Balestrini, R., et al. (2013). Genome of arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* 110, 20117–20122. doi: 10.1073/pnas.1313452110

Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–1111. doi: 10.1093/bioinformatics/btp120

Tunc-Ozdemir, M., Miller, G., Song, L., Kim, J., Sodek, A., Koussevitzky, S., et al. (2009). Thiamin confers enhanced tolerance to oxidative stress in *Arabidopsis*. *Plant Physiol.* 151, 421–432. doi: 10.1104/pp.109.140046

Valverde, C., and Wall, L. G. (2003). Ammonium assimilation in root nodules of actinorhizal *Discaria trinervis*, regulation of enzyme activities and protein levels by the availability of macronutrients (N, P and C). *Plant Soil* 254, 139–153. doi: 10.1023/A:1024907130981

Van de Velde, W., Zehirov, G., Szatmari, A., Debreczeny, M., Ishihara, H., Kevei, Z., et al. (2010). Plant peptides govern terminal differentiation of bacteria in symbiosis. *Science* 327, 1122–1126. doi: 10.1126/science.1184057

van Gelderen, K., Kang, C., and Pierik, R. (2018). Light signaling, root development, and plasticity. *Plant Physiol.* 176, 1049–1060. doi: 10.1104/pp.17.01079

van Velzen, R., Holmer, R., Bu, F., Rutten, L., van Zeijl, A., Liu, W., et al. (2018). Comparative genomics of the nonlegume *Parasponia* reveals insights into evolution of nitrogen-fixing rhizobium symbioses. *Proc.*

Natl. Acad. Sci. U.S.A. 115, E4700–E4709. doi: 10.1073/pnas.172139
5115

Walsh, K. B., Ng, B. H., and Chandler, G. E. (1984). Effects of nitrogen nutrition on
xylem sap composition of Casuarinaceae. *Plant Soil* 81, 291–293. doi: 10.1007/
BF02197162

Winter, G., Todd, C. D., Trovato, M., Forlani, G., and Funck, D. (2015).
Physiological implications of arginine metabolism in plants. *Front. Plant Sci.*
6:534. doi: 10.3389/fpls.2015.00534

Yano, K., Shibata, S., Chen, W. L., Sato, S., Kaneko, T., Jurkiewicz, A., et al. (2009).
CERBERUS, a novel U-box protein containing WD-40 repeats, is required for
formation of the infection thread and nodule development in the legume-
rhizobium symbiosis. *Plant J.* 60, 168–180. doi: 10.1111/j.1365-313X.2009.
03943.x

Yokawa, K., and Baluška, F. (2015). Pectins, ROS homeostasis and UV-B responses
in plant roots. *Phytochemistry* 112, 80–83. doi: 10.1016/j.phytochem.2014.
08.016

Zdyb, A., Salgado, M. G., Demchenko, K. N., Płaszczyca, M., Stumpe, M.,
Herrfurth, C., et al. (2018). Allene oxide synthase, allene oxide cyclase and
jasmonic acid levels in *Lotus japonicus* nodules. *PLoS One* 13:e0190884. doi:
10.1371/journal.pone.0190884

Conflict of Interest Statement: The authors declare that the research was
conducted in the absence of any commercial or financial relationships that could
be construed as a potential conflict of interest.

Dryas as a Model for Studying the Root Symbioses of the Rosaceae

Benjamin Billault-Penneteau[1], Aline Sandré[1], Jessica Folgmann[1], Martin Parniske[1]* and Katharina Pawlowski[2]*

[1] Institute of Genetics, Faculty of Biology, LMU Munich, Martinsried, Germany, [2] Department of Ecology, Environment and Plant Sciences, Stockholm University, Stockholm, Sweden

The nitrogen-fixing root nodule symbiosis is restricted to four plant orders: Fabales (legumes), Fagales, Cucurbitales and Rosales (Elaeagnaceae, Rhamnaceae, and Rosaceae). Interestingly all of the Rosaceae genera confirmed to contain nodulating species (i.e., *Cercocarpus*, *Chamaebatia*, *Dryas*, and *Purshia*) belong to a single subfamily, the Dryadoideae. The *Dryas* genus is particularly interesting from an evolutionary perspective because it contains closely related nodulating (*Dryas drummondii*) and non-nodulating species (*Dryas octopetala*). The close phylogenetic relationship between these two species makes *Dryas* an ideal model genus to study the genetic basis of nodulation by whole genome comparison and classical genetics. Therefore, we established methods for plant cultivation, transformation and DNA extraction for these species. We optimized seed surface sterilization and germination methods and tested growth protocols ranging from pots and Petri dishes to a hydroponic system. Transgenic hairy roots were obtained by adapting *Agrobacterium rhizogenes*-based transformation protocols for *Dryas* species. We compared several DNA extraction protocols for their suitability for subsequent molecular biological analysis. Using CTAB extraction, reproducible PCRs could be performed, but CsCl gradient purification was essential to obtain DNA in sufficient purity for high quality *de novo* genome sequencing of both *Dryas* species. Altogether, we established a basic toolkit for the culture, transient transformation and genetic analysis of *Dryas* sp.

Keywords: *Dryas*, model-plant, *Dryas drummondii*, *Dryas octopetala*, Rosaceae, genome comparison

INTRODUCTION

Nitrogen and phosphate are key nutrients for plant growth, but their availability is limited, especially in alkaline and calcareous soils (Vitousek et al., 2010; Lambers et al., 2012). Plants, with their limited capacity to retrieve nutrients from the soil, profit from interactions with beneficial microorganisms. Root endosymbiosis with arbuscular mycorrhizal fungi or nitrogen fixing bacteria are examples of this kind of interaction, where the microsymbiont is accommodated within root cells, leading to a gain of function that enables the plant to survive or even thrive in previous uninhabitable environments. Among terrestrial plants, the vast majority (80%) is able to develop arbuscular mycorrhiza (AM), a symbiosis with fungi of the Glomeromycotan (Delaux et al., 2013); one of the main benefits of the AM symbiosis is the ability of the fungi to improve the host plants' access to phosphate. Despite the advantages of this symbiosis, some plant lineages have lost genes essential for its establishment. One such lineage is the Brassicaceae family that comprises the

model plant *Arabidopsis thaliana* (Cosme et al., 2018). Another symbiotic system evolved for the acquisition of nitrogen through the cooperation with nitrogen fixing *Frankia* or rhizobia bacteria, the nitrogen-fixing root nodule symbiosis (RNS). The symbiosis is named after the specialized root organs formed by the plant, root nodules, which provide physiological conditions for the bacteria to fix atmospheric nitrogen and export a product of nitrogen fixation to the host plant. This endosymbiosis is restricted to plant species in the related orders Fabales, Fagales, Cucurbitales and Rosales, forming the FaFaCuRo clade (Soltis et al., 1995; Werner et al., 2014; Griesmann et al., 2018). Legumes (Fabales) and *Parasponia* sp. (Cannabaceae, Rosales) interact with rhizobia, while all other RNS forming plant species interact with the actinobacterium *Frankia* and consequently are called actinorhizal plants.

Genetic dissection of rhizobial symbiosis, mainly two model legumes – *Medicago truncatula* (Barrel medic) and *Lotus corniculatus* L. (bird's-foot trefoil) *var. japonicus,* – has revealed symbiosis-related genes that are essential for nodule organogenesis, bacterial infection, and nitrogen fixation (see, e.g., Geurts et al., 2016). Recent phylogenomic studies have revealed that although most current members of the FaFaCuRo clade cannot form root nodules, the common ancestor of the FaFaCuRo clade was able to enter a symbiosis (Griesmann et al., 2018; van Velzen et al., 2018). However, it is not clear whether this ancestral symbiosis involved the formation of root nodules (Parniske, 2018). Due to their common origin, several similarities exist between actinorhizal and rhizobial symbioses, and the transfer of knowledge from legumes to actinorhizal plants improves our understanding of the main processes underlying the symbiosis with *Frankia* bacteria. Nevertheless, important aspects of actinorhizal symbioses remain unknown (Pawlowski and Bisseling, 1996; Perrine-Walker et al., 2011; Van Nguyen and Pawlowski, 2017). Consequently, research on actinorhizal plants could allow us to better understand the evolution of divergent symbiotic processes in RNS.

The Rosaceae family, belonging to the order Rosales, is globally the 4th most important plant family in terms of economic value (Vallée et al., 2016). Surprisingly, from *ca.* hundred Rosaceae genera only four, forming the basal subfamily Dryadoideae (Xiang et al., 2016), have been described to contain actinorhizal species that are able to enter a nitrogen-fixing RNS with *Frankia* bacteria (Pawlowski and Demchenko, 2012). The most basal genus of the Rosaceae, *Dryas*, is one of the most dominant dwarf shrubs among the arctic plant genera in terms of biomass. The taxonomy within the *Dryas* genus is controversial due to the existence of hybrids that occur naturally in the wild (Packer, 1994; Philipp and Siegismund, 2003). In areas where different *Dryas* species cohabit, natural hybrids were described between *Dryas integrifolia* and *Dryas octopetala* (Philipp and Siegismund, 2003), or between *Dryas. drummondii* and *D. integrifolia* (known as *D. x wyssiana*). The German botanist Franz Sündermann also created *D. x suendermannii* by crossing *D. drummondii* with *D. octopetala* (Packer, 1994), as part of the collection of the "Botantischen Alpengarten Sündermann" at Lindau, Germany, where the hybrids are maintained by clonal propagation.

Currently, three *Dryas* species, *D. drummondii*, *D. integrifolia*, and *D. octopetala*, are recognized; however, the genus is in need of taxonomic revision (Porslid, 1947; Böcher et al., 1968; Hultén, 1968; Yurtsev, 1997; Philipp and Siegismund, 2003; Skrede et al., 2006). The genus *Dryas* is unique in that it contains closely related nodulating and non-nodulating species which makes it an ideal model to study the evolution of root symbioses. Nodulation was reported for the first time in 1967 (Lawrence et al., 1967) in the North American species *D. drummondii* (Newcomb, 1981; Kohls et al., 1994; **Figure 3A**). The other species appear to be non-nodulating (Becking, 1970; Markham, 2009), but all *Dryas* species form ectomycorrhiza (Melville et al., 1988; Ryberg et al., 2009; Bjorbækmo et al., 2010; Botnen et al., 2014). The only exception may be *D. drummondii* because it has been described as ectomycorrhizal only on one occasion (Fitter and Parsons, 1987). The question whether this species can form ectomycorrhiza requires further examination.

The arctic-alpine species *D. octopetala* has a particularly wide distribution; it can be used for mapping refugial isolation and postglacial expansion during the glaciation in the Pleistocene in northern Europe (Philipp and Siegismund, 2003; Skrede et al., 2006). The plants typically grow in alkaline calcareous soils (Crocker and Major, 1955) and thus face nutritional limitations especially in terms of nitrogen and phosphorus (Lambers et al., 2012). The presence and abundance of *Dryas* species across all arctic and alpine tundra's makes this genus a key player in arctic phylo- and bio-geography (Tremblay and Schoen, 1999; Skrede et al., 2006), landscape ecology (Eichel et al., 2016; Eichel et al., 2017) and mycology community ecology (Väre et al., 1992; Ryberg et al., 2009; Bjorbækmo et al., 2010; Brunner et al., 2017). The genus is of particular importance in research on climate change (see, e.g., McGraw et al., 2014; Gillespie et al., 2016; Panchen and Gorelick, 2017). Therefore, *Dryas* species also are integral to Citizen Science Projects, e.g., the Spatial Food Web Ecology Group of the University of Helsinki use *Dryas* sp. as base of two Ecosystem Ecology projects, the Arctic Parasitoid Project and the Global *Dryas* Project[1], and the Climate Impact Research Centre in Abisko also includes *Dryas octopetala* in their target plants[2].

Dryas species are diploid with an estimated haploid genome size of 250 Mbp (Griesmann et al., 2018) distributed over nine chromosomes (Potter et al., 2007), less than *Malus × domestica* (apple) which is diploid or triploid with 750 Mbp, or *Rosa* which is tetraploid or triploid with 600 Mbp (Jung et al., 2013). The genome sequence of *D. drummondii* obtained from DNA purified with the protocol described here is publicly available; all data have been deposited in GigaDB (Griesmann et al., 2018). This small genome, combined with a generation time of less than a year, makes *Dryas* suitable as model genus for the Rosaceae family.

In this study, we focused on *D. drummondii* and *D. octopetala*. We omitted *D. integrifolia* because of its high similarity with *D. octopetala* (Skrede et al., 2006), the latter being more accessible and better researched. The close relation between *D. drummondii* and *D. octopetala* allows genomic comparisons in order to

---

[1]https://www.helsinki.fi/en/researchgroups/spatial-food-web-ecology/research
[2]https://www.arcticcirc.net/our-projects/

identify genes specifically involved in plant root endosymbiosis. We present the advances achieved in the development and adaptation of protocols in order to use *Dryas* as a model genus in Rosaceae research as well as to study the evolution of root symbioses.

MATERIALS AND METHODS

Dryas Seeds and Ecotypes

Seeds of *D. drummondii* DA462 and *D. octopetala* DA460 were purchased from the seed producer Jelitto (Schwarmstedt, Germany). The Nymphenburg Botanical Garden of Munich supplied seeds for *D. drummondii* BGM. Ecotypes Albe. (origin Clearwater County, Alberta, Canada, collected in 2000); and Alas. (origin Alaska, United States, collected in 2002) were found in and supplied by the KEW Millennium Seed Bank (Royal Botanic Gardens, Kew, London, United Kingdom). *D. octopetala* ecotype E548 was harvested in the Italian Alps (approximate GPS coordinates: 46°24′36.7″N 11°37′48.2″E) by Anna Heuberger.

Primers

Primers were designed based on the first draft genome of *D. drummondii* (Griesmann et al., 2018). They are as follows: GADPH forward 5′-CCCCAGTACGAATGCTCCCATGTTT G-3′, GADPH reverse 5′-TTAGCCAAAGGAGCAAGACAGTT GGTGG-3′; EF1-a forward 5′-TGGGTTTGAGGGTGACAACA TGA-3′; EF1-a reverse 5′-GTACACATCCTGAAGTGGGAGAC GGAGG-3′; 26S rRNA forward 5′-TACTGCAGGTCGGCAAT CGG-3′, 26S rRNA reverse 5′-TCATCGCGCTTGGTTGAAAA-3′. ITS primers were designed based on Cheng et al. (2016): ITS forward 5′-CCTTATCAYTTAGAGGAAGGAG-3′, ITS reverse 5′-RGTTTCTTTTCCTCCGCTTA-3′.

Seed Storage and Sterilization

Based on advice from seed producers and on results from Nichols (1934) who observed that without prior refrigeration, germination of several alpine species was considerably reduced, we assumed that seeds of *Dryas* species might require cold stratification prior to germination and therefore stored them at 4°C. *Dryas* seeds were surface sterilized by immersion in 30% H_2O_2 (10 min for *D. octopetala*; 15 min for *D. drummondii*) and washed three times with sterile H_2O. These experiments were performed with four biological replicates per species, with at least 100 seeds per replicate.

Growth Systems

Sterilized seeds were transferred on 1% agar-water plates and incubated in the dark at 22°C for 12 and 8 days for *D. octopetala* and *D. drummondii,* respectively. Several sources of agar were tested such as Bacto™ agar (Becton Dickinson and company) and agar Kalys HP 696 (Kalys SA, Bernin, France). The germination assays were set up in the dark because this reportedly increased germination rates Bliss (1958). After germination, seedlings of *Dryas* spp. were grown on plates, in a hydroponic system or in pots.

Growth on plates was performed on $^1/_4$ Hoagland's pH 5.8 (using the protocol for N-free medium; Hoagland and Arnon, 1950) and adding 1 KNO_3 to a final concentration of 1 mM) with 0.4% of Gelrite (Duchefa, Haarlem, Netherlands), at 22°C, 55% of humidity with 16 h-light/8 h-dark cycles. After 1 week, plantlets were transferred either into pots or Weck jars (containing production substrate A210; Stender AG, Germany) or into a hydroponic system.

The hydroponic system consisted of a standard 1 mL pipette tip box in two parts: the bottom part contained 250 mL of growth medium ($^1/_4$ Hoagland, 1 mM KNO_3, pH 5.8) and the tip holder with 24 holes in which the plantlets were inserted (**Figure 2**). To avoid seedlings or young plantlets falling into the medium compartment, the holes were covered with adhesive tape and plantlets were introduced through thin slits cut into the tape. The growth medium was changed twice per week. The hydroponic system was kept in a growth cabinet at 22°C, 55% humidity with 16 h-light/8 h-dark cycles for a maximum of 4 months.

Plants in pots were transferred to the greenhouse (day temperature 21–24°C, night temperature 18–21°C, with additional lighting from 6:00 to 10:00 h and from 15:00 to 22:00 h). The pots were filled either with sand:vermiculite (2:1) or with propagating substrate (A210 Stender AG, Germany). Note that temperature and light conditions were applied as available in our plant growth facilities and not experimentally optimized for *Dryas*.

Cutting Propagation

For clonal propagation, young and soft shoots of *Dryas* spp. were cut after the third internode (2–5 cm) above the woody part of the shoot. These explants were directly transferred into moist production substrate A210 (Stender AG, Germany), then kept under plastic cover in the greenhouse. High humidity was maintained under the cover by spraying with water every 2 days for 2 weeks; thereafter, spraying was stopped and cuttings were kept in moist soil under the cover until new leaves had developed and the covers were removed. Three series of ca. 20 cuttings per species were cultivated in the greenhouse during different seasons.

Hairy Root Transformation

We established a protocol for hairy root transformation in *Dryas* spp. by adapting *Lotus* protocols. The *Agrobacterium rhizogenes* strain AR1193 (Stougaard et al., 1987) was used because it had been shown to be very efficient for some plant species such as pea (Clemow et al., 2011) and because it was one of the strains available in our lab previously successfully tested for *Lotus japonicus* hairy root transformation.

Agrobacterium rhizogenes AR1193 bacteria carrying a Golden Gate LIIIβ F A-B (Binder et al., 2014) plasmid containing the *mCherry* gene under control of the *Ubiquitin* promoter (*AtUbi10*pro) as transformation marker (Pimprikar et al., 2016), were grown in liquid culture (LB medium with 50 μg mL^{-1} each of rifampicin, carbenicillin and kanamycin) at 28°C overnight. Bacteria were collected via a centrifugation step (15 min at 4.369 × *g*) and resuspended in water to obtain the wanted OD_{600} (0.01; 0.1; 1; 7.2). Cut hypocotyls of

10–12 days old axenically grown *Dryas* spp. seedlings were dipped in the bacteria suspension and placed on ¼ Hoagland (1 mM KNO₃, pH 5.8), 0.4% Gelrite (Duchefa, Haarlem, The Netherlands) plates. The plates were kept for 4 days in the dark at 22°C, then under a 16 h-light/8 h-dark cycle with 55% humidity. To prevent overgrowth of bacteria and dehydration, the plants were transferred onto new plates every week. Four to six weeks after transformation, roots were screened using a Leica MZ16 FA stereomicroscope (Leica Microsystems GmbH, Wetzlar, Germany) using the N3 filter from Leica (BP 546/12;600/40).

DNA Extraction and PCR Reactions

The CsCl gradient DNA extraction method was performed according to Ribeiro et al. (1995). Six to ten grams of leaves (mix of young and old from the same plant) were ground by hand using pistil and mortar with 4 g of PolyclarAT in liquid nitrogen. For the other extraction methods, the two youngest leaves of a shoot with the apical meristem were used as starting material. After being shock frozen in liquid nitrogen, they were ground with a Retsch Mill MM400 (Fa. Retsch, Haan, Germany) two times at 30 Hz for 30 s in 2 mL Eppendorf tubes containing 2 mm diameter stainless steel beads each. The "classical CTAB" extraction method is described in Doyle and Doyle (1987), whereas the "PVP/NaCl" extraction method was developed by Khanuja et al. (1999) based on the classical CTAB method.

PCRs were performed on 1 μl of DNA (20–700 ng of DNA, usually ca. 100 ng) using GoTaq® DNA polymerase (Promega, Germany), SYBR Green buffer and 0.2 μM of each primer. Amplifications were carried for 5 min at 95°C, followed by 35 cycles (30 s at 95°C, 30 s at 60°C, and 40 s at 72°C), and a final extension for 1 min at 72°C. Electrophoresis of a 4 μL of PCR reaction was performed on a 3% agarose gel for 100 min at 130 V. DNA was visualized with UVP UV solo touch from Analytik Jena© (Jena, Germany) after incubation of the gel for 10 min in an Ethidium bromide bath at 2 ng mL⁻¹.

RNA Extraction

Leaves, seedlings and root systems were shock frozen in liquid nitrogen. RNA of ground material (with the same procedure previously described for DNA extraction) was extracted using the Spectrum™ Plant Total RNA Kit (Sigma-Aldrich, CA, United States) without adaptation in the protocol except for older and thicker leaves for extraction of which Polyclar AT was added. The RNA was treated with DNAse (Invitrogen™ TURBO DNA-free Kit; Carlsbad, CA, United States) and tested for purity and integrity with a Bioanalyzer from Agilent (Agilent Technologies, Palo Alto, CA, United States). Twelve independently isolated RNA samples were analyzed.

RESULTS AND DISCUSSION

To establish *Dryas* as new model genus in the laboratory, we developed cultivation protocols under controlled conditions.
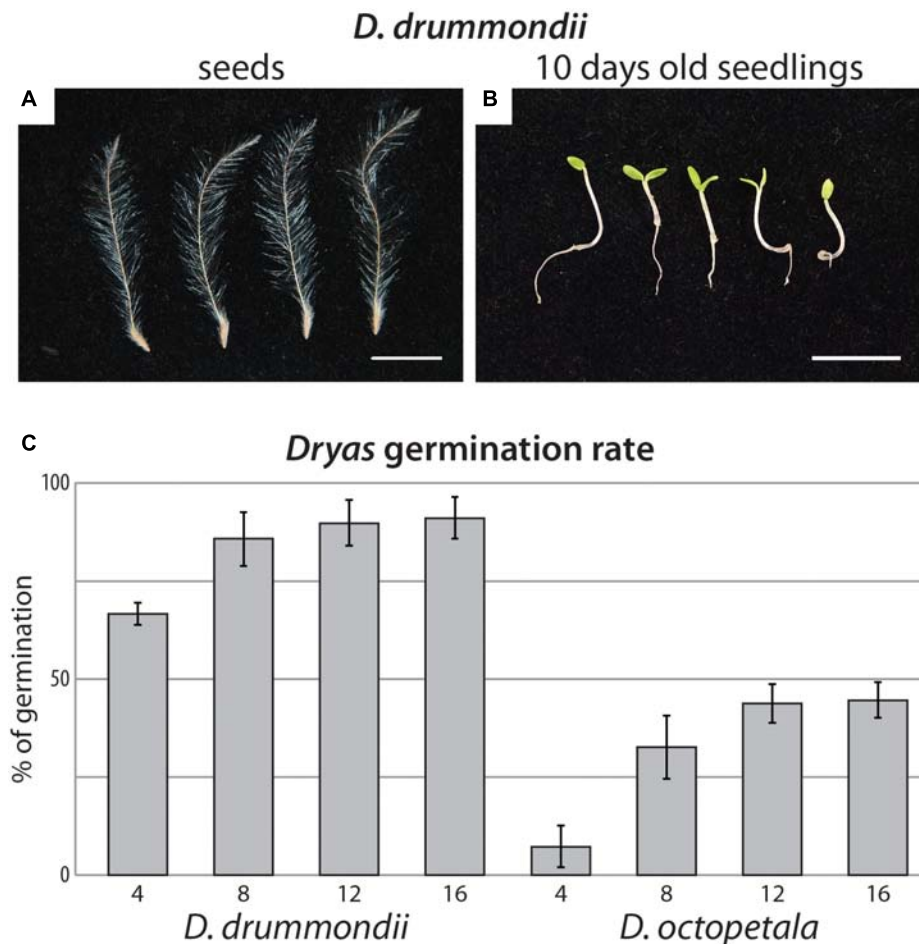
Dryas Seeds and Germination

Fungal contamination of seeds was often observed, whether they were collected in the field or obtained from a professional seed producer. In our study, white fungal hyphae growing out of the seeds led to seedling death at an early stage. *Dryas* seeds were quite sensitive to different surface sterilization procedures: any traces of ethanol would completely inhibit germination, while the thinness of the seed coat rendered the use of sulphuric acid for scarification risky. Furthermore, the contaminating fungi were quite resistant to NaOCl. However, after stratification of *Dryas* seeds at 4°C (**Figure 1A**), most efficient sterilization and highest germination rates were observed using hydrogen peroxide. Indeed approximately 100% of *D. octopetala* seeds and between 99 and 100% of *D. drummondii* seeds were free of contaminants after the procedure. Four replicates, with at least 100 seeds per replicate, were observed every 2 days. For *D. drummondii*, the maximal germination rate (85%) was obtained 8 days post-sterilization, whereas the maximum germination rate of *D. octopetala*, *ca.* 40%, was only reached 12 days post-sterilization (**Figure 1C**).

The time difference of 4 days to reach maximum germination observed between *D. drummondii* and *Dryas octopetala* was consistently observed at least within the samples tested: seeds of all *D. drummondii* seed sources examined germinated within 8 days and *D. octopetala* seeds from all sources available within 12 days. This phenotype was consistent not only for seeds produced in the greenhouse or botanical garden during the same month, but also for commercial seeds, the age of which was unknown. However, given that *D. octopetala* has a very wide distribution, it is possible that the two seed sources examined do not encompass the entire variability of the species.

Dryas Growth Systems

We examined different growth conditions and systems including Gelrite and agar plates, hydroponic systems and classical pots. These distinct growth systems combine diverse advantages for research such as axenic culture, conditions for root system observations and for inoculation with the microsymbiont. *Frankia* strains able to nodulate *D. drummondii* have not yet been successfully cultured (Normand et al., 2017), necessitating infection with crushed nodules. As these nodules carry a rich fungal and bacterial microbiome on the surface, inoculation of *Dryas* with these nodules while maintaining a gnotobiotic system is challenging. On the other hand, plants grown in pots do not represent the most suitable system for root analyses. The process of cleaning the soil from the roots stresses the plant and furthermore, the harvesting of root systems entails the risk of breaking thin and fragile lateral roots and root hairs. To circumvent this drawback, plants can be grown in Petri dishes. For *Dryas* species this system was suited for early stages of development; for experiments exceeding 4–5 weeks, root and shoot growth required more space. Furthermore, shielding of plates never totally protected the roots from light, and a long exposure of roots to light tends to interfere with the analysis of root responses to any treatment. Exposure of roots to direct light modifies their transcriptome (Hemm et al., 2004) and often

FIGURE 1 | *Dryas* spp. from seeds to seedlings. **(A)** Isolated anemochorous silky-feathery achenes from *Dryas drummondii*. **(B)** 10-day-old seedlings of *D. drummondii*. Scale bars denote 1 cm. **(C)** Time course of *Dryas* sp. germination after seed surface sterilization. Displayed are means (*n* = 4 biological replicates with at least 100 seeds per replicate) and standard errors.

leads to stress responses, which can perturb the analyses and cause misleading effects. Hydroponic systems offer the possibility to observe the roots in a non-invasive way while also shielding them from light. They can be used with or without an inert substrate that mimics physical soil contact. The fact that *Dryas* species can grow in well-aerated soil but can also tolerate flood periods (West et al., 1993), suggested the use of hydroponics as a method of choice.

Once germinated and grown on 1% agar plates with classical plant media like B5 or MS (Murashige and Skoog, 1962; Gamborg et al., 1968; Duchefa, Haarlem, Netherlands) or Fåhræus medium (Fåhraeus, 1957) with 1 mM KNO$_3$, *Dryas* species seedlings turned reddish, likely due to μlanin production, a response typically interpreted as stress- or defense-related. This anthocyanin production was less pronounced when the seedlings were grown on 0.4% Gelrite with ¼ strength Hoagland medium containing 1 mM KNO$_3$ (**Figure 1B**). Moreover, after 2 weeks on plates, *Dryas* spp. plantlets grown on ¼ strength Hoagland medium showed darker green cotyledons and further developed root systems than on B5 medium. Thus, among the

tested media for *Dryas* seedlings on gel-forming media, the best results were obtained with 0.4% Gelrite containing ¼ strength Hoagland solution.

After germination on plates, *Dryas* species. plantlets were transferred to a hydroponic system (**Figure 2A**) with ¼ Hoagland solution. In this system *Dryas* species plants grew and developed without obvious stress symptoms like accumulation of anthocyanins; they formed well-developed primary and lateral roots, and the speed of shoot development resembled that of pot-grown plants (**Figure 2B**). Altogether, *Dryas* species plants adapted very well to the hydroponic system tested. The absence of gel and soil substrates offers the opportunity to perform non-invasive observations of roots.

Sexual Propagation of *Dryas* Species

Dryas is a perennial plant genus. *D. drummondii* has been found to flower in its fifth year (Lawrence et al., 1967), indicating a long generation time that renders crossing experiments difficult. However, nodulation, plant growth and flowering processes in *Dryas* sp. seem to be extremely dependent on the
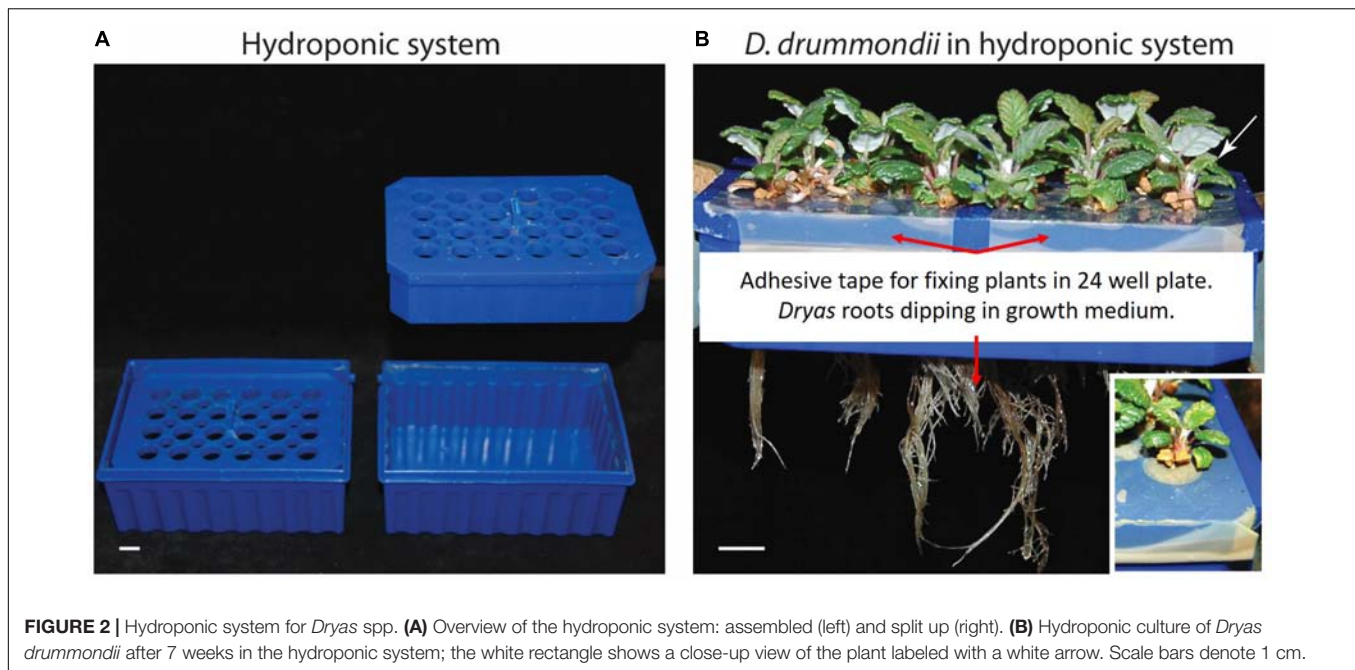
FIGURE 2 | Hydroponic system for *Dryas* spp. **(A)** Overview of the hydroponic system: assembled (left) and split up (right). **(B)** Hydroponic culture of *Dryas drummondii* after 7 weeks in the hydroponic system; the white rectangle shows a close-up view of the plant labeled with a white arrow. Scale bars denote 1 cm.

environment and on light quality and intensity (Kohls et al., 1994). Therefore, we attempted to reduce the generation time under greenhouse conditions.

In our study, flower and seed production did not occur when plants were grown at a distance of 2 m from standard high-pressure mercury vapor lamps (providing 90 μmol m^{-2} s^{-1} at the plant level) used in initial trials (fluorescent lamps were not tried for flowering as seedlings from both species were growing significantly more slowly under them than under mercury vapor lamps for the first 5 weeks after germination). However, when the plants were placed at a distance of 2 m under high-pressure sodium vapor lamps (150 μmol m^{-2} s^{-1}) for 16 h per day, flowering was induced (**Figures 3B–D**). High-pressure sodium lamps provide light with a richer emission in yellow-orange and a red/far red ratio shifted to the far red compared to standard high-pressure mercury vapor lamps and fluorescent lamps. The flowers produced seeds (**Figure 1A**) within less than a year after germination, whether the plants originated from cuttings or from sexual propagation. This is a far shorter generation time than the 5 years required for *D. drummondii* according to Lawrence et al. (1967). In the field, *Dryas* spp. flower primordia are formed during the summer, i.e., far in advance of flowering, which occurs in the next year shortly after snowmelt, with most individuals flowering within a month (Lawrence et al., 1967). This behavior suggests that the development of floral primordia and blooming depends on photoperiod or vernalization (or both). It is surprising that light from high-pressure sodium lamps, characterized by a lower red/far red ratio, leads to induction of flowering in an arctic/alpine species like *D. octopetala,* and of a species that has been described as extremely shade-sensitive *(D. drummondii*; Cooper, 1931). We observed that plants grew much better outdoors than in the glasshouse, but did not identify the limiting parameters. At any rate, changes of light period and

temperature might further speed up the induction of flowering and shorten the generation time.

Seed set under greenhouse conditions occurred at ca. 75% of all *D. drummondii* flowers and at ca. 65% of all *D. octopetala* flowers. In their natural habitat, *Dryas* species combine autogamy and allogamy; seed set is improved when insects are available for pollination (Kevan, 1975; Roslin et al., 2013; Tiusanen et al., 2016, 2019). In the greenhouse, while some insects were usually present, seeds would be formed by flowers covered with paper bags, indicating that all seed batches used gave rise to plants that could perform self-fertilization.

Manual pollination was performed in the greenhouse and in a botanical garden to obtain hybrids. While the large and open flowers of *D. octopetala* (**Figure 3D**) made manual pollination possible, *D. drummondii* flowers were never completely open during full bloom (**Figure 3C**). Therefore, crossings of female *D. octopetala* with male *D. drummondii* were attempted. However, these attempts were not successful. Flowers of species contain multiple stamens, and often, after manual pollination it turned out that not all of them had been removed.

Altogether, sexual reproduction of *Dryas* spp. was feasible in a laboratory context when sufficient light intensity of a suitable spectrum was provided.

Clonal Propagation of *Dryas* spp.

Given that *Dryas* spp. are partially allogamous and experimental studies require homogenous plant material, the use of *Dryas* as a model genus requires an easy protocol for vegetative propagation. In the wild, clonal growth of *Dryas* species enables individuals to persist and grow in extreme environments where sexual proliferation is often unsuccessful (Wookey et al., 1995), and where individual clones of *D. octopetala* commonly live for more than 100 years (Kihlman, 1890; Crawford, 1989). Thus, clonal
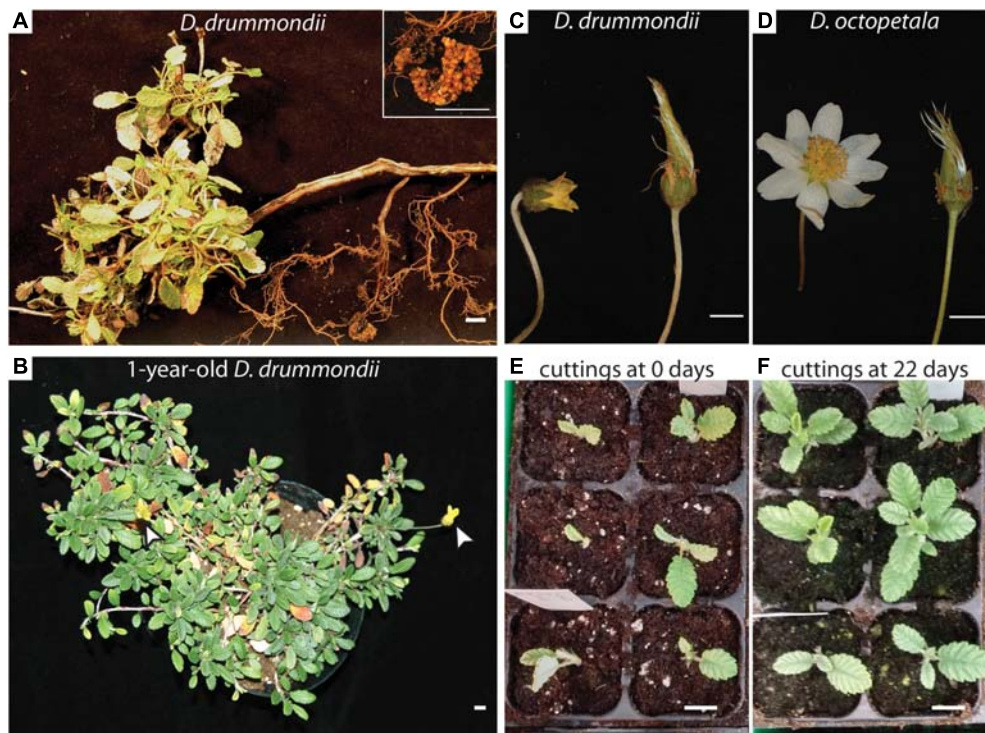
FIGURE 3 | Sexual and vegetative propagation of *Dryas* spp. **(A)** Part of a *Dryas drummondii* plant with a dense leaf rosette and lignified roots, coming from a woody shoot, bearing a nodule (shown magnified in the inset). The plant was collected in the Nymphenburg Botanical Garden of Munich. **(B)** 1-year-old *D. drummondii* plant flowering in the greenhouse under a higher-pressure sodium lamp (white arrows indicate flowers). **(C–D)** Flowers of *D. drummondii* **(C)** and *D. octopetala* **(D)** in full bloom and in fructification. **(E,F)** Cuttings of *D. drummondii* at day 0 **(E)** and after 22 days of growth **(F)**. Scale bars denote 1 cm.

propagation of *Dryas* species was expected to be easy. For clonal propagation by cuttings, three series of ca. 20 cuttings per species were grown in a greenhouse at different times of the year. Two to three cm of *Dryas* stems containing one node were transferred into moist soil (**Figure 3E**) in a small growth container with a transparent plastic lid for conservation of high humidity levels. Under these conditions, 65–95% of the *Dryas* cuttings developed roots within 3 weeks in the absence of hormonal treatments (**Figure 3F**). Once the shoots had successfully rooted, the plants were transferred into single pots and grown under standard greenhouse conditions. This easy protocol for vegetative propagation of *Dryas* species by cuttings in the glasshouse represents an important tool for performing experiments on a high number of plants that have the same genotype, and it obviates the requirement for seeds.
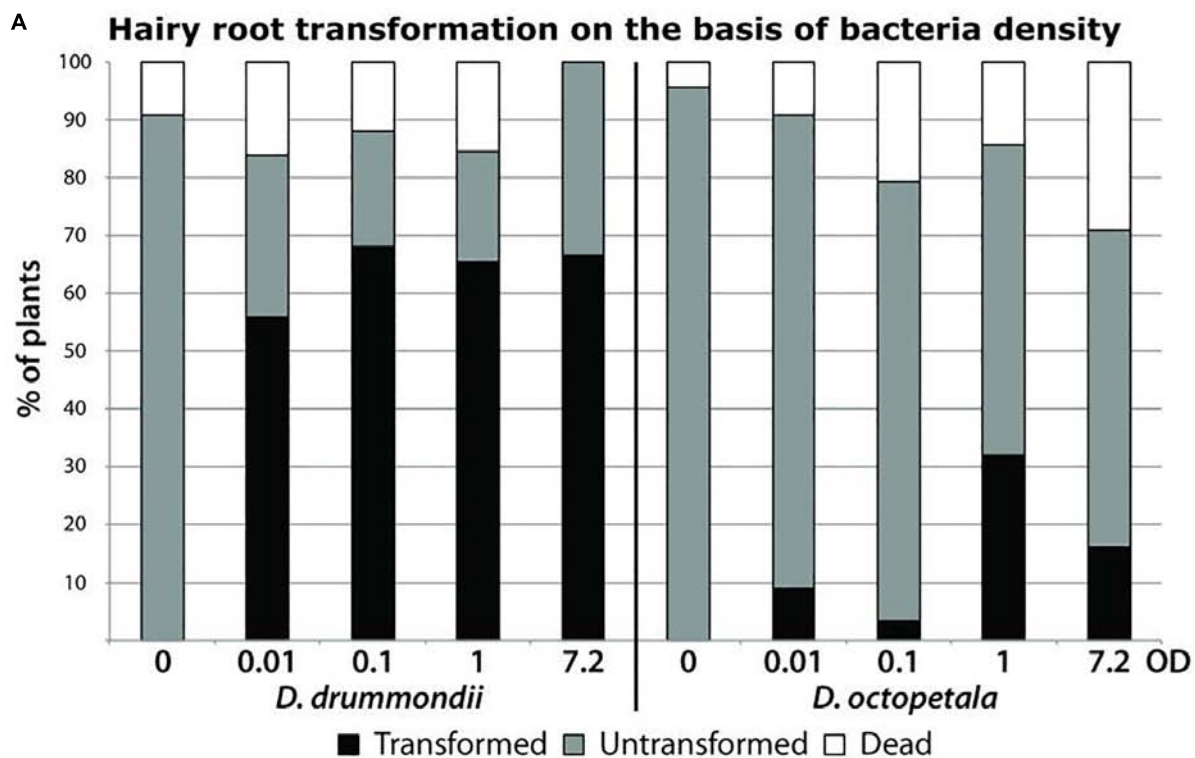
Hairy Root Transformation of *Dryas* spp.

For a model plant, a protocol for genetic modification is important in order to analyze the expression of marker gene promoter-reporter gene fusions, or to perform reverse genetics. Hairy root transformation mediated by *Agrobacterium rhizogenes* is the most commonly used technique to introduce chimeric constructs into plant roots. The fact that this method does not transform the shoot is no hindrance to the study of root symbioses; *A. rhizogenes*-mediated hairy root transformation is routinely used not only in the model legumes *Lotus corniculatus*

var. *japonicus* (Díaz et al., 2005) and *Medicago truncatula* (Boisson-Dernier et al., 2001) but also for actinorhizal plants like *Datisca glomerata* (Markmann et al., 2008), *Casuarina glauca* (Diouf et al., 1995) and (Imanishi et al., 2011) and non-FaFaCuRo plants like tomato (Ron et al., 2014).

In order to develop a hairy root transformation protocol, we inoculated axenically grown *Dryas* spp. seedlings with *A. rhizogenes* at different cell densities. 5 weeks after transformation, the composite plants on plate were evaluated. For *D. drummondii*, a transformation efficiency of 55–70% was obtained under all conditions tested, while *D. octopetala* plants died more frequently in response to infection with *A. rhizogenes*. The use of higher bacterial densities had a negative effect on plant survival, while lower bacterial densities reduced transformation efficiency. Here, the best compromise between low mortality and transformation rate for *D. octopetala* was observed when the *A. rhizogenes* suspension was adjusted to an OD_{600} of 1. However, the transformation efficiency was still low with only 30% (**Figure 4A**). The experiment was repeated three times using an *A. rhizogenes* suspension adjusted to an OD_{600} of 1 on ca. 70 seedlings per species. In all cases, the results were the same: 55–70% transformation for *D. drummondii*, maximally 30% transformation for *D. octopetala*.

Previous studies have shown that hairy roots induced by different bacterial strains can vary in morphology and production of secondary metabolites (Thwe et al., 2016); it was also
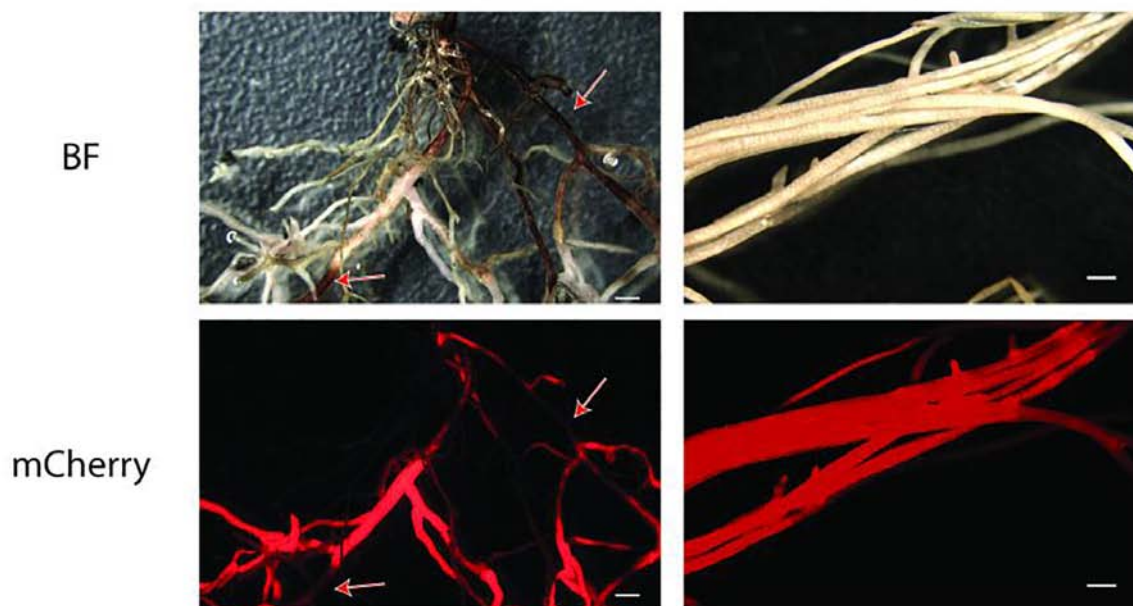
FIGURE 4 | *Agrobacterium rhizogenes*-mediated transformation of *Dryas* spp. **(A)** Success rates of hairy root transformation of *D. drummondii* and *D. octopetala* depended on the bacteria density. Percentages of dead (white boxes), surviving untransformed (gray boxes) and transformed (black boxes) root systems were determined 5 weeks after transformation on plants grown in Petri dishes. Transformation was determined based on mCherry fluorescence. **(B)** Visualization of the mCherry transformation marker of *D. drummondii* hairy roots after 7 weeks of growth in Weck Jars containing sand:vermiculite (left panel) vs. growth in the hydroponic system (right panel). Red arrows point at lignified part of the roots. BF = bright field; mCherry = mCherry fluorescence. Scale bars denote 1 mm.
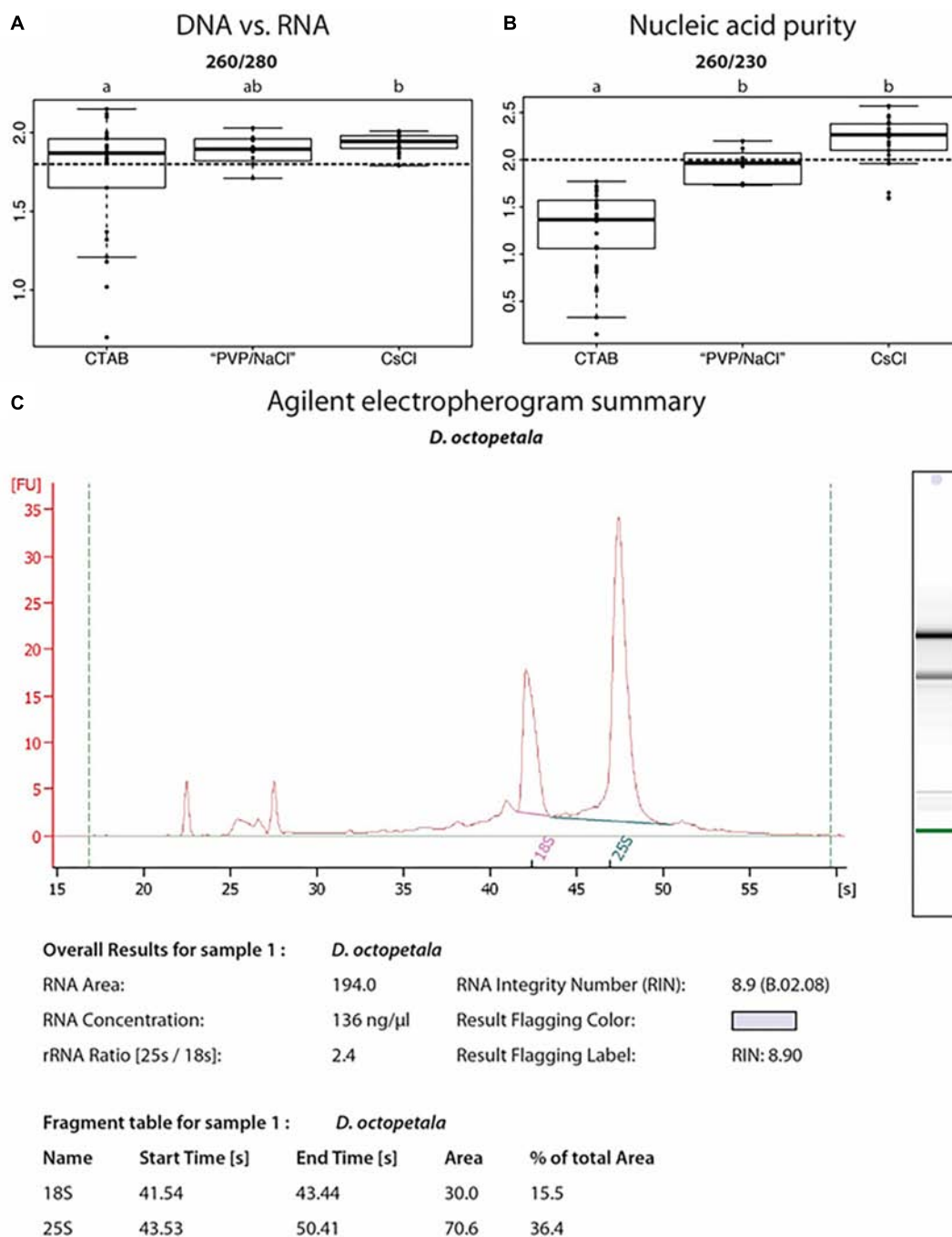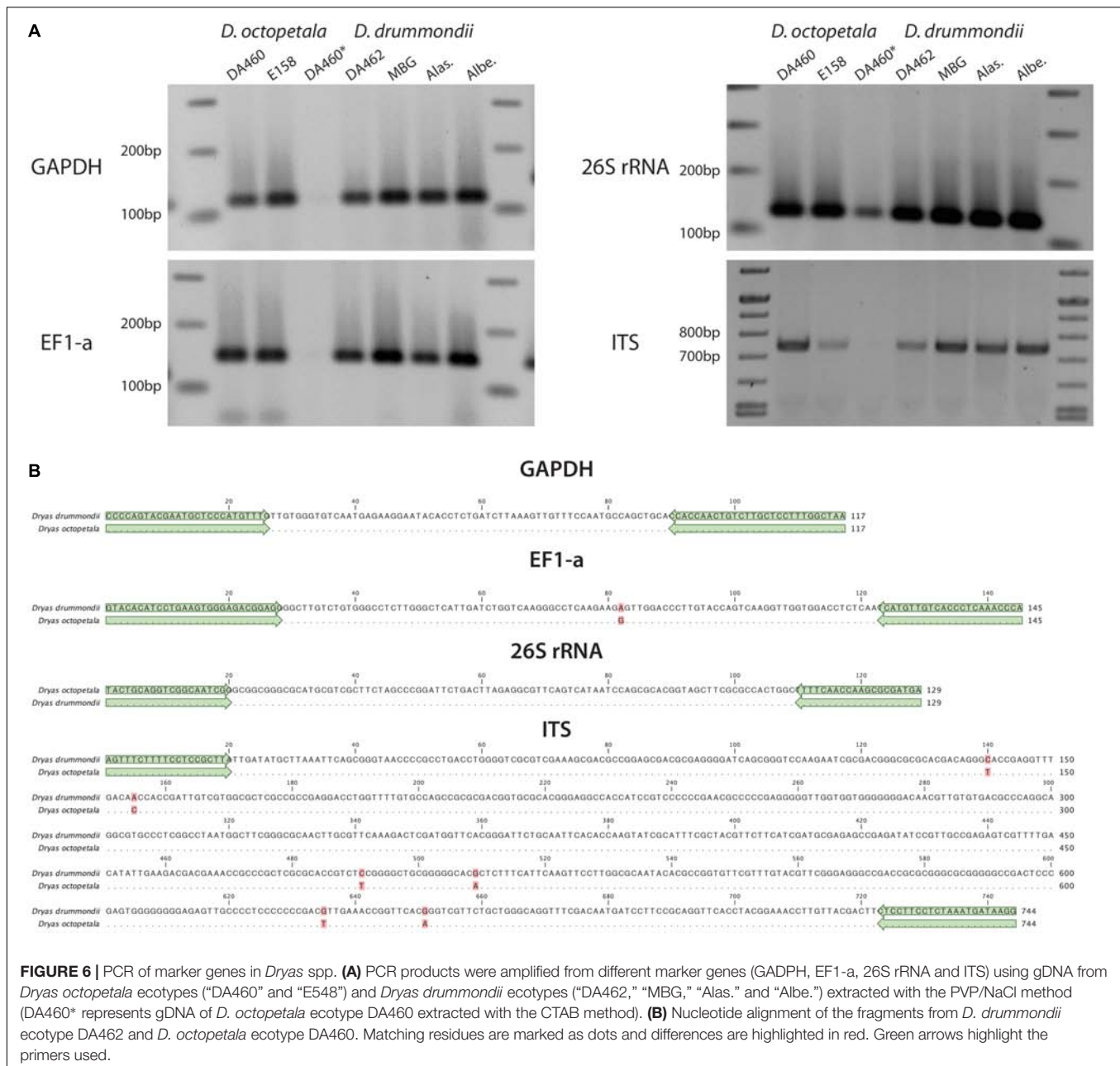
FIGURE 5 | Nucleic acid extractions from *Dryas* spp. **(A,B)** Ratios of UV absorbance at 260 nm vs. 280 or 230 nm, from DNA samples isolated from *Dryas drummondii* and *Dryas octopeatala* using three different methods: the classical CTAB method = "CTAB"; an adapted CTAB method for difficult plants = "PVP/NaCl" and a method involving a Caesium chloride gradient centrifugation = "CsCl." An OD_{260}/OD_{280} ratio **(A)** for nucleic acids vs. protein of at least 1.8 (dashed line) is generally accepted as denoting "pure DNA." OD_{260}/OD_{230} **(B)** values for nucleic acids vs. polysaccharides should be higher than 2.0 (dashed line; Green and Sambrook, 2012). All DNA isolations were performed on 30 biological replicates per method. **(C)** Agilent Bioanalyzer electropherogram analysis of RNA isolated from *D. octopetala*, showing RNA integrity as determined by an RNA Integrity Number (RIN) of 8.9.

shown that plant defense reactions, phytohormone signaling and secondary metabolism could be affected by high expression levels of the agrobacterial *rolB* gene (Bulgakov et al., 2018). Thus, the difference in the reactions of two closely related species to the same *A. rhizogenes* strain is interesting. At any rate, since only one *A. rhizogenes* strain was used in this study, the use of other strains might leave room for further optimization of hairy root transformation of *D. octopetala*.

FIGURE 6 | PCR of marker genes in *Dryas* spp. **(A)** PCR products were amplified from different marker genes (GADPH, EF1-a, 26S rRNA and ITS) using gDNA from *Dryas octopetala* ecotypes ("DA460" and "E548") and *Dryas drummondii* ecotypes ("DA462," "MBG," "Alas." and "Albe.") extracted with the PVP/NaCl method (DA460* represents gDNA of *D. octopetala* ecotype DA460 extracted with the CTAB method). **(B)** Nucleotide alignment of the fragments from *D. drummondii* ecotype DA462 and *D. octopetala* ecotype DA460. Matching residues are marked as dots and differences are highlighted in red. Green arrows highlight the primers used.

Up to 7 weeks after transfer to pots or to the hydroponic system, transgenic roots showed healthy growth and expressed the transformation marker mCherry driven by the ubiquitin promoter (**Figure 4B**). However, roots growing in particle substrates such as sand:vermiculite, developed sections with increased lignification, hence more autofluorescence and opacity. This led to the quenching of the mCherry signal as highlighted by the red arrows in **Figure 4B**. In contrast, when plants were grown in the hydroponic system, lignification was less pronounced. Thus, the hydroponic system is well suitable for the observation of fluorescent proteins in *Dryas* spp. hairy roots.

The ability to clone *Dryas* spp. genes combined with the capacity to introduce chimeric constructs into root systems opens the possibility to study *Dryas* genetics in depth, allowing cross-species complementation, as well as transient expression, protein localization, and reverse genetics using CRISPR/Cas or RNAi methods.

Nucleic Acid Extraction From *Dryas*

For molecular biological studies, DNA and RNA have to be isolated with high purity, integrity and yield to be used for sequencing or reverse transcription, respectively. This was particularly challenging since the woody nature of *Dryas* species and the composition of the leaves adapted to harsh environmental conditions led to the presence of contaminants interfering with nucleic acid extraction protocols.

We tested different DNA extraction protocols on *D. drummondii* and *D. octopetala,* performing at least 30 extractions per method. DNA isolated from *Dryas* spp. with classical CTAB extraction protocols had an UV absorbance ratio at 260/280 of *ca.* 1.8, but the 260/230 ratio was always below 1.8, indicating polysaccharide contamination (**Figures 5A,B**). Several established DNA extraction methods were tested (**Supplementary Table S1**), but none of them led to a yield and purity sufficient for robust PCRs and *de novo* whole genome sequencing. However, a CTAB protocol adapted for recalcitrant plant material (Khanuja et al., 1999; "PVP/NaCl") by addition of PVP, followed by a high salt lysis buffer and extraction with chloroform:isoamyl alcohol (24:1, v/v), resulted in good quality DNA suitable for PCRs with reproducible results (**Figure 5B**). Yet, the DNA yield and quality required for genome sequencing was so far only achieved using a modified Dellaporta et al. (1983) protocol followed by a CsCl gradient centrifugation as described by Ribeiro et al. (1995). The DNA extracted using this last method was used for *de novo* whole genome sequencing performed in collaboration with the Beijing Genomics Institute (BGI, China). The first version of the *D. drummondii* genome was used in a phylogenomic comparison study by Griesmann et al. (2018).

The Spectrum[TM] Plant Total RNA Kit (Sigma-Aldrich) was used in order to extract RNA from different organs of *Dryas* spp. When Polyclar AT was added during the grinding step for recalcitrant samples (e.g., mature leaves and lignified roots), this method resulted in RNA of suitable integrity and purity for the performance of reverse transcription-quantitative PCR as indicated by the RNA integrity number (**Figure 5C**). RNA extracted from roots, leaves and seedlings following this method was used by the BGI in order to assist gene prediction for the *D. drummondii* genome (Griesmann et al., 2018). The transcripts were mapped to the protein-coding gene models, identified using the MAKER-P pipeline (version 2.31; Campbell et al., 2014), in order to obtain gene characteristics (size and number of exons/introns per gene, distribution of genes, features of splicing sites, etc.).

All method comparisons are summarized in **Supplementary Table S1**.

PCR Amplification of *Dryas* spp. gDNA Fragments Using the *D. drummondii* Genome for Primer Design

We tested the suitability of the DNA preparations resulting from different protocols as templates for PCR. Based on the published *D. drummondii* genome (Griesmann et al., 2018). Primers were designed based on the first version of the *D. drummondii* genome. The targets were regions in the internal transcribed spacer (ITS) of nuclear ribosomal DNA, 26S ribosomal RNA (26S rRNA), glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and the

elongation factor 1-alpha (EF1-a). Using *D. octopetala* gDNA as template, fragments were amplified and sequenced as well. The amplification confirms that the DNA preparations were of sufficient quality, while high sequence conservation in these regions highlights the similarity between *D. drummondii* and *D. octopetala*. Indeed, the size of amplicons of both species were similar and their sequences presented few single nucleotide polymorphisms (**Figure 6**).

Dryas as Model Genus for the Rosaceae

With the basic but indispensable procedures and protocols for cultivation, vegetative and sexual propagation, hairy root transformation of and nucleic acid isolation from *Dryas* spp. described in this study, *Dryas* emerges as a new model genus to study important traits associated with survival in arctic and alpine conditions, including the formation of root symbioses with bacteria and ectomycorrhizal fungi.

AUTHOR CONTRIBUTIONS

KP: proposal of *Dryas* as a promising model system to study root symbioses. MP, BB-P, and KP: conceptualization. BB-P, AS, and KP: methodology. JF: establishment of *Dryas* hairy root transformation. KP and BB-P: high quality nucleic acid extraction. BB-P and AS: visualization. BB-P: writing – original draft. BB-P, AS, MP, and KP: writing – review and editing. MP: funding acquisition. MP: supervision of AS and JF. KP and MP: supervision of BB-P.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.00661/full#supplementary-material

REFERENCES

Becking, J. (1970). Frankiaceae fam. nov. (*Actinomycetales*) with one new combination and six new species of the genus *Frankia* Brunchorst 1886, 174. *Int. J. Syst. Evol. Microbiol.* 20, 201–220. doi: 10.1099/00207713-20-2-201

Binder, A., Lambert, J., Morbitzer, R., Popp, C., Ott, T., Lahaye, T., et al. (2014). A modular plasmid assembly kit for multigene expression, gene silencing and silencing rescue in plants. *PLoS One* 9:e88218. doi: 10.1371/journal.pone.0088218

Bjorbækmo, M. F. M., Carlsen, T., Brysting, A., Vrålstad, T., Høiland, K., Ugland, K. I., et al. (2010). High diversity of root associated fungi in both alpine

and arctic *Dryas octopetala*. *BMC Plant Biol.* 10:244. doi: 10.1186/1471-2229-10-244

Bliss, L. (1958). Seed germination in arctic and alpine species. *Arctic* 11, 180–188. doi: 10.1002/ece3.3539

Böcher, T. W., Holmen, K., and Jakobsen, K. (1968). *The Flora of Greenland*. Copenhagen: P. Haase & Son Publishers.

Boisson-Dernier, A., Chabaud, M., Garcia, F., Bécard, G., Rosenberg, C., and Barker, D. G. (2001). *Agrobacterium rhizogenes*-transformed roots of *Medicago truncatula* for the study of nitrogen-fixing and endomycorrhizal symbiotic associations. *Mol. Plant Microbe Interact.* 14, 695–700. doi: 10.1094/mpmi.2001.14.6.695

Botnen, S., Vik, U., Carlsen, T., Eidesen, P. B., Davey, M. L., and Kauserud, H. (2014). Low host specificity of root-associated fungi at an Arctic site. *Mol. Ecol.* 23, 975–985. doi: 10.1111/mec.12646

Brunner, I., Frey, B., Hartmann, M., Zimmermann, S., Graf, F., Suz, L. M., et al. (2017). Ecology of alpine macrofungi-combining historical with recent data. *Front. Microbiol.* 8:2066. doi: 10.3389/fmicb.2017.02066

Bulgakov, V. P., Vereshchagina, Y. V., Bulgakov, D. V., Veremeichik, G. N., and Shkryl, Y. N. (2018). The *rolB* plant oncogene affects multiple signaling protein modules related to hormone signaling and plant defense. *Sci. Rep.* 8:2285. doi: 10.1038/s41598-018-20694-6

Campbell, M. S., Law, M., Holt, C., Stein, J. C., Moghe, G. D., Hufnagel, D. E., et al. (2014). MAKER-P: a tool kit for the rapid creation, management, and quality control of plant genome annotations. *Plant Physiol.* 164, 513–524. doi: 10.1104/pp.113.230144

Cheng, T., Xu, C., Lei, L., Li, C., Zhang, Y., and Zhou, S. (2016). Barcoding the kingdom plantae: new PCR primers for ITS regions of plants with improved universality and specificity. *Mol. Ecol. Res.* 16, 138–149. doi: 10.1111/1755-0998.12438

Clemow, S. R., Clairmont, L., Madsen, L. H., and Guinel, F. C. (2011). Reproducible hairy root transformation and spot-inoculation methods to study root symbioses of pea. *Plant Methods* 7:46. doi: 10.1186/1746-4811-7-46

Cooper, W. S. (1931). A third expedition to Glacier Bay. Alaska. *Ecology* 12, 61–95. doi: 10.2307/1932934

Cosme, M., Fernández, I., Van der Heijden, M. G. A., and Pieterse, C. M. J. (2018). Non-mycorrhizal plants: The exceptions that prove the rule. *Trends Plant Sci.* 23, 577–587. doi: 10.1016/j.tplants.2018.04.004

Crawford, R. M. M. (1989). *Studies in Plant Survival: Ecological Case Histories of Plant Adaptation to Adversity*. Oxford: Blackwell Scientific Publications.

Crocker, R. L., and Major, J. (1955). Soil development in relation to vegetation and surface age at Glacier Bay, Alaska. *J. Ecol.* 43, 427–448.

Delaux, P.-M., Séjalon-Delmas, N., Bécard, G., and Ané, J.-M. (2013). Evolution of the plant–microbe symbiotic 'toolkit'. *Trends Plant Sci.* 18, 298–304. doi: 10.1016/j.tplants.2013.01.008

Dellaporta, S. L., Wood, J., and Hicks, J. B. (1983). A plant DNA minipreparation: version II. *Plant Mol. Biol. Rep.* 1, 19–21. doi: 10.1007/bf02712670

Díaz, C. L., Grønlund, M., Schlaman, H. R., and Spaink, H. P. (2005). "Induction of hairy roots for symbiotic gene expression studies," in *Lotus japonicus Handbooks*, ed. A. J. Márquez (Dordrecht: Springer), 261–277. doi: 10.1007/1-4020-3735-x_26

Diouf, D., Gherbi, H., Prin, Y., Franche, C., Duhoux, E., and Bogusz, D. (1995). Hairy root nodulation of *Casuarina glauca*: a system for the study of symbiotic gene expression in an actinorhizal tree. *Mol. Plant Microbe Interact.* 8, 532–537.

Doyle, J. J., and Doyle, J. L. (1987). Genomic plant DNA preparation from fresh tissue-CTAB method. *Phytochem. Bull.* 19, 11–15.

Eichel, J., Corenblit, D., and Dikau, R. (2016). Conditions for feedbacks between geomorphic and vegetation dynamics on lateral moraine slopes: a biogeomorphic feedback window. *Earth Surf. Process. Landf.* 41, 406–419. doi: 10.1002/esp.3859

Eichel, J., Draebing, D., Klingbeil, L., Wieland, M., Eling, C., Schmidtlein, S., et al. (2017). Solifluction meets vegetation: the role of biogeomorphic feedbacks for turf-banked solifluction lobe development. *Earth Surf. Process. Landf.* 42, 1623–1635. doi: 10.1002/esp.4102

Fåhraeus, G. (1957). The infection of clover root hairs by nodule bacteria studied by a simple glass slide technique. *J. Gen. Microbiol.* 16, 374–381. doi: 10.1099/00221287-16-2-374

Fitter, A. H., and Parsons, W. F. J. (1987). Changes in phosphorus and nitrogen availability on recessional moraines of the Athabasca Glacier. Alberta. *Can. J. Bot.* 65, 210–213. doi: 10.1139/b87-028

Gamborg, O. L., Miller, R. A., and Ojima, K. (1968). Nutrient requirements of suspension cultures of soybean root cells. *Exp. Cell Res.* 50, 151–158. doi: 10.1016/0014-4827(68)90403-5

Geurts, R., Xiao, T. T., and Reinhold-Hurek, B. (2016). What does it take to evolve a nitrogen-fixing endosymbiosis? *Trends Plant Sci.* 21, 199–208. doi: 10.1016/j.tplants.2016.01.012

Gillespie, M. A. K., Baggesen, N., and Cooper, E. J. (2016). High arctic flowering phenology and plant–pollinator interactions in response to delayed snow melt and simulated warming. *Environ. Res. Lett.* 11:115006. doi: 10.1088/1748-9326/11/11/115006

Green, M. R., and Sambrook, J. (2012). *Molecular Cloning: A Laboratory Manual*, 4th Edn. New York, NY: Cold Spring Harbor Laboratory Press.

Griesmann, M., Chang, Y., Liu, X., Song, Y., Haberer, G., Crook, M. B., et al. (2018). Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. *Science* 361:eaat1743. doi: 10.1126/science.aat1743

Hemm, M. R., Rider, S. D., Ogas, J., Murry, D. J., and Chapple, C. (2004). Light induces phenylpropanoid metabolism in *Arabidopsis* roots. *Plant J.* 38, 765–778. doi: 10.1111/j.1365-313x.2004.02089.x

Hoagland, D. R., and Arnon, D. I. (1950). *The Water-Culture Method For Growing Plants Without Soil. Circular California Agricultural Experiment Station 347*, 2nd edn. Berkeley: University of California.

Hultén, E. (1968). *Flora of Alaska and Neighboring Territories*. Stanford, CA: Stanford University Press.

Imanishi, L., Vayssières, A., Franche, C., Bogusz, D., Wall, L., and Svistoonoff, S. (2011). Transformed hairy roots of *Discaria trinervis*: a valuable tool for studying actinorhizal symbiosis in the context of intercellular infection. *Mol. Plant Microbe Interact.* 24, 1317–1324. doi: 10.1094/MPMI-03-11-0078

Jung, S., Ficklin, S. P., Lee, T., Cheng, C.-H., Blenda, A., Zheng, P., et al. (2013). The genome database for Rosaceae (GDR): year 10 update. *Nucleic Acids Res.* 42, D1237–D1244. doi: 10.1093/nar/gkt1012

Kevan, P. G. (1975). Sun-tracking solar furnaces in high arctic flowers: significance for pollination and insects. *Science* 189, 723–726. doi: 10.1126/science.189.4204.723

Khanuja, S. P., Shasany, A. K., Darokar, M. P., and Kumar, S. (1999). Rapid isolation of DNA from dry and fresh samples of plants producing large amounts of secondary metabolites and essential oils. *Plant Mol. Biol. Rep.* 17:74.

Kihlman, A. O. (1890). *Pflanzenbiologische Studien aus Russisch Lappland: ein Beitrag zur Kenntniss der regionalen Gliederung an der polaren Waldgrenze*, Vol. 3. Helsingfors: Weilin & Göös Buchdruckerei AG.

Kohls, S. J., Thimmapuram, J., Buschena, C. A., Paschke, M. W., and Dawson, J. O. (1994). Nodulation patterns of actinorhizal plants in the family Rosaceae. *Plant Soil* 162, 229–239. doi: 10.1007/bf01347710

Lambers, H., Bishop, J. G., Hopper, S. D., Laliberte, E., and Zuniga-Feest, A. (2012). Phosphorus-mobilization ecosystem engineering: the roles of cluster roots and carboxylate exudation in young P-limited ecosystems. *Ann. Bot.* 110, 329–348. doi: 10.1093/aob/mcs130

Lawrence, D. B., Schoenike, R., Quispel, A., and Bond, G. (1967). The role of *Dryas drummondii* in vegetation development following ice recession at Glacier Bay, Alaska, with special reference to its nitrogen fixation by root nodules. *J. Ecol.* 55, 793–813.

Markham, J. H. (2009). Does *Dryas integrifolia* fix nitrogen? *Botany* 87, 1106–1109. doi: 10.1139/b09-071

Markmann, K., Giczey, G., and Parniske, M. (2008). Functional adaptation of a plant receptor-kinase paved the way for the evolution of intracellular root symbioses with bacteria. *PLoS Biol.* 6:e68. doi: 10.1371/journal.pbio.0060068

McGraw, J. B., Turner, J. B., Chandler, J. L., and Vavrek, M. C. (2014). Disturbances as hot spots of ecotypic variation: a case study with *Dryas octopetala*. *Arct. Antarct. Alp. Res.* 46, 542–547. doi: 10.1657/1938-4246-46.3.542

Melville, L., Massicotte, H., Ackerley, C., and Peterson, R. (1988). An ultrastructural study of modifications in *Dryas intergrifolia* and *Hebeloma cylindrosporum* during ectomycorrhiza formation. *Botan. Gazette* 149, 408–418. doi: 10.1086/337733

Murashige, T., and Skoog, F. (1962). A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiol. Plant.* 15, 473–497. doi: 10.1111/j.1399-3054.1962.tb08052.x

Newcomb, W. (1981). Fine structure of the root nodules of *Dryas drummondii* Richards (*Rosaceae*). *Can. J. Bot.* 59, 2500–2514. doi: 10.1139/b81-300

Nichols, G. E. (1934). The influence of exposure to winter temperatures upon seed germination in various native American plants. *Ecology* 15, 364–373. doi: 10.2307/1932351

Normand, P., Nguyen, T. V., Battenberg, K., Berry, A. M., Vanden Heuvel, B., Fernandez, M. P., et al. (2017). Species proposal of *Candidatus* Frankia californiensis, the uncultured symbiont in nitrogen-fixing root nodules of a phylogenetically broad group of hosts endemic to western North America. *Int. J. Syst. Evol. Microbiol.* 67, 3706–3715. doi: 10.1099/ijsem.0.002147

Packer, J. G. (1994). A mountain avens, *Dryas* x sündermannii Kellerer ex Sündermann, in Alberta. *Can. Field Natural.* 108, 77–79.

Panchen, Z. A., and Gorelick, R. (2017). Prediction of Arctic plant phenological sensitivity to climate change from historical records. *Ecol. Evol.* 7, 1325–1338. doi: 10.1002/ece3.2702

Parniske, M. (2018). Uptake of bacteria into living plant cells, the unifying and distinct feature of the nitrogen-fixing root nodule symbiosis. *Curr. Opin. Plant Biol* 44, 164–174. doi: 10.1016/j.pbi.2018.05.016

Pawlowski, K., and Bisseling, T. (1996). Rhizobial and actinorhizal symbioses: what are the shared features? *Plant Cell* 8:1899. doi: 10.1105/tpc.8.10.1899

Pawlowski, K., and Demchenko, K. N. (2012). The diversity of actinorhizal symbiosis. *Protoplasma* 249, 967–979. doi: 10.1007/s00709-012-0388-4

Perrine-Walker, F., Gherbi, H., Imanishi, L., Hocher, V., Ghodhbane-Gtari, F., Lavenus, J., et al. (2011). Symbiotic signaling in actinorhizal symbioses. *Curr. Protein Pept. Sci.* 12, 156–164.

Philipp, M., and Siegismund, H. R. (2003). What can morphology and isozymes tell us about the history of the *Dryas integrifolia–octopetala* complex? *Mol. Ecol.* 12, 2231–2242. doi: 10.1046/j.1365-294x.2003.01875.x

Pimprikar, P., Carbonnel, S., Paries, M., Katzer, K., Klingl, V., Bohmer, M. J., et al. (2016). A CCaMK-CYCLOPS-DELLA complex activates transcription of RAM1 to regulate arbuscule branching. *Curr. Biol.* 26, 987–998. doi: 10.1016/j.cub.2016.01.069

Porslid, A. E. (1947). The genus *Dryas* in North America. *Can. Field Nat.* 61, 175–192.

Potter, D., Eriksson, T., Evans, R. C., Oh, S., Smedmark, J., Morgan, D. R., et al. (2007). Phylogeny and classification of *Rosaceae*. *Plant Syst. Evol.* 266, 5–43.

Ribeiro, A., Akkermans, A., van Kammen, A., Bisseling, T., and Pawlowski, K. (1995). A nodule-specific gene encoding a subtilisin-like protease is expressed in early stages of actinorhizal nodule development. *Plant Cell* 7, 785–794. doi: 10.1105/tpc.7.6.785

Ron, M., Kajala, K., Pauluzzi, G., Wang, D., Reynoso, M. A., Zumstein, K., et al. (2014). Hairy root transformation using *Agrobacterium rhizogenes* as a tool for exploring cell type-specific gene expression and function using tomato as a model. *Plant Physiol.* 166, 455–569. doi: 10.1104/pp.114.239392

Roslin, T., Wirta, H., Hopkins, T., Hardwick, B., and Várkonyi, G. (2013). Indirect interactions in the High Arctic. *PLoS One* 8:e67367. doi: 10.1371/journal.pone.0067367

Ryberg, M., Larsson, E., and Molau, U. (2009). Ectomycorrhizal diversity on *Dryas octopetala* and *Salix reticulata* in an alpine cliff ecosystem. *Arct. Antarct. Alp. Res.* 41, 506–514. doi: 10.1657/1938-4246-41.4.506

Skrede, I., Eidesen, P. B., Portela, R. P., and Brochmann, C. (2006). Refugia, differentiation and postglacial migration in arctic-alpine Eurasia, exemplified by the mountain avens (*Dryas octopetala* L.). *Mol. Ecol.* 15, 1827–1840. doi: 10.1111/j.1365-294x.2006.02908.x

Soltis, D. E., Soltis, P. S., Morgan, D. R., Swensen, S. M., Mullin, B. C., Dowd, J. M., et al. (1995). Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. *Proc. Natl. Acad. Sci. U.S.A.* 28, 2647–2651. doi: 10.1073/pnas.92.7.2647

Stougaard, J., Abildsten, D., and Marcker, K. A. (1987). The *Agrobacterium rhizogenes* pRi TL-DNA segment as a gene vector system for transformation of plants. *Mol. Gen. Genet.* 207, 251–255. doi: 10.1007/bf00331586

Thwe, A., Arasu, M. V., Li, X., Park, C. H., Kim, S. J., Al-Dhabi, N. A., et al. (2016). Effect of different *Agrobacterium rhizogenes* strains on hairy root induction and phenylpropanoid biosynthesis in tartary buckwheat (*Fagopyrum tataricum* Gaertn). *Front. Microbiol.* 7:318. doi: 10.3389/fmicb.2016.00318

Tiusanen, M., Hebert, P. D., Schmidt, N. M., and Roslin, T. (2016). One fly to rule them all—muscid flies are the key pollinators in the Arctic. *Proc. R. Soc. B* 283:20161271. doi: 10.1098/rspb.2016.1271

Tiusanen, M., Huotari, T., Hebert, P. D. N., Andersson, T., Asmus, A., Bety, J., et al. (2019). Flower-visitor communities of an arcto-alpine plant–global patterns in species richness, phylogenetic diversity and ecological functioning. *Mol. Ecol.* 28, 318–335. doi: 10.1111/mec.14932

Tremblay, N., and Schoen, D. (1999). Molecular phylogeography of *Dryas integrifolia*: glacial refugia and postglacial recolonization. *Mol. Ecol.* 8, 1187–1198. doi: 10.1046/j.1365-294x.1999.00680.x

Vallée, G. C., Muñoz, D. S., and Sankoff, D. (2016). Economic importance, taxonomic representation and scientific priority as drivers of genome sequencing projects. *BMC Genomics* 17:782. doi: 10.1186/s12864-016-3100-9

Van Nguyen, T., and Pawlowski, K. (2017). "*Frankia* and actinorhizal plants: symbiotic nitrogen fixation," in *Rhizotrophs: Plant Growth Promotion to Bioremediation*, ed. S. Mehnaz (Singapore: Springer), 237–261. doi: 10.1007/978-981-10-4862-3_12

van Velzen, R., Holmer, R., Bu, F., Rutten, L., van Zeijl, A., Liu, W., et al. (2018). Comparative genomics of the nonlegume *Parasponia* reveals insights into evolution of nitrogen-fixing rhizobium symbioses. *Proc. Natl. Acad. Sci. U.S.A.* 115, E4700–E4709. doi: 10.1073/pnas.1721395115

Väre, H., Vestberg, M., and Eurola, S. (1992). Mycorrhiza and root-associated fungi in Spitsbergen. *Mycorrhiza* 1, 93–104. doi: 10.1007/bf00203256

Vitousek, P. M., Porder, S., Houlton, B. Z., and Chadwick, O. A. (2010). Terrestrial phosphorus limitation: mechanisms, implications, and nitrogen–phosphorus interactions. *Ecol. Appl.* 20, 5–15. doi: 10.1890/08-0127.1

Werner, G. D., Cornwell, W. K., Sprent, J. I., Kattge, J., and Kiers, E. T. (2014). A single evolutionary innovation drives the deep evolution of symbiotic N_2-fixation in angiosperms. *Nat. Commun.* 5:4087. doi: 10.1038/ncomms5087

West, R., Andrew, R., and Pettit, M. (1993). Taphonomy of plant remains on floodplains of tundra rivers, present and pleistocene. *New Phytol.* 123, 203–221. doi: 10.1111/j.1469-8137.1993.tb04546.x

Wookey, P., Robinson, C., Parsons, A., Welker, J., Press, M., Callaghan, T., et al. (1995). Environmental constraints on the growth, photosynthesis and reproductive development of *Dryas octopetala* at a high Arctic polar semi-desert. *Svalbard. Oecol.* 102, 478–489. doi: 10.1007/BF00341360

Xiang, Y., Huang, C.-H., Hu, Y., Wen, J., Li, S., Yi, T., et al. (2016). Evolution of *Rosaceae* fruit types based on nuclear phylogeny in the context of geological times and genome duplication. *Mol. Biol. Evol.* 34, 262–281.

Yurtsev, B. A. (1997). Analysis of evolutionary differentiation in some key arctic–alpine taxa: *Dryas, Oxytropis* sect. Arctobia and *Taraxacum* sect. Arctica. *Opera. Botan.* 132, 27–38.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read
for greatest visibility
and readership

FAST PUBLICATION
Around 90 days
from submission
to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative,
and constructive
peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers
acknowledged by name
on published articles

REPRODUCIBILITY OF RESEARCH
Support open data
and methods to enhance
research reproducibility

DIGITAL PUBLISHING
Articles designed
for optimal readership
across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics
track visibility across
digital media

EXTENSIVE PROMOTION
Marketing
and promotion
of impactful research

LOOP RESEARCH NETWORK
Our network
increases your
article's readership