



Gene Expression in Nitrogen-Fixing Symbiotic Nodule Cells in *Medicago truncatula* and Other Nodulating Plants

Peter Mergaert,^{a,1} Attila Kereszt,^b and Eva Kondorosi^b

^aInstitute for Integrative Biology of the Cell, UMR 9198, CEA, CNRS, Université Paris-Saclay, 91198 Gif-sur-Yvette, France

^bInstitute of Plant Biology, Biological Research Centre, 6726 Szeged, Hungary

ORCID IDs: 0000-0002-5919-7317 (P.M.); 0000-0002-4421-6554 (A.K.); 0000-0002-4065-8515 (E.K.)

Root nodules formed by plants of the nitrogen-fixing clade (NFC) are symbiotic organs that function in the maintenance and metabolic integration of large populations of nitrogen-fixing bacteria. These organs feature unique characteristics and processes, including their tissue organization, the presence of specific infection structures called infection threads, endocytotic uptake of bacteria, symbiotic cells carrying thousands of intracellular bacteria without signs of immune responses, and the integration of symbiont and host metabolism. The early stages of nodulation are governed by a few well-defined functions, which together constitute the common symbiosis-signaling pathway (CSSP). The CSSP activates a set of transcription factors (TFs) that orchestrate nodule organogenesis and infection. The later stages of nodule development require the activation of hundreds to thousands of genes, mostly expressed in symbiotic cells. Many of these genes are only active in symbiotic cells, reflecting the unique nature of nodules as plant structures. Although how the nodule-specific transcriptome is activated and connected to early CSSP-signaling is poorly understood, candidate TFs have been identified using transcriptomic approaches, and the importance of epigenetic and chromatin-based regulation has been demonstrated. We discuss how gene regulation analyses have advanced our understanding of nodule organogenesis, the functioning of symbiotic cells, and the evolution of symbiosis in the NFC.

INTRODUCTION

In response to nitrogen starvation and the presence of specific compatible nitrogen-fixing bacteria in the rhizosphere, plants of the nitrogen-fixing clade (NFC) in the Rosid clade form symbiotic organs on their roots known as nodules. Nodules are infected by a large population of these bacteria, which convert atmospheric nitrogen gas to ammonia. The ammonia is taken up by the plant to satisfy its nitrogen needs for growth. The NFC, which originated over 100 million years ago (MYA), is composed of four orders, the Fabales, Cucurbitales, Fagales, and Rosales (Soltis et al., 1995). Nodulation is widespread in the Leguminosae family of the Fabales (although not present in all legume species), while it is scattered among species of the three other orders. Leguminosae and species of the *Parasponia* genus of the Rosales are infected with phylogenetically diverse bacteria, collectively named rhizobia, belonging to the Rhizobiales of the alphaproteobacteria (α -rhizobia) or the Burkholderiales of the betaproteobacteria (β -rhizobia). The other nodulating genera, collectively known as actinorhizal plants, interact with nitrogen-fixing actinobacteria of the genus *Frankia*.

Although they are diverse, all rhizobia, as typical Gram-negative bacteria, have (broadly speaking) a similar type of cell envelope structure, which is a crucial attribute for effective nodule colonization (Haag et al., 2013). At least one other shared feature of rhizobia — with one reported exception (Giraud et al., 2007) — is the capacity to produce so-called Nod factors, the bacterial signals that trigger the nodulation program of the host plant via the activation of the common symbiosis signaling pathway (CSSP).

Nod factors are lipochitooligosaccharides with a characteristic molecular structure and are similar only to the Myc factors produced by mycorrhizal fungi (Oldroyd, 2013; Cope et al., 2019). The capacity to produce Nod factors is encoded by the nodulation or *nod* genes located on symbiotic islands or on mobile plasmids in the rhizobial genomes (Poole et al., 2018).

By contrast, *Frankia* are Gram-positive bacteria with a fundamentally different cell envelope from rhizobia. Although very little is known about the requirements of cell envelopes for symbiosis in *Frankia*, these envelope differences likely have profound effects on the colonization of nodules. The nodule-inducing factors produced by *Frankia* strains are currently unknown (Cissoko et al., 2018), although some particular *Frankia* strains possess genes homologous to the *nod* genes of rhizobia (Persson et al., 2015). It is thus conceivable that *Frankia* strains produce nodule-inducing signals that are structurally homologous to Nod factors.

The nodule can be viewed as a plant organ that relies on the integration or recruitment of existing processes (meristem formation, endoreduplication, polar growth, endocytosis and exocytosis, nitrogen, and sugar metabolism) in a new context, in conjunction with the deployment of seemingly completely new features (infection threads and uptake of bacteria, meristem organization, formation of new organellar structures, maintenance of bacteria inside plant cells and immune suppression and metabolic integration of a bacterial endosymbiont). Thus, depending on whether one considers the glass half-empty or half-full, nodules can be viewed as “novel” or “not-so-novel” plant organs. The latter point of view — particularly motivated by the finding that nodule formation is controlled by the CSSP, which is widely conserved in land plants that form symbiotic relationships with arbuscular mycorrhizal fungi — has led to the suggestion that engineering nodulation and symbiotic nitrogen-fixation to currently nonnodulating

¹ Address correspondence to peter.mergaert@i2bc.paris-saclay.fr.
www.plantcell.org/cgi/doi/10.1105/tpc.19.00494

plants would require a minimal set of genes and is therefore an attainable objective (Markmann and Parniske, 2009; Oldroyd and Dixon, 2014; Stokstad, 2016). However, nodule formation is accompanied by massive transcriptional reprogramming involving the activation and repression of hundreds of genes. Remarkably, this nodule transcriptional program is entirely different from the mycorrhizal transcriptional program controlled by the same CSSP. This highlights the uniqueness of nodules as plant organs and emphasizes the challenges faced when trying to transfer nodulation to nonlegume crops. Here we review what we have learned from transcriptome approaches aimed at characterizing the specific features of nodules, focusing mostly but not exclusively on the transcriptional reprogramming taking place during the formation of symbiotic nodule cells in the model legume, *Medicago truncatula*.

NODULE DEVELOPMENT AND MORPHOLOGY: NODULES ARE UNIQUE STRUCTURES

Nodules in different lineages of the NFC can differ from one another in terms of development, infection mechanisms, final structure, and metabolism. However, nodule organogenesis and functioning commonly calls upon several processes, which do not have equivalents in other plant developmental programs. Nodule formation in *M. truncatula* is among the best described (Figure 1). This process is governed by a number of unique signaling cascades that use inter-specific (bacterium to plant and vice versa) and intra-specific (plant to plant) signaling molecules including flavonoids (plant to bacterium) and Nod factors (bacterium to plant; Oldroyd, 2013; Figure 1A), Type III, IV, and VI secretion system effectors (bacterium to plant; Nelson and Sadowsky, 2015; Teulet et al., 2019), bacterial surface polysaccharides (exopolysaccharides, lipopolysaccharides; bacterium to plant; Haag et al., 2013), NCR (nodule-specific cysteine-rich) peptides (plant to bacterium; Kondorosi et al., 2013), phytohormones (intra-plant; Gamas et al., 2017; Schiessl et al., 2019), CLE and CEP peptides (intra-plant; Taleski et al., 2018), and small RNAs (intra-plant and bacterium to plant; Tsikou et al., 2018; Ren et al., 2019). Each of these signaling systems seems to be variations on a theme of widely used regulatory mechanisms in plants. For example, similar to Myc factors, Nod factors activate the CSSP (Figure 1A), which thus controls not only nodulation but also mycorrhization (Oldroyd, 2013). However, the output of the signaling in response to Nod factors and Myc factors is completely different. It is a long-lasting conundrum how this differential output is achieved. Other key signals of nodulation are used in plant development (e.g. phytohormones) or in biotic interactions (e.g. immunity) but with a unique output in nodulation. Thus, nodulation relies upon a unique rewiring of signaling pathways.

The nodule primordium and the persistent meristem of indeterminate *M. truncatula* nodules have features reminiscent of the formation of lateral roots and the root apical meristem (Couzigou et al., 2012; Xiao et al., 2014; Franssen et al., 2015), including overlapping genetic and hormonal control mechanisms (Schiessl et al., 2019). However, the root apical meristem is organized in a different manner and leads to the formation of one central vascular bundle, while in legumes, the nodule meristem generates several peripheral vascular bundles (Figure 1B).

Nevertheless, actinorhizal and *Parasponia* nodules have a central vasculature (Downie, 2014; Behm et al., 2014). The other cell types derived from the nodule meristem are also fundamentally different from root cells. Endoreduplication drives the differentiation of postmeristematic cells into symbiotic nodule cells, which become infected by rhizobium bacteria (Kondorosi et al., 2013). Endoreduplication is common in plants and can be observed in differentiating root cells and other plant tissues in a species-specific manner, as well as in many other biotrophic interactions in plants (Edgar et al., 2014; Wildermuth et al., 2017). In nodules, endoreduplication is tightly coupled to infection and the passage of so-called infection threads (van Brussel et al., 1992; Yang et al., 1994; Suzuki et al., 2014). Determinate nodules, which form in some legume clades, differ fundamentally from the indeterminate nodules of *M. truncatula*, because they do not develop a meristem. Nevertheless, cells of determinate nodules also undergo endoreduplication after infection with rhizobia (González-Sama et al., 2006). Consequently, and contrary to meristem-containing indeterminate nodules, these nodules have a determinate growth, which is limited by early cell divisions during the primordium stage and the subsequent cell enlargement of rhizobium-infected cells driven by their endoreduplication. Well-studied examples of determinate nodules are formed by *Lotus japonicus* and soybean (*Glycine max*).

Infection threads are tubular structures containing infecting rhizobia that conduct these bacteria inside the nodule tissues (Figure 1C). Infection thread formation initiates in growing root hairs due to the activation of the CSSP by rhizobial Nod factors. These structures share some similarities with pollen tubes and root hairs, which both display polar growth, as do infection threads. The polar growth of infection threads, pollen tubes, and root hairs indeed relies on common molecular mechanisms (Yang, 2008; Ke et al., 2012; Kiirika et al., 2012; Montiel et al., 2012; Morieri et al., 2013). However, contrary to pollen tubes and root hairs, which are autonomous cells, infection threads are trans-cellular structures that grow from cell to cell, hijacking the cellular machinery of their temporal host cells. Infection thread growth is a process also derived from the eukaryotic cell cycle (van Brussel et al., 1992; Yang et al., 1994; Timmers et al., 1999). Before penetration by an infection thread, the targeted vacuolated cortical root cell undergoes cytological reorganization, forming a so-called pre-infection thread or PIT (Figure 1C; van Brussel et al., 1992; Timmers et al., 1999). The PIT is structurally related to the phragmosome, a cytoplasmic bridge through the vacuole formed by cells in preparation for mitosis. An analysis of the expression of cell cycle marker genes revealed that PIT-forming cells indeed enter into the cell cycle and duplicate their genomes but are arrested in the G2 phase, before the initiation of mitosis (Yang et al., 1994). More recently, it was shown that root hairs also prepare for infection thread formation by activating cell cycle genes (Breakspear et al., 2014). PITs are reminiscent of the prepenetration apparatus in root cells before infection with the hyphae of arbuscular mycorrhizal fungi (Genre et al., 2008).

Infection threads grow toward young, differentiating symbiotic cells where they release rhizobia through an endocytotic process (Figure 1D). Endocytosis, a well-known process in plants (Paez Valencia et al., 2016), is essential for controlling the turnover of plasma membrane proteins such as receptors and transporters.

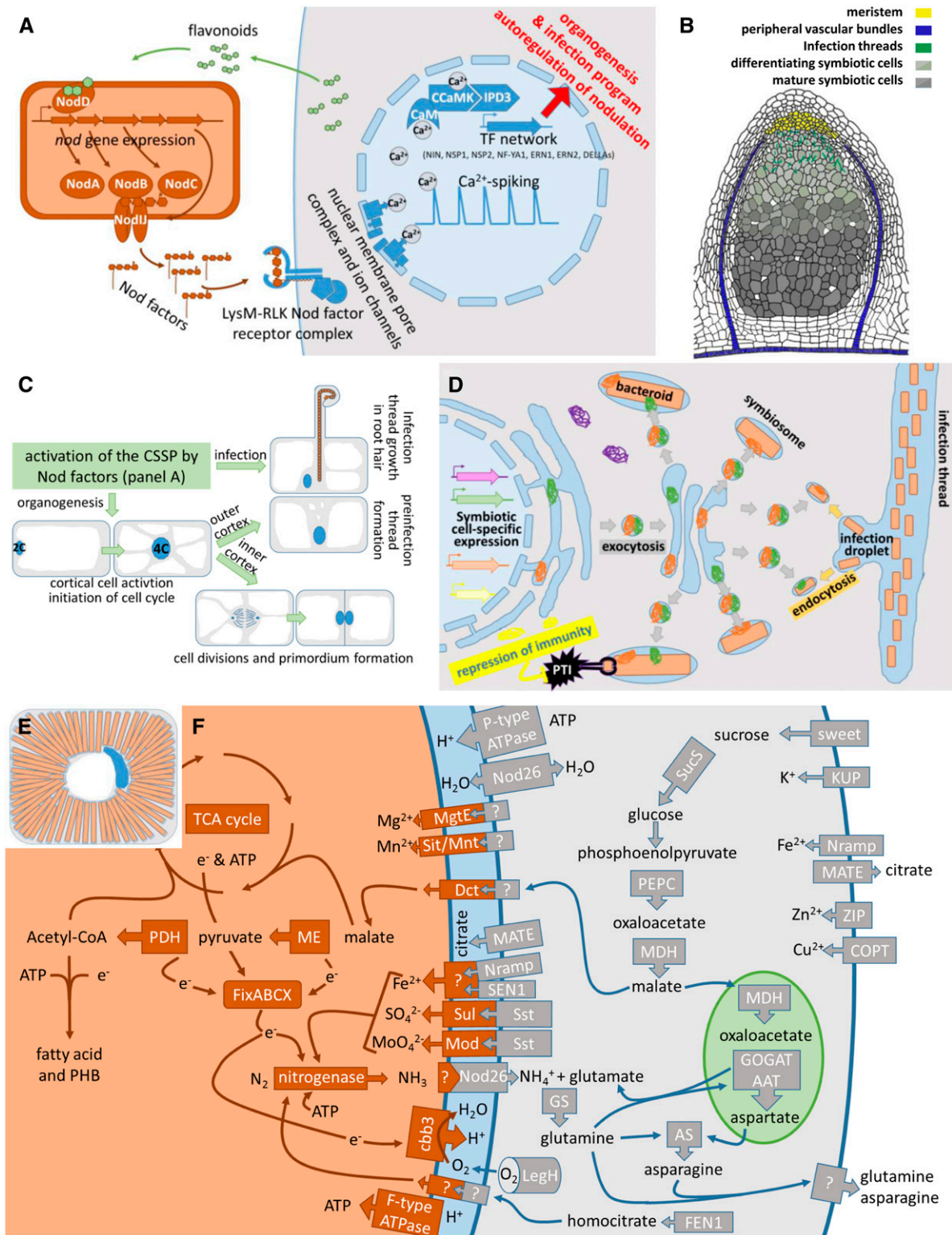


Figure 1. Unique Features of Nodules.

(A) Nod factor activation of the CSSP. Brown, signaling in the rhizobium symbiont resulting in Nod factor production. Gray, CSSP in the host cells activates nodule organogenesis, infection, and autoregulation of nodulation.

However, the uptake of whole bacteria by plant cells as during rhizobium (or *Frankia*) infection of nodule cells is unknown, except for the case of the nitrogen-fixing symbiosis of the plant genus *Gunnera* with cyanobacteria. Nevertheless, the infection of nodule cells shares some similarities with the infection of plant cells by fungal pathogens and mycorrhizal symbionts (Harrison and Ivanov, 2017; Parniske, 2018). Endocytotic uptake does not release rhizobia freely in the cytosol of nodule cells, instead releasing them inside organelle-like vesicles known as symbiosomes, which have a plasmalemma-like membrane (Figure 1D). Within symbiosomes, the rhizobia grow and differentiate into nitrogen-fixing structures called bacteroids. The latter process is enabled by the exocytosis pathway, which traffics specific proteins to the symbiosomes (Figure 1D). The exocytosis pathway is strongly conserved in eukaryotes, but some components are nodule specific and are encoded by duplicated genes or by nodule-specific splice variants (Stonoha-Arther and Wang, 2018).

Among the unique features of nodules, the nitrogen-fixation process by symbiotic nodule cells is certainly the most prominent. This process requires the maintenance of thousands of bacterial endosymbiont-containing symbiosomes within individual plant cells (Figure 1E). During this process, symbiosomes acquire a specific vesicular identity and avoid a vacuolar identity, which would lead to the elimination of the symbiont (Limpens et al., 2009; Sinharoy et al., 2013). In addition, symbiotic cells contain specific machinery that suppresses innate immunity responses against their many endosymbionts (Figure 1D; Gourion et al., 2015; Wang et al., 2016; Domonkos et al., 2017; Yu et al., 2018). In certain legumes, such as *Medicago* spp in the inverted repeat-lacking clade (IRLC) and *Aeschynomene* spp in the Dalbergoid branch, symbiosome formation also involves the production of symbiosis-specific peptides called NCRs, which target the bacteroids and control their extreme, terminal differentiation (Mergaert, 2018). Finally, the functioning of symbiotic cells requires bacterial metabolism to be integrated with the metabolism of the host cells (Figure 1F). Bacteroids generate ammonia in large excess to their own needs and release this ammonia to the plant. Major features of the metabolism of symbiotic cells include (1) the supply of carbon sources derived from photosynthesis in the shoots to the energy-

devouring bacteroids, (2) the maintenance of a low oxygen concentration to avoid the irreversible oxidation and inactivation of the nitrogenase enzyme, and (3) the assimilation of ammonium provided by nitrogen-fixing bacteroids. This specific metabolism requires the activation of a suite of enzymes and transporters and the massive production of leghemoglobin to bind oxygen (Figures 1D and 1F; Udvardi and Poole 2013; Poole et al., 2018; Roy et al., 2019).

NODULE TRANSCRIPTOME DATABASES

Transcriptional reprogramming associated with nodule formation has been studied in many legume species, including *M. truncatula* (see below), *Medicago sativa* (O'Rourke et al., 2015), pea (*Pisum sativum*; Alves-Carvalho et al., 2015), chickpea (*Cicer arietinum*; Kant et al., 2016; Kudapa et al., 2018), *Trifolium pratense* (Trujillo et al., 2019), *L. japonicus* (Colebatch et al., 2004; Høgslund et al., 2009; Takanashi et al., 2012; Verdier et al., 2013; Kelly et al., 2018), *G. max* (Brechenmacher et al., 2008; Severin et al., 2010), common bean (*Phaseolus vulgaris*; O'Rourke et al., 2014), *Sesbania rostrata* (Capoen et al., 2007), *Lupinus* spp (Keller et al., 2018), *Aeschynomene evenia* (Gully et al., 2018), and peanut (*Arachis hypogaea*), *Arachis duranensis*, *Arachis ipaensis* (Clevenger et al., 2016; Peng et al., 2017; Karmakar et al., 2019), and in the nonlegume *Parasponia andersonii* (van Velzen et al., 2018), as well as in the actinorhizal plant species *Casuarina glauca* (Hocher et al., 2011), *Alnus glutinosa* (Hocher et al., 2011), *Datisca glomerata* (Demina et al., 2013; Battenberg et al., 2018; Salgado et al., 2018), and *Ceanothus thyrsiflorus* (Battenberg et al., 2018; Salgado et al., 2018).

Transcriptomic approaches to study nodule formation have been extensively applied in *M. truncatula*, in large part due to the development of whole genome Affymetrix *Medicago* GeneChips. These GeneChips have been used in 274 different experimental conditions, which can be found in the *Medicago truncatula* Gene Expression Atlas (MtGEA) repository (<https://mtgea.noble.org/v3/index.php>; Benedito et al., 2008). MtGEA compiles a broad range of developmental and environmental conditions, including different plant organs and tissue types, biotic-, abiotic stress

Figure 1. (continued).

(B) Organization of nodule tissues. Tissues are color-coded. The dividing meristem cells (yellow) give rise at the edge to several vascular bundles (blue) concentrically organized around the central nodule tissue. The meristem cells (2C to 4C DNA content) also give rise to the central nodule tissue, which is mainly composed of infected symbiotic cells formed from postmeristematic cells by successive rounds of endoreduplication (shades of gray from light to dark indicate cells with increasing ploidy from 8C to 64C). Infection threads (green) deliver rhizobia to early differentiating symbiotic cells.

(C) Infection threads, filled with rhizobium bacteria, are initiated in a root hair and have polar growth (red). The underlying cortical cells are activated and initiate a cell cycle. Outer cortical cells, adjacent to an infection thread-carrying cell, do not complete the cell cycle but form a cytoplasmic bridge (related to a phragmosome) that will be penetrated by the growing infection thread, which is referred to as a preinfection thread. Inner cortical cells complete the cell cycle and the newly generated cells constitute the nodule primordium. Drawings are based on van Brussel et al., 1992.

(D) An infection thread reaching a target cell in a nodule primordium or growing nodule releases rhizobia by endocytosis. The exocytosis pathway of the host cell targets proteins and peptides to bacteria located in vesicles called symbiosomes, resulting in the differentiation of these bacteria into nitrogen-fixing bacteroids. PTI is actively repressed in symbiotic cells. Specific expression in symbiotic cells is required for the production of symbiosome-targeted secretory proteins (green and orange), for immune suppression functions (yellow), and for metabolism [purple, see **(F)**].

(E) A fully mature symbiotic cell is entirely filled with bacteroid-containing symbiosomes and has a central vacuole and a polyploid nucleus (blue).

(F) Integration of the bacteroid metabolism (brown) with symbiotic cell metabolism (gray) relies on complementarity of the metabolic functions in both symbiotic partners and a complex exchange of small molecules mediated by a multitude of transporters. Some metabolic steps require a third partner, the plastid (green). The drawing is based on Temple et al., 1998, Udvardi and Poole, 2013, and Poole et al., 2018.

conditions, and hormone treatments. It is therefore currently the richest *M. truncatula* resource for the analysis of expression patterns and specificity of genes of interest and the discovery of genes expressed under certain conditions or in organs of interest, even if a limitation is the absence in the database of a subset of the currently annotated *M. truncatula* genes, in particular genes encoding peptides and noncoding RNAs. Ordering the MtGEA data set with a clustering algorithm results in the identification of tissue-specific gene sets (Figure 2A). The heat map discloses large clusters of genes that are expressed in different plant organs, leaves, stems, flowers, seeds, hypocotyls, roots, mycorrhiza, and

nodules. This analysis also revealed that nodules and mycorrhiza are primarily typified by the expression of different sets of genes (Figure 2B), despite the shared CSSP module of regulators that controls both nodule and mycorrhiza formation and a number of common target genes that are activated immediately downstream of the CSSP during both types of interactions (Hogekamp et al., 2011). For research on rhizobium-legume symbiosis, the MtGEA data set can be complemented by expression data derived from laser-capture microdissection (LCM)-transcriptome analysis available in the SYMBIMICS database (<https://iant.toulouse.inra.fr/symbimics/>; Roux et al., 2014). Together, the MtGEA and

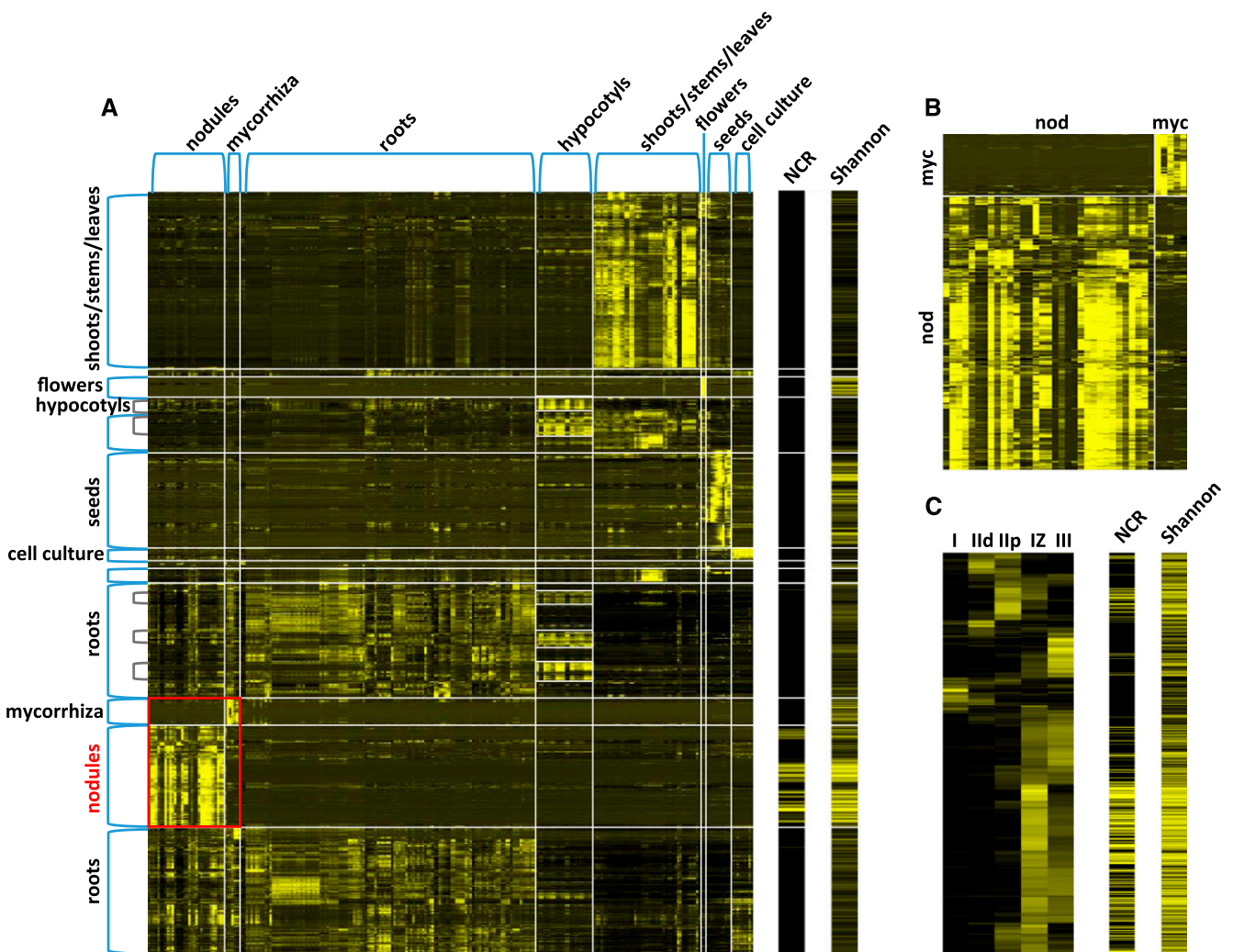


Figure 2. Gene Expression in *M. truncatula*.

(A) Hierarchical clustering of the MtGEA data set (Benedito et al., 2008). Genes are shown in rows and experiments in columns. On the top, the experiment types are indicated. To the left, gene clusters are labeled based on organ-specific expression. In the expression heatmap, yellow indicates high expression and black indicates low expression. In the “NCR” column, the NCR genes are indicated by a yellow line. In the “Shannon” columns, yellow indicates high specificity (i.e. low Shannon entropy) and black indicates low specificity, with more uniform expression (i.e. high Shannon entropy).

(B) An enlargement of the red squared area of the heatmap in panel A showing nodule-specific (nod) and mycorrhizal-specific (myc) genes (lines) and experiments (columns).

(C) The SYMBIMICS data set (Roux et al., 2014) for nodule-specific genes from MtGEA [gene cluster labeled in red in (A)]. The columns correspond to the nodule tissues: I, meristem; IId, distal infection zone or nodule zone II; IIp, proximal infection zone; IZ, interzone between nodule zones II and III; III, fixation zone or nodule zone III. Color-codes as in (A). The data sets used to generate the heatmaps are available upon request.

SYMBIMICS databases are priceless resources for studying *M. truncatula* biology, particularly nodulation.

Data sets similar to MtGEA are available for two alfalfa subspecies, *M. sativa* ssp. *sativa* (B47) and *M. sativa* ssp. *falcata* (F56; Alfalfa Gene Index and Expression Atlas Database or AGED; <http://plantgrn.noble.org/AGED/>; O'Rourke et al., 2015), *L. japonicus* (LjGEA; <https://ljgea.noble.org/v2/>; <https://lotus.au.dk/expat/>; Verdier et al., 2013), *P. vulgaris* (Common Bean Gene Expression Atlas or PvGEA; <http://plantgrn.noble.org/PvGEA/>; O'Rourke et al., 2014), soybean (SoyBase; <https://soybase.org/soyseq/>; Severin et al., 2010), and *Arachis* spp (PeanutBase; <https://www.peanutbase.org/>; Clevenger et al., 2016). Except for LjGEA, which reports ~80 experimental conditions, including organs and physiological conditions, the other databases are less well furnished and focus on the principal plant organs. LjGEA was generated with Affymetrix gene chips while the AGED, PvGEA, SoyBase, and PeanutBase data sets were generated by RNA-sequencing (RNA-seq).

NOVEL INSIGHTS FROM NODULE TRANSCRIPTOME ANALYSES

Because nodules form de novo on roots starting from fully differentiated cortical, endodermal, pericycle, and epidermal cells, the nodule transcriptome is usually defined by the genes showing differential expression in nodules versus roots. A recurring theme in each transcriptome study on legumes, *Parasponia*, and actinorhizal plants is that nodule formation is accompanied by the differential expression of hundreds to thousands of genes. These sets of nodule-enhanced genes are composed in part of genes that are more strongly expressed in nodules than in other organs, but importantly, these gene sets also contain large numbers of absolutely nodule-specific genes. Conversely, a transcriptome analysis in *M. truncatula* revealed that the larger part of the gene expression changes during lateral root initiation also occur during nodule organogenesis (Schiessl et al., 2019). These common differentially expressed genes are related to auxin biosynthesis and signaling and transcriptional regulation. This observation suggests that lateral root organogenesis has been recruited into the symbiotic program to which a large panel of nodule-specific features was added. The nodule-specific gene sets are composed of genes with homologs in other plants or paralogs expressed in other plant tissues as well as orphan genes. These findings further underscore the unique nature of nodules and the very complex genetic program governing the formation of this symbiotic organ. Thus, from this perspective, the glass might be half-empty.

The Specificity of the Nodule Transcriptome

The analysis presented in Figure 2 suggests that some organs, particularly nodules, are associated with the expression of genes that are extraordinarily specifically expressed. This is obvious when zooming in on the expression of NCR genes (Figure 3A; Guefrachi et al., 2014). This extremely large gene family encodes nodule-specific cysteine-rich peptides, which target endosymbionts and induce them into a terminally differentiated state, are required for maintaining bacteroids, or are involved in the selection of well-adapted rhizobia in nodules (Mergaert, 2018). All

but five of the over 300 analyzed NCR genes are exclusively expressed in nodules and in no other plant organ or in response to any other biotic interaction or abiotic stress. To express the apparent tissue specificity of genes expressed in nodules and of the NCRs in a quantitative manner, a metric called Shannon entropy was used (Guefrachi et al., 2014). Shannon entropy is a parameter derived from information theory that is applied to gene expression analysis to characterize the uniformity (high Shannon entropy) or specificity (low Shannon entropy) of the expression pattern of a gene of interest over a number of tested conditions. Taking into account ten different tissues or organs of *M. truncatula*, the Shannon entropy was calculated for all genes in MtGEA. Ordering the MtGEA genes according to expression profile (Figure 3B) or ranking them according to their Shannon entropy value (Figure 3C) makes it clear that flowers, seeds, pods, and mycorrhiza and most particularly nodules are characterized by large gene sets with very low Shannon entropy that are thus (almost) uniquely expressed in these organs. Thus, nodulation in *M. truncatula* strongly depends on dedicated gene sets that do not function elsewhere in the plant (Guefrachi et al., 2014). As discussed below, the large majority of these genes are expressed in differentiating or mature symbiotic cells (Figure 3C).

Shannon entropy analysis of the AGED, PvGEA, LjGEA, and SoyBase data sets (Figure 3D) confirmed the notion that nodule formation in *M. sativa*, *P. vulgaris*, *L. japonicus*, and soybean is also associated with the expression of large numbers of low Shannon entropy genes that are (nearly) uniquely expressed in that organ. Soybean/*P. vulgaris*, *L. japonicus*, and *M. truncatula*/*M. sativa* belong to different clades of legumes: the Milletioids, Robinioids, and IRLC, respectively. Thus, the reliance on large gene sets that are uniquely expressed in the nodule is not restricted to *Medicago* spp and is not a specific feature of the particular nodule types formed by these species, which are of the indeterminate type and induce terminal bacteroid differentiation. Both of these features are not present in bean, soybean, or *L. japonicus* nodules. However, the nodule-specific gene sets in *Medicago* spp appear to be significantly larger than those in the other legumes (Figure 3). This difference is largely due to the extremely high number of *Medicago* genes encoding NCR peptides (Mergaert et al., 2003). In *M. truncatula*, the NCR gene family is composed of ~600 genes and in tetraploid *M. sativa*, the number of NCR genes increases to 1300 (O'Rourke et al., 2015), whereas these genes are absent in the three other legumes.

Besides the NCR gene family, there are many other nodule-specific, low entropy genes. Among these are well-studied genes, some with essential functions in symbiosis (e.g. genes encoding leghemoglobin, SNARPs [small nodulin acidic RNA binding protein, also known as LEED..PEEDs], the phospholipase C-like DNF2 (defective in nitrogen fixation), the cysteine-rich receptor-like kinase SymCRK, metabolite transporters, and so on) and many genes with only predicted or uncharacterized functions (e.g. putative transporter encoding genes, glycine-rich proteins [GRPs], early nodulins ENOD8 and ENOD16, calmodulin-like proteins, and so on), or genes of unknown function. Importantly, the nodule-specific transcriptome contains several families of orphan genes or taxonomically restricted genes (genes without homologs in other lineages), including the NCRs, SNARPs, and GRPs, which are restricted to the IRLC legumes (Kevei et al., 2002;

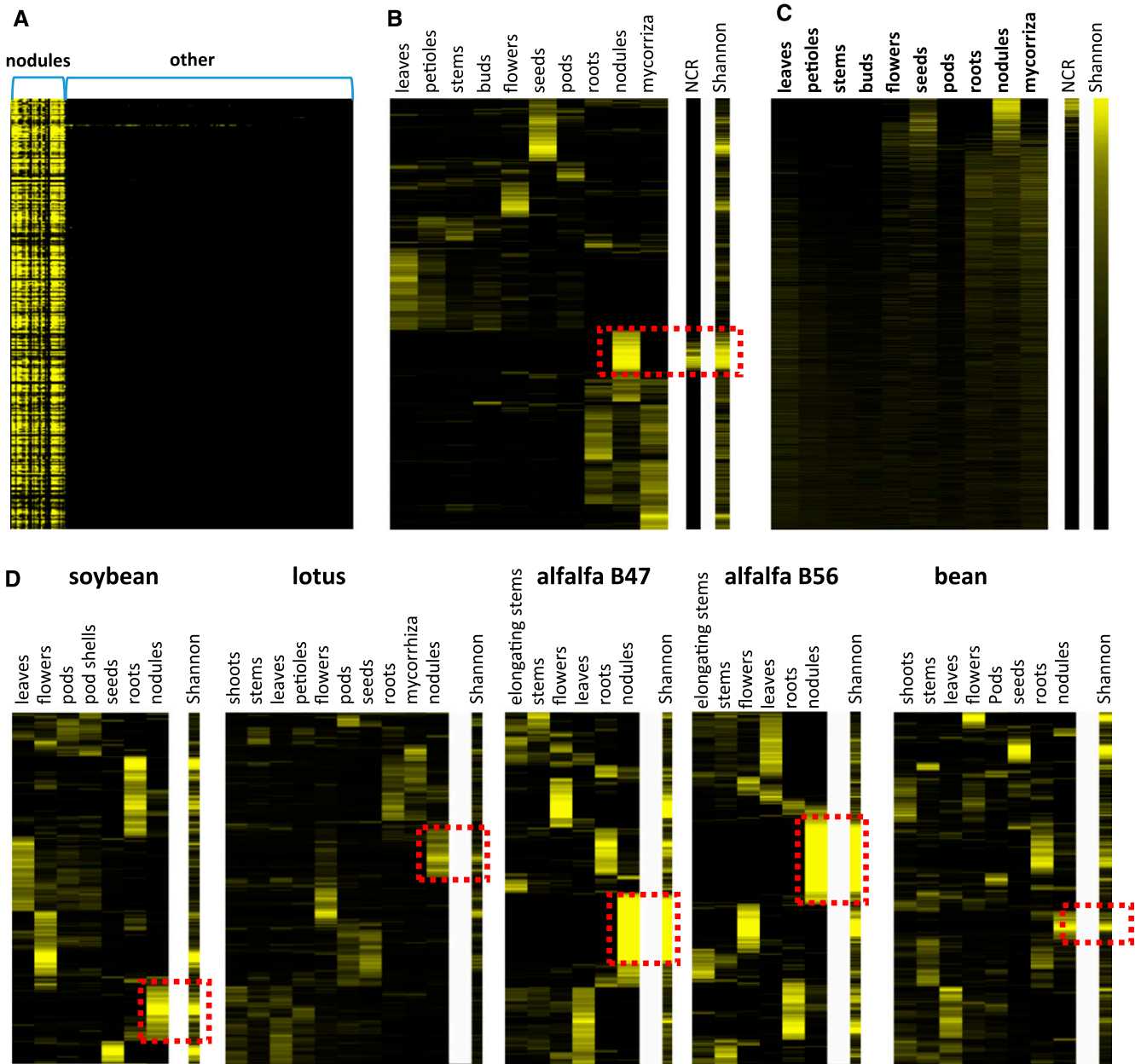


Figure 3. Nodule-Specific Gene Expression.

(A) Expression of NCR genes in *M. truncatula*.

(B) Extreme specificity of nodule-specific genes in *M. truncatula*. Clusters of tissue-specific genes. Expression patterns of genes ordered according to specificity (Shannon entropy).

(C) Expression patterns of *M. truncatula*, genes ordered according to specificity (Shannon entropy).

(A) to **(C)** are adapted from Guefrachi et al. (2014), licensed under the Creative Commons Attribution 4.0 International License.

(D) Nodule-specific gene expression in different legumes.

The heatmaps show hierarchical cluster analysis of gene expression in the indicated legumes using data from the SoyBase (Severin et al., 2010), LjGEA (Verdier et al., 2013), AGED (O'Rourke et al., 2015), and PvGEA (O'Rourke et al., 2014) databases. Genes are shown in rows and tissue types in columns. The Shannon entropy of the genes (calculated based on the expression profile) is displayed to the right of each gene expression heatmap. The nodule-specific gene clusters mostly comprising low Shannon entropy genes are highlighted by a red rectangle. Color codes: in the expression heatmaps, yellow indicates high expression and black indicates low expression; in the "NCR" columns, the NCR genes are indicated by a yellow line; in the "Shannon" columns, yellow indicates high specific expression (i.e. low Shannon entropy) and black indicates low specificity, more uniform expression (i.e. high Shannon entropy). The data sets used to generate the heatmaps are available upon request.

(C) Expression patterns of *M. truncatula* genes ordered according to specificity (Shannon entropy).

Mergaert et al., 2003; Laporte et al., 2010; Trujillo et al., 2014), the NCR genes of *Aeschynomene* spp (Czernic et al., 2015), the NAD1 (nodules with activated defense 1) genes conserved only in plants from the NFC (Wang et al., 2016), and the nodulin 20 family, first identified in soybean (Richter et al., 1991), which is specific for the Phaseoleae (Trujillo et al., 2019). Such genes could have arisen from the duplication and rearrangement of ancestral genes followed by rapid diversification. Alternatively, orphan genes could have evolved de novo from noncoding genomic regions believed to be used as raw materials for the evolution of new genes (Tautz and Domazet-Lošo 2011).

A number of low-entropy nodule-specific genes encode proteins with unique combinations of existing domains, likely resulting from gene recombination. For example, the symbiosis-specific thioredoxin Trx s, Calmodulin-like protein, and ENOD8 families evolved from fusions of known protein domains with an N-terminal signal peptide, resulting in unique secretory proteins targeted to symbiosomes and bacteroids. There are a remarkably high number of low-entropy genes with signal peptides, suggesting that symbiotic cells target many proteins and peptides to the symbiosomes (Figure 1D; e.g. NCRs, GRPs, SNARPs, calmodulin-like proteins, ENOD8, ENOD16, ENOD20, DNF2, SymCRK, Trx s, nodulin 25, MtN22, MtN11, lipase/lipoxygenases [also known as nodule-specific PLAT domain proteins or NPDs], basic blue-like protein, NCR-like peptides, MBP-1, and CAP antimicrobial peptides in the Dalbergioid lineage, nodulin 20 family proteins of the Phaseoleae, APN1 of *L. japonicus*, and so on; Dickstein et al., 2002; Liu et al., 2006; Alunni et al., 2007; Laporte et al., 2010; Maunoury, 2010; Trujillo et al., 2014; Yamaya-Ito et al., 2018; Pecrix et al., 2018; Pislariu et al., 2019; Trujillo et al., 2019). The high level of protein secretion to symbiosomes and bacteroids in symbiotic cells of *M. truncatula*, particularly in developing and young mature cells, is further supported by the strongly developed endoplasmic reticulum in these cells (Maunoury et al., 2010) and by the requirement of symbiotic cell-specific regulators of the secretory pathway for normal bacteroid differentiation and nodule development (reviewed in Stonoha-Arther and Wang, 2018).

Other genes resulted from duplications and modifications with neofunctionalization, such as leghemoglobin genes. Oxygen binding leghemoglobins are essential for protecting the oxygen-sensitive enzyme nitrogenase. Their recruitment for symbiosis required mutations (replacement of a few amino acids) of nonoxygen-transporting hemoglobins present in all plants to create an oxygen-transporting hemoglobin, as well as nodule-specific expression of the corresponding genes. Other genes also resulted from duplication, whereby one paralog obtained a new, nodule-specific expression profile but without a new function. Genes encoding subunits of a nodule-specific signal peptidase complex, including *Dnf1*, fall in this category (Wang et al., 2010). This endoplasmic reticulum-localized enzyme complex of the secretory pathway is required for removing the signal peptide during trafficking of secretory proteins. The NPDs in the IRLCs and the MBP-1 and CAP peptides in the Dalbergioids are encoded by multigene families including nodule-specific subsets as well as subsets with constitutive expression (Trujillo et al., 2019). Other examples are the many small metabolite transporter genes that are essential for the formation of functional nodules (Figure 1F; Sugiyama et al., 2017; Frare et al., 2018; Senovilla et al., 2018;

Chen et al., 2019). These genes belong to small multigene families, with one copy having nodule-specific or nodule-enhanced expression. The encoded transporters assure the transport of metabolites to and from the symbiotic cells and bacteroids and are essential for the metabolic integration of bacteroids with symbiotic nodule cells. A similar situation holds true for the many genes encoding the enzymes involved in the specific carbon and nitrogen metabolism of symbiotic cells (Figure 1F). Carbon metabolism provides the carbon source to bacteroids in the form of dicarboxylic acids and synthesizes the carbon backbone for assimilation of the ammonium derived from nitrogen fixation in bacteroids. Nitrogen metabolism assures that this nitrogen is assimilated into amino acids or in some legumes into ureides. A large body of evidence indicates that the key enzymes of carbon and nitrogen metabolism in symbiotic cells are encoded by nodule-specific genes. Like small-metabolite transporters, metabolic genes are also members of small gene families, with one copy having a nodule-specific or -enhanced expression (e.g. Miller et al., 1998; Hakoyama et al., 2009; Nouwen et al., 2017).

The large number of rhizobia in symbiotic cells implies the massive presence of PAMPs (pathogen-associated molecular patterns) and the potential to induce strong PAMP triggered immunity (PTI). Several symbiotic cell-specific proteins are required to suppress the PTI response and to enable the long-term maintenance of rhizobia (Figure 1D). In the absence of the receptor-like kinase SymCRK, the phospholipase C-like DNF2, the PLAT Domain protein NPD1, the transcription factor RSD, or the endoplasmic reticulum located small protein NAD1, a strong immune response is induced, as evidenced by the induction of immunity marker genes, nodule necrosis, and a complete halting of symbiosis (Bourcy et al., 2013; Sinharoy et al., 2013; Berrabah et al., 2014; Wang et al., 2016; Domonkos et al., 2017; Pislariu et al., 2019). The genes encoding these immune repressors are low entropy genes that are only expressed in symbiotic nodule cells. *SymCRK*, *DNF2*, *NPD1*, and *RSD* are nodule-specific paralogs, whereas *NAD1* is unique and appears to be specific to the NFC (Wang et al., 2016), suggesting that it was present in the ancestor of the NFC.

Examples of Functional Insights from Nodule Transcriptomics

As mentioned above, nodule formation has been studied by transcriptomics in many legumes and nonlegume species, and several studies have focused on specific aspects of the symbiotic process in these NFC plants (e.g. Høglund et al., 2009; Libault et al., 2009a; Libault et al., 2010; Yuan et al., 2017; van Velzen et al., 2018). Undoubtedly, transcriptomics have been used most extensively in *M. truncatula* to study nodule development, from very early responses induced by Nod factors and gene activation in root hairs (Comber et al., 2008a; Czaja et al., 2012; Breakspear et al., 2014; Larrainzar et al., 2015; van Zeijl et al., 2015; Jardinaud et al., 2016) to the development of mature nodules (El Yahyaoui et al., 2004; Maunoury et al., 2010; Moreau et al., 2011). Analyses focusing on the early responses to Nod factors and rhizobia have used clever tricks to overcome the limited sensitivity in detecting transcriptional responses in the small number of Nod factor- or rhizobium-responsive cells in roots. These solutions included

preparing RNA from the roots of supernodulating plant mutants, which have a strongly enhanced nodulation response (Combiér et al., 2008a; Larrainzar et al., 2015). Alternatively, analysis has been performed using as a starting material purified root hairs, the first target cells of Nod factors and rhizobia, obtained by breaking them off from roots frozen in liquid nitrogen (Breakspear et al., 2014; Damiani et al., 2016; Liu et al., 2019a) or root epidermal cells obtained by LCM (Jardinaud et al., 2016). Analysis of the early symbiotic signaling responses highlighted the importance of cytokinin, auxin, ethylene, strigolactone, abscisic acid, gibberellic acid, and brassinosteroid phytohormone signaling pathways. This observation suggests that an important rewiring of the interconnected phytohormone pathways is a key step in the initiation of nodule organogenesis (Czaja et al., 2012; Breakspear et al., 2014; Larrainzar et al., 2015; van Zeijl et al., 2015; Jardinaud et al., 2016; reviewed in Gamas et al., 2017; Buhian and Bensmihen, 2018; Liu et al., 2018; Liu et al., 2019a). These studies confirmed the activation of cell cycle-related processes early in the nodulation signaling pathway (Breakspear et al., 2014; Larrainzar et al., 2015; Jardinaud et al., 2016), which is in agreement with the known role of the cell cycle in the establishment of a nodule primordium (Foucher and Kondorosi, 2000), cell differentiation (Vinardell et al., 2003; Suzuki et al., 2014), and infection (van Brussel et al., 1992; Yang et al., 1994; Suzuki et al., 2014). Another striking finding is the identification of the gene encoding the rapid alkanization factor MtRALF1 peptide as a Nod factor-induced gene controlling infection thread development (Combiér et al., 2008a). Peptides of this family control cell expansion in plants, notably during root hair and pollen tube growth (Haruta et al., 2014; Murphy and De Smet, 2014; Ge et al., 2017; Mecchia et al., 2017), the two plant cell types that show the highest similarity to the infection thread structure. Notably, these RALF peptides interact with receptor-like kinases containing a so-called malectin-like domain within their ectodomain (Franck et al., 2018). Such a domain is also present and plays a key role in the ectodomain of the receptor-like kinase SYMRK/DMI2/NORK (Antolín-Llovera et al., 2014; Pan et al., 2018), a component of the CSSP, suggesting that MtRALF1 targets the Nod factor receptor complex, a hypothesis that might be worth pursuing.

Using time series of wild-type nodule growth combined with collections of both rhizobial and plant mutants arrested at different stages of nodule development, transcriptome dynamics during nodule organogenesis was dissected into distinct transcriptional waves composed of the induction and repression of genes that are consecutively initiated during organogenesis (Starker et al., 2006; Maunoury et al., 2010; Moreau et al., 2011; Lang and Long, 2015). Four major waves of gene activation were uncovered using these approaches. The first wave is associated with early signaling and bacterial infections, the second with symbiotic cell differentiation, the third with the differentiation of bacteroids, and the fourth with the activation of nitrogen fixation.

In addition to studies that focused on the overall transcriptome dynamics associated with nodule development, other studies have used transcriptomics to address more specific questions. The transcriptomes in nonfunctional nodules formed by *M. truncatula* mutants in transcriptional regulators were used to characterize the stages of nodule development affected by the mutations at the molecular level and to delineate the sets of

potential target genes of these transcriptional regulators (see below; Vernié et al., 2008; Satgé et al., 2016; Sinharoy et al., 2013; Deng et al., 2019). Another type of study used transcriptomics to compare the nutritional status of whole plants in response to symbiotic nitrogen fixation or mineral nitrogen as nitrogen source (Ruffel et al., 2008) or to investigate specific physiological conditions in nodules. During nodule senescence, nodule cells break down and their macromolecular components are degraded and remobilized to other plant organs (Van de Velde et al., 2006). The process can be developmentally induced by the aging of nodules or when the host plants need to shift their nutrient resources for seed production. Nodule senescence can also be induced by particular physiological conditions, including reduced photosynthesis, nutrient starvation, or a shift in the nitrogen status of the plant. Nodule senescence under these different conditions has been studied intensively by transcriptomics (Van de Velde et al., 2006; Seabra et al., 2012; Cabeza et al., 2014a, 2014b; Liese et al., 2017; Deng et al., 2019), which has led to the identification of marker genes and actors in the process. Cys proteinases appear to be key players in nodule senescence: manipulating their expression causes delayed or premature senescence, thus affecting nitrogen fixation (Pierre et al., 2014; Deng et al., 2019).

Other studies have used transcriptomics and dedicated Affymetrix gene chips to focus on the expression of NCR and other defensin-like peptide genes in *M. truncatula* (Nallu et al., 2013; Tesfaye et al., 2013). The results confirmed the notion that NCRs are nodule-specific peptides and revealed a distinct subset of defensin-like genes unrelated to the NCRs that are specifically expressed in seeds. Like *M. truncatula*, many defensin-like peptides are also expressed in both the inflorescences and seeds of *Arabidopsis* (*Arabidopsis thaliana*). The peptides in these organs may play roles in defense. Indeed, the abundant expression of defensin-like peptides may reflect a specific, strongly protected immune status of reproductive organs. In an RNA-seq study focusing on small, secreted peptides that act via receptor-mediated signaling, 365 peptides belonging to 29 different putative functional peptide families were identified whose expression changed during nodule formation, suggesting that peptide signaling plays a prominent role in the coordination of nodulation (de Bang et al., 2017). This data set produced in this study represents a rich resource for future work aiming to decipher the functions of these peptides.

The genotypes of both the *Rhizobium* and legume influence the outcome of the interaction and the efficiency of the nitrogen-fixation process (Sugawara et al., 2013; Kazmierczak et al., 2017). Transcriptomics were used to gain insights in the molecular basis of these genotype-by-genotype variations in *M. truncatula*-*Sinorhizobium* symbiosis (Heath et al., 2012; Nallu et al., 2014; Burghardt et al., 2017). These studies revealed a high degree of genotype-by-genotype variability in the transcriptomes that was more dependent on the host genotype than the bacterial symbiont genotype. In particular, NCR expression is highly variable, suggesting that the NCRs and the differentiation of the bacteroids induced by NCRs are important determinants of the efficiency of symbiosis (Kazmierczak et al., 2017). In agreement with this finding, the orthologous NCR genes within a collection of *M. truncatula* accessions showed a high degree of nucleotide variation with signatures of diversifying selection pressure, suggesting

that NCR genes are rapidly and recently evolving, possibly to help plants adapt to diversifying rhizobium symbionts (Branca et al., 2011; Nallu et al., 2014; Zhou et al., 2017).

Many transcriptome studies have been performed using whole nodules and did not provide information on the cell type-specific expression of nodule-enhanced or nodule-specific genes. Indeterminate *M. truncatula* nodules are complex organs composed of many cell types. Differentially expressed genes could be specifically expressed in one or several of these cell types. However, some studies have provided spatial resolution at the level of cell type or nodule tissue (Maunoury et al., 2010; Limpens et al., 2013; Roux et al., 2014). These studies demonstrated that the large majority of nodule-enhanced or nodule-specific genes are expressed in differentiating or mature symbiotic cells. By analyzing different stages of nodule development in conjunction with a large panel of symbiotic mutants arrested in nodule development, two major transcriptome switches (regulating a large proportion of nodule-enhanced or -specific genes) were revealed, one that correlated with symbiotic cell maturation and another one with bacteroid differentiation (Maunoury et al., 2010). An impressive improvement in transcriptome analysis was achieved by combining LCM of nodule tissues with Affymetrix gene chips (Limpens et al., 2013) or RNA-seq analysis (Roux et al., 2014; Jardinaud et al., 2016), providing spatial resolution to global gene expression analysis in nodules (Figure 4). The dissection of nodules by LCM-RNA-seq confirmed the conclusion that a large majority (over 90%) of the 1200 nodule-enhanced or -specific genes are expressed in differentiating or mature symbiotic cells (Figure 3C; Roux et al., 2014). This observation reinforces the notion that symbiotic nodule cells are unique cell types that have complex needs for their formation and maintenance and for the metabolic integration of endosymbionts.

New technologies could in the (near) future further improve the spatial resolution to the level of nodule tissue subdomains or even single cells, which would resolve nonuniform responses that are averaged over the population of cells in multi-cell samples or unmask unique biological properties of single cells that are too diluted in populations to be discovered in multi-cell samples. These new methods include single cell-type or single cell transcriptome analysis. Single cell types or single cells or the corresponding nuclei could be obtained through LCM, through isolating nuclei tagged in specific cell types (INTACT), or through flow cytometry and sorting of nuclei based on genome size (Figure 4; Libault, 2018). The latter technique could also be combined with fluorescent labeling of nuclei in specific cells in transgenic plant lines with cell type-specific promoters driving the expression of a fluorescent marker. Other new technologies, including spatial transcriptomics and Slide-seq, that could be applied to nodule gene expression analysis, are based on the transfer of RNA from tissue sections onto a surface coated with spatially structured molecular barcodes that are sequenced together with the transferred RNA, generating spatially resolved RNA-seq transcriptome data (Figure 4; Ståhl et al., 2016; Rodrigues et al., 2019).

Genome Organization of Nodule-Specific Genes

Nodule-specific genes often form small, intermediate, or even very large gene families, as observed for the NCR genes in many IRLC

legumes (Montiel et al., 2017). After their birth, the NCR genes in these plants must have spread massively through the genome by repetitive gene duplications. NCR genes appear in small clusters, which are spread across the genome in *M. truncatula* (Graham et al., 2004; Alunni et al., 2007; Zhou et al., 2017), indicating that long-distance duplications followed by local duplications must have played a role in the spreading of NCR genes. A well-documented mechanism for generating gene duplications is based on transposon-mediated nonhomologous recombination and retrotranscription (Galindo-González et al., 2017). Many NCR genes in *M. truncatula* are in the vicinity of transposable elements, indicating transposons were (and possibly still are) involved in the explosion of the NCR gene family in the *M. truncatula* genome (Graham et al., 2004; Satgé et al., 2016).

Besides NCR genes, other nodule-specific genes in *M. truncatula*, particularly those expressed in differentiating and mature symbiotic cells, also display local duplications, including genes encoding GRPs, SNARPs, NPDs, calmodulin-like proteins, ENOD8, leghemoglobins, the ammonium channel NOD26, and others (Dickstein et al., 2002; Liu et al., 2006; Alunni et al., 2007; Trujillo et al., 2014; Frare et al., 2018; Pecrix et al., 2018; Trujillo et al., 2019). These clusters are distributed all over the *M. truncatula* genome and strikingly, in addition to locally duplicated genes, they often contain different types of nodule-specific genes (Liu et al., 2006; Pecrix et al., 2018). These genome clusters of nodule-specific genes are called symbiosis-related islands (SRIs). Almost 300 SRIs were identified in the *M. truncatula* genome, which on average contain more than seven nodule-enhanced genes and more than five distinct gene families. Besides the above-mentioned genes, several other well-known, important symbiotic genes are located within SRIs (for example, *RHIZOBIUM-DIRECTED POLAR GROWTH [RPG]*, *ENOD40*, *ENOD11*, *ENOD12*, *INCOMPLETE ROOT HAIR ELONGATION [IRE]*, *SymCRK*, *RSD*, and *SYMBIOTIC REMORIN1[SYMREM1]*). This clustered gene organization is thought to contribute to the nodule-specific regulation of SRI-localized genes (Pecrix et al., 2018).

The Evolution of Nodulation

To date, no systematic meta-analysis has been performed that compares the details of the transcriptomes of all analyzed NFC plants to identify (for example) the common nodule-enhanced genes. The upstream regulatory hierarchy that orchestrates the genetic program underlying nodulation in the host, the CSSP, is functionally conserved between legumes, *Parasponia*, and actinorhizal plants (Gherbi et al., 2008; Markmann et al., 2008; Svistoonoff et al., 2013; Clavijo et al. 2015; Granqvist et al., 2015; Battenberg et al. 2018; van Velzen et al., 2018; van Zeijl et al., 2018). However, the extent of conservation of the other actors in nodulation downstream of these master regulators is currently unknown. Nevertheless, analyses have recently started to address this question by defining orthologous groups between NFC plant species and identifying common nodule-enhanced genes. One study compared root and nodule transcriptomes in the actinorhizal plants *C. thysiflorus* and *D. glomerata* and the legume *M. truncatula* and identified a set of only 51 genes commonly induced in the nodules of these three NFC plants (Battenberg et al. 2018). A comparison focusing on the low-entropy nodule-specific

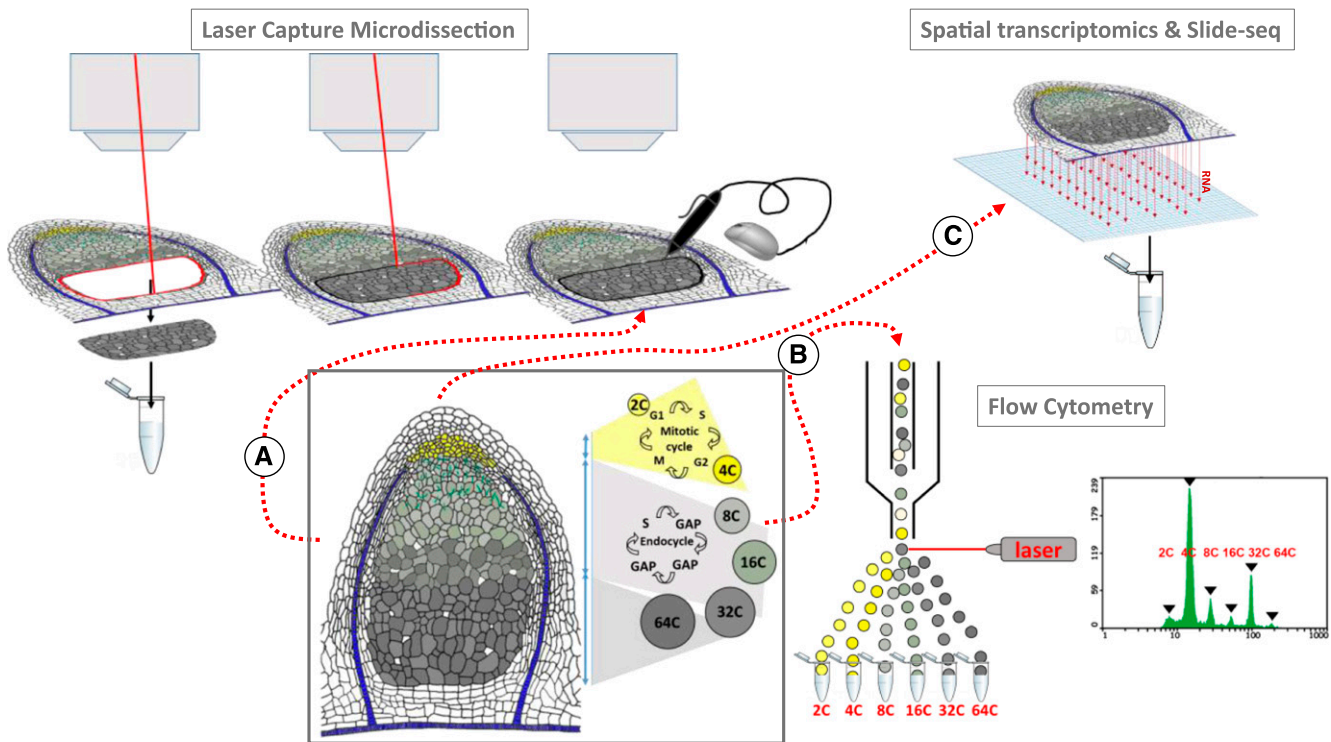


Figure 4. Spatial Resolution of Gene Expression Analysis in Nodules.

(A) Purified nodule cell types can be obtained by LCM. This technology uses a laser guided by a microscope to cut cells in a precise, contamination-free manner out of their surrounding tissues from an organ section.

(B) A second methodology exploits the varying DNA content in the nuclei of differentiating symbiotic nodule cells and separates these nuclei using flow cytometry and a cell sorter. Flow-cytometry is a laser-assisted technology that analyzes the biophysical features of small particles, such as cells or nuclei, in a capillary stream. The graph on the right shows a typical flow cytometry measurement of the DNA content in nodule nuclei. A cell sorter coupled to the flow cytometer separates these nuclei. Both methods, LCM and flow cytometry, yield purified materials that are compatible with downstream molecular biology methods such as gene expression analysis (RNA-seq, microarray hybridization, RT-quantitative PCR) and chromatin analysis (DNA methylation analysis, chromatin immunoprecipitation-seq). Both methods can be used to isolate single cells or single cell types from nodules and can also be combined with the use of fluorescent markers expressed in cells of interest.

(C) In spatial transcriptomics or Slide-seq, a nodule section is placed on a support that carries a defined array of barcodes. After the RNA is transferred from the tissue to the array, the barcode sequence is introduced during cDNA synthesis, and RNA-seq libraries are produced from the whole sample, preserving the spatial information through barcodes.

gene sets in soybean, *L. japonicus*, *P. vulgaris*, and *M. truncatula* found only 12 common orthologous genes between the four species and 34 to 70 common genes in pairwise comparisons (Wu et al., 2019). These results suggest that nodule-specific gene sets appeared late in evolution and independently in the legume clades. However, another study comparing the *P. andersonii*-nodule transcriptome obtained by RNA-seq with comparable *M. truncatula* data found a more substantial overlap (van Velzen et al., 2018). Among the 1719 and 2753 nodule-enhanced genes in *P. andersonii* and *M. truncatula*, respectively (Roux et al., 2014; genes that are upregulated at least twofold in young mature nodules vs. uninfected roots), 290 putative orthologs were found. This common set of genes activated in *P. andersonii* and *M. truncatula* nodules includes genes encoding several known symbiosis regulators, signaling proteins, and TFs, as well as metabolic proteins such as enzymes and metabolite transporters. The common genes were further classified into two functional categories based on their expression profiles in nitrogen-fixing *P.*

andersonii nodules and in uninfected nonfunctional nodules. Of the 290 common genes, 126 were upregulated in both functional and nonfunctional nodules, suggesting they play roles related to nodule organogenesis. The 164 remaining genes were only induced in infected and functional nodules and are therefore associated with infection or nitrogen fixation (van Velzen et al., 2018).

At first glance, the list of 290 common genes between these two nodule transcriptomes appears to be relatively small in light of the 1719 and 2753 nodule-enhanced genes in these species, suggesting that the functions of these nodules primarily involve different gene sets. However, the two plants diverged ~100 MYA, which may have brought about this functional divergence. Moreover, this comparison has to be placed in the context of the evolution of nodulation and symbiotic nitrogen fixation. Two opposing evolutionary models for nodulation have been proposed (Figures 5A and 5B). Until recently, the prevailing hypothesis (based on phylogenetic ancestral state reconstruction and the strong differences in nodules of different lineages) considered the

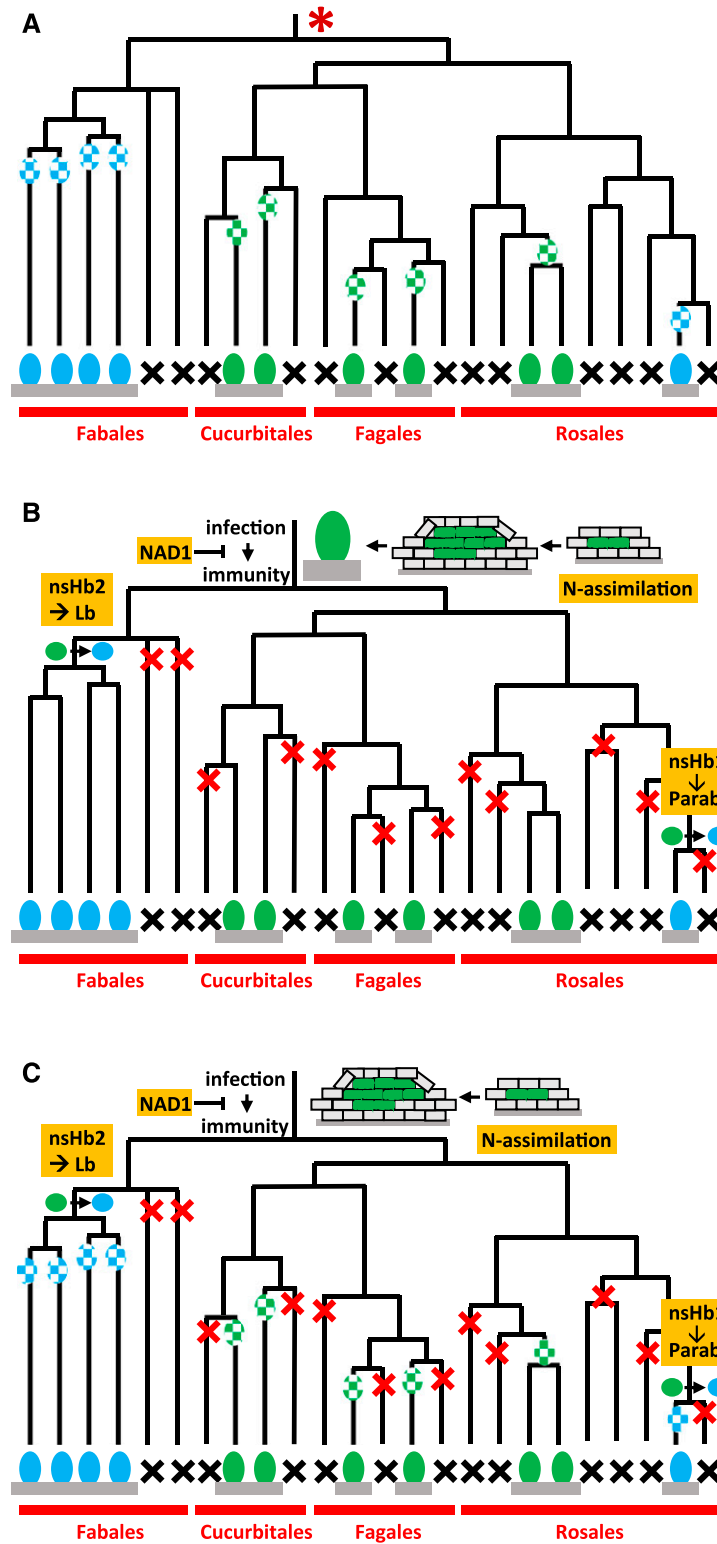


Figure 5. Models for the Evolution or Nodulation in the Nitrogen-Fixing Clade.

(A) The convergent evolution hypothesis proposes the evolution of a cryptic innovation or predisposition for nodulation (red asterisk) followed by independent acquisitions of nodule formation (indicated by colored dots on tree branches). Green and blue dots signify the acquisition of nodulation with *Frankia* and *Rhizobium*, respectively.

evolution of nodulation to be a two-step process. The initial step was the acquisition of an uncharacterized predisposition for nodulation by the ancestor of the NFC over 100 MYA, which was followed (at least 30 MY later) by the independent acquisition or convergent evolution of nodulation in six to ten different lineages of the NFC (Figure 5A; Soltis et al., 1995; Werner et al., 2014; Doyle, 2016). This hypothesis implies multiple independent recruitments for nodulation of the CSSP, that preexisted and functioned in mycorrhization in these lineages, an evolutionary scenario known as deep homology that has repeatedly been identified for animal structures (e.g. eyes and limbs; Shubin et al., 2009). Under this scenario, nodulation evolved independently in *Medicago* and *Parasponia*, implying that the 290 common genes in *P. andersonii* and *M. truncatula* nodules arose through complex convergent evolution.

Therefore, van Velzen et al. proposed the challenging hypothesis that nodulation arose already in the common ancestor of the NFC (van Velzen et al., 2018, 2019). Besides the substantial overlap in the nodule transcriptomes of *P. andersonii* and *M. truncatula*, including 126 putative organogenesis genes, this new hypothesis is also based on the independent and recent losses or pseudogenizations of the essential symbiosis genes *NFP*, *NIN*, and *RPG* in nonnodulating plants of multiple lineages of the NFC (Griesmann et al., 2018; van Velzen et al., 2018). Together, these observations can be interpreted as a single gain of nodulation at the origin of the NFC and subsequent multiple losses in the current nonnodulating lineages (Figure 5B). Since the divergence of the four orders of the NFC occurred very rapidly (~104 MYA) after the divergence of the NFC itself from its sister clades (~110 MYA), a period of only a few MYs (Wang et al., 2009; Li et al., 2015), this hypothesis implies that nodule organogenesis evolved in a very short time, within these few MYs. Alternatively, perhaps nodulation and nitrogen-fixing symbiosis are much more ancient than previously thought and have been lost altogether in the NFC sister clades.

A possible evolutionary scenario leading to nodule formation is that the existing interaction with mycorrhiza in the ancestor for nodulation was extended to a bacterial nitrogen-fixing symbiont, resulting in the infection of root cortical cells with bacteria rather than fungi. This interaction used the already present CSSP and recruited additional genes. One of the new key functions could have been the suppression of immunity against bacteria in infected root cells. NAD1, a conserved protein specific to the NFC, might represent an early recruited immunity suppressor (Wang

et al., 2016). Perhaps the ammonium channel NOD26, which is involved in the uptake of ammonium produced by bacteroids from the symbiosomes into the symbiotic cells, represents another early innovation (Frare et al., 2018). This transporter, along with ammonium assimilation enzymes, is required for the metabolic integration of the nitrogen fixing symbiont. At the regulatory level, the TF NIN, which is required for nodule organogenesis and infection in present-day plants, might have been recruited early on by the CSSP from a preexisting TF family of NIN-like proteins (NLPs) implicated in the regulation of nitrogen homeostasis in plants (Soyano and Hayashi, 2014; Clavijo et al., 2015). Another innovation must have been the activation of the cell cycle in differentiated root cells to increase the number of infected cells by cell proliferation and to direct organogenesis. This step could have been accomplished through the recruitment of cytokinin signaling by NIN, a crucial step in the initiation of root cortical cell division in legumes (Schauser et al., 1999; Vernié et al., 2015; Gamas et al., 2017), although the role of cytokinin signaling in nodulation has not been confirmed in the other clades of the NFC (van Zeijl et al., 2018; Gauthier-Coles et al., 2019). Subsequently, the evolution of nodulation may have occurred via a gradual process that created an ever increasing mass of infected symbiotic cells, providing more and more nitrogen to the host until the full nitrogen needs of the plant were met (Figure 5B).

The new model of van Velzen et al. (2019) further proposes that symbiosis with *Frankia* occurred at the NFC origin and subsequent independent switches to rhizobia occurred in the ancestor of legumes and *Parasponia*. This proposition is based on the finding that legumes and *Parasponia* use paralogous and not homologous hemoglobin genes in nodules to protect the oxygen-sensitive nitrogenase enzyme of the rhizobium symbionts (Sturms et al., 2010). Nitrogen fixation by *Frankia* does not require an external oxygen scavenger because these bacteria are able to create their own oxygen barrier by forming a shell of hopanoid lipids. Legumes recruited class 2 hemoglobins and *Parasponia* independently recruited a class 1 hemoglobin to function as oxygen transporters in nitrogen-fixing symbiotic nodule cells. Thus, in this second evolutionary scenario, nodulation with *Frankia* in proto-legumes and *Parasponia* shares a relatively short common evolutionary history from the origin of nodulation to the diversification of the two lineages, followed by 100 MY of separation during which both lineages independently switched from hosting *Frankia* to rhizobia. The symbiont exchange could have been achieved through a temporary phase of coinfection of nodules by

Figure 5. (continued).

(B) The model of a single acquisition of nodulation with *Frankia* at the origin of the NFC and multiple losses in the nonnodulating lineages (red crosses). This acquisition required the suppression of PAMP-triggered immunity against the infecting bacteria, a process possibly involving the NAD1 protein. The evolution of nodulation might have involved a rapid succession of stepwise innovations including the replacement of mycorrhizal fungal infection by *Frankia* infection, followed by cortical cell divisions resulting in the production of small foci of infected cells before the development of nodules. Two independent switches of symbionts occurred (indicated by green to blue dot), once in the *Parasponia* lineage of the Rosales and once in the ancestor of the legumes in the Fabales. These switches to a *Rhizobium* symbiont were possible due to the recruitment and mutation of a nonsymbiotic, nonoxygen transporting class 1 or class 2 hemoglobin (nsHb1 or nsHb2), forming the oxygen-transporting *Parasponia* hemoglobin (Parab) or leghemoglobin (Lb), respectively.

(C) An alternative model proposes that the ancestor of the NFC acquired the possibility to extend mycorrhizal infection with *Frankia* infection of root cortical cells. The subsequent stages imply multiple independent acquisitions of nodule organogenesis as in **(A)** and two independent symbiont switches as in **(B)**. The trees are schematic representations based on published chronograms (Doyle, 2016). The present status of clades is indicated by a blue or green nodule for *Rhizobium*-infected or *Frankia*-infected nodulating clades and by a black cross for nonnodulating clades.

the primary *Frankia* symbiont and a secondary rhizobium symbiont, followed by the full overtaking of *Frankia* by the rhizobia. Indeed, coinfection of nodules by a nodule-inducing bacterium and a nonnodulating bacterium is often observed in legumes and actinorhizal plants (Martínez-Hidalgo and Hirsch, 2017). Furthermore, the replacement of the symbiont likely required new host functions for recognizing the new symbionts (which have a distinct cell envelope) and for suppressing immune responses against the rhizobia. From the bacterial side, the rhizobia acquired the nodulation genes required for Nod factor synthesis (Figure 1A), although the origin of these genes is still an enigma. Nod factors were required to replace the *Frankia* factors that triggered the nodulation program, and they also helped suppress immune responses by the host (Liang et al., 2013).

One could also envision a third, in-between scenario in which the ancestor of the NFC established an interaction with *Frankia* spp without nodule formation and organogenesis, instead forming nitrogen-fixing structures similar to arbuscular mycorrhiza (structures formed without organogenesis) with little or no cell division, facilitating bacterial uptake and forming infection foci (Figure 5C; see Parniske 2018 for an inspiring discussion on this scenario). In other words, in this third scenario, the hypothetical “predisposition of nodulation” of scenario 1 would correspond to the replacement (or extension) of a mycorrhizal fungus with a filamentous bacterial symbiont of the *Frankia* genus. During the second stage, *Frankia* were replaced by rhizobia independently in legumes and *Parasponia*, as described in scenario 2. Finally, during the third stage, nodule organogenesis evolved independently in multiple lineages. The timing of stage two and three could also be inverted. The present day absence of nodulation in many lineages of the NFC could correspond to the loss of nodulation or the loss of the more ancient nitrogen-fixing *Frankia* symbiosis without organogenesis. It is harder to reconcile this third scenario with the presence of putative organogenesis-related genes in the common gene set of the *Medicago* and *Parasponia* transcriptomes. However, this overlap could represent the combined result of the early shared evolution toward a simple infection focus in the root cortex and the later convergent recruitment of orthologous gene functions after the separation of the two lineages. Moreover, the three symbiosis genes (*NFP*, *NIN*, and *RPG*) that were lost in nonnodulating plants of the NFC are involved in infection and are thus not necessarily (exclusive) markers of nodule organogenesis. On the other hand, the third scenario fits well with the existing differences in infection, ontogeny, and physiology of nodules formed by plants in the different lineages of the NFC. For example, the nodules of *Parasponia* and actinorhizal plants have a central vascular bundle, whereas the nodules of legumes contain multiple peripheral vascular bundles (Behm et al., 2014; Downie, 2014). In addition, legumes and nonlegumes of the NFC differ in their reliance on cytokinin signaling (van Zeijl et al., 2018; Gauthier-Coles et al., 2019). Finally, the nodule-specific transcriptomes of different plants are largely dissimilar (Battenberg et al. 2018; Wu et al., 2019; despite the overlap of a subset of nodule-enhanced genes in *Medicago* and *Parasponia*; van Velzen et al., 2018), suggesting that organogenesis in these plants followed distinct evolutionary paths.

Investigating whether the large overlap between nodule transcriptomes observed between *P. andersonii* and *M. truncatula*

could be extended to all nodulating plants could be informative. The available nodule transcriptomes from legumes cover all the major clades of the Papilionoideae subfamily, the hologalegina (*Medicago*, *Pisum*, *Lotus*, *Sesbania*), the millettoids (*Phaseolus*, *Glycine*), and the genistoids (*Lupinus*), but no data are available for species of the other large clade of the Leguminosae, the Mimosoideae-Caesalpineae-Cassieae clade. In the nonlegume NFC plants, the Cucurbitales are represented by the *Datsica* transcriptome, the Fagales by the *Alnus* and *Casuarina* transcriptomes, and the Rosales by *C. thyrsoiflorus* and the rhizobium-infected *Parasponia* transcriptomes. A systematic comparative analysis identifying orthologous genes in the available transcriptome data, complemented with new data from the unexplored branches of the legume family, would certainly provide insights into the evolutionary scenario of nodulation. Such analysis would further highlight the mechanisms in nodulation that are ancestral and those that are more recent innovations specific to particular lineages of the NFC.

REGULATION OF THE NODULE TRANSCRIPTOME

Transcription Factors

The stimulation of the CSSP by Nod factors results in the activation of a TF network composed of IPD3 (CYCLOPS), IPD3L, NIN, NSP1, NSP2, DELLA proteins, ERN1, ERN2, NF-YA1, and potentially other yet-to-be identified TFs (Schauser et al., 1999; Kaló, 2005; Smit et al., 2005; Combier et al., 2006; Andriankaja et al., 2007; Marsh et al., 2007; Middleton, 2007; Asamizu et al., 2008; Cerri et al., 2012; Singh et al., 2014; Fonouni-Farde et al., 2016; Jin et al., 2016; for reviews on the topic, see Oldroyd, 2013; Soyano and Hayashi, 2014; Jin et al., 2018; Roy et al., 2019). This TF network activates the infection process, nodule organogenesis, and autoregulation of nodulation (a long-distance negative-feedback pathway that controls nodule number based on the number of already formed or initiated nodules and the nitrogen needs of the plant; Figure 1A). The NIN TF emerged from these studies as a master regulator of nodulation that coordinates these processes (Soyano et al., 2013; Vernié et al., 2015; Liu et al., 2019a, 2019b), which is in agreement with the proposed early recruitment of NIN during the evolution of nodulation in the NFC. Several direct target genes of this network have been identified (e.g. *ENOD11*, *CRE1*, *NPL1* [*NODULATION PECTATE LYASE 1*], *CLE*), but how this TF network further connects to the downstream nodule transcriptome and the massive reprogramming in the symbiotic cells (Figure 2B) remains to be clarified. This regulation most likely involves the generation of secondary signals (such as phytohormones and others) as well as a downstream TF network.

Nevertheless, some TFs involved in late stages of nodule formation have been identified, but they have no direct link to transcription in symbiotic cells or their roles remain to be pinpointed. A few early studies have identified potential regulators in symbiotic nodule cells, which were not further characterized (Heard and Dunn, 1995; Heard et al., 1997; Jørgensen et al., 1999). Other transcriptional regulators, such as the *Medicago* Krüppel TF MtZPT2, the *Medicago* basic helix-loop-helix TF MtbHLH1, NOOT in *Medicago* and *Lotus* (belonging to the NPR1-like BTB/POZ-

Ankyrin domain protein family), and soybean GmbHLHm1 (formerly GmSAT1), are important for nodule formation but are not directly involved in symbiotic cell-specific gene regulation. Their genes are not expressed in symbiotic cells but are instead expressed in roots and nodule vascular bundles or nodule parenchyma, where they affect nodule vasculature patterning and nutrient transport between nodules and roots (Frugier et al., 2000; Godiard et al., 2011; Couzigou et al., 2012; Chiasson et al., 2014; Magne et al., 2018).

Other characterized TFs have demonstrated roles in symbiotic cell formation. The TF EFD, belonging to the ethylene response factor family, controls symbiotic cell formation, bacteroid differentiation, and the expression of a subset of nodule-specific genes, including some NCR genes (Vernié et al., 2008). However, this regulation of nodule-specific genes could be indirect, as *MtRR4* is the only gene found to be directly regulated by EFD (Vernié et al., 2008). *MtRR4* is a type A cytokinin response regulator induced by cytokinins that is part of a negative feedback loop involved in repressing the cytokinin pathway. Cytokinins are key signals that activate cell divisions in the root cortex to establish a nodule primordium (Gamas et al., 2017). The repression of cytokinin signaling mediated by EFD through *MtRR4* might be essential for promoting symbiotic cell differentiation. Furthermore, a group of three class 2 KNOX homeodomain TFs directly or indirectly regulate *EFD* and *MtRR4* expression (Azaraksh et al., 2015; Di Giacomo et al., 2017). One of these KNOX TFs, *MtKNOX3*, is strongly induced in nodules and is active in differentiating symbiotic cells. This TF could be involved in regulating other genes, including symbiotic cell-specific genes.

The C₂H₂ TF RSD is expressed in differentiating symbiotic cells and is required for symbiosome formation. However, data suggest that RSD act as a repressor of a small set of genes to influence vesicle transport in symbiotic cells and avoid mis-targeting of specific vesicles and their cargo to symbiosomes (Sinharoy et al., 2013). Thus, the present data on RSD are not in agreement with a role of this TF in positively activating symbiotic cell-specific gene expression. Similarly, the *Medicago* TF MtbHLH2 might be expressed in infected nodule cells but is thought to act as a negative regulator of genes that stimulate nodule senescence (Deng et al., 2019). A small set of APETALA2 TFs in soybean, *L. japonicus*, and *P. vulgaris* might have a negative effect on nodulation, including symbiotic cell functioning. These TF were identified because they are targeted by a nodule-enhanced microRNA (miRNA), and their upregulation in symbiotic cells may be correlated with the activation of senescence-associated genes (Wang et al., 2014; Holt et al., 2015; Nova-Franco et al., 2015).

In a large-scale effort in soybean, the expression profiles of ~1000 genes encoding putative transcriptional regulators were analyzed during nodule formation via RT-qPCR (Libault et al., 2009a). This systematic screening identified 39 genes that are upregulated during nodule formation. A gene encoding a MYB TF was found to be expressed in vascular tissues and not in symbiotic nodule cells. Although the remaining genes were not further characterized, they constitute a collection of putative regulators in symbiotic nodule cells.

Overall, very little is thus known about how the massive transcriptome switch in symbiotic nodule cells is directly regulated. Libault et al. (2009b) and Young et al. (2011) cataloged all TFs in the

M. truncatula genome. Analyzing the MtGEA and Symbimics databases using nodule-specific or nodule-enhanced expression and expression in symbiotic cells (IZ and zone III) as criteria resulted in the identification of a set of interesting TFs that might be involved in gene-specific regulation in symbiotic cells (Supplemental Figure). These TFs include NIN (the above-mentioned master regulator of nodulation) and IPD3. IPD3 directly interacts with the CSSP component calcium- and calmodulin-dependent protein kinase (CCaMK) and functions in Nod factor activation of the CSSP, infection thread formation, and the release of rhizobia from infection threads, organogenesis, and symbiotic cell formation (Yano et al., 2008; Maunoury et al., 2010; Horváth et al., 2011; Ovchinnikova et al., 2011; Singh et al., 2014). The transcriptome data show that *IPD3* is most strongly expressed in differentiating and mature symbiotic cells of the interzone and zone III. This finding corresponds with the results of *IPD3* promoter-GUS fusion analysis (Messinese et al., 2007). *CCaMK* is also expressed in these nodule zones but has a uniform expression pattern across all nodule tissues. These expression patterns suggest that *IPD3* and *NIN* also play important roles in controlling gene expression in differentiating and mature symbiotic cells. Therefore, *NIN* and *IPD3* might regulate symbiotic cell-specific genes directly or indirectly (through downstream TFs). In agreement with this hypothesis, the nodule transcriptome is not activated in the nodule-like structures formed in the *M. truncatula ipd3* mutant TR3 (Maunoury et al., 2010). Interestingly, genetic analyses also uncovered a role for other components of the CSSP in multiple stages of nodulation (Moling et al., 2014; Capoen et al., 2005; Limpens et al., 2005). These observations suggest that the CSSP is a regulatory module that is used and re-used to control all stages of nodulation, from the first response to Nod factors to infection thread formation to the endocytotic uptake of rhizobia and to symbiotic cell differentiation and functioning. During each step, this pathway appears to activate different gene sets depending on the cellular context.

Cis-Regulatory Elements in the Promoters of Nodule-Specific Genes

Leghemoglobin genes, the prototypical symbiotic cell-specific genes, were the first legume genes cloned (Sullivan et al., 1981). Their identification was rapidly followed by the identification of other nodule-specific genes encoding nodulins (Legocki and Verma, 1980; Sandal et al., 1987). The promoters of these genes were experimentally screened for *cis*-acting elements required for high level, nodule-specific expression (Stougaard et al., 1987; Jensen et al., 1988; Metz et al., 1988; Szabados et al., 1990; Szczygłowski et al., 1994; Frühling et al., 2000; Nallu et al., 2013). These elements were used to demonstrate the existence of nodule-specific trans-acting factors. Unfortunately, these efforts were not further pursued, and the putative TFs were not characterized (Jensen et al., 1988; Metz et al., 1988).

The precise identification of *cis*-acting elements should in principle facilitate the straightforward screening of TFs using methods such as yeast-one-hybrid analysis. Thus far, this strategy has not led to the identification symbiotic cell-specific genes. Such an approach has only been performed on *cis*-acting elements of the Nod factor-responsive promoter of the very early

activated gene *ENOD11*, resulting in the identification of the TF ERN1 (Andriankaja et al., 2007). Certainly, this approach holds promise for elucidating the transcriptional network in symbiotic nodule cells.

Chromatin-Based Regulation of Nodule-Specific Gene Expression

The tight regulation of nodule-specific genes clearly requires specific TFs, but in addition, regulation at the chromatin level might also be important for keeping these genes silent during all stages of plant development. Genes with high tissue-specific expression are often actively silenced throughout much of the plant lifecycle via epigenetic mechanisms. Chromatin-based silencing of nodule-specific genes could be released during symbiotic cell differentiation, making them accessible to TFs.

The soybean protein CPP1 (cysteine-rich polycomb-like protein 1) is a DNA binding protein that interacts with the promoter of the leghemoglobin gene *Gmlbc3* (Cvitanich et al., 2000). CCP1 contains conserved CXC domains homologous to a domain in some polycomb proteins. CCP proteins are widely distributed in plants and animals where they form small families. CCP proteins Tso1 in *Arabidopsis*, TESMIN in mammals, Mip120 and Tombola in *Drosophila*, and LIN-54 in *Caenorhabditis elegans* function in chromatin complexes and are involved in cell cycle regulation, cell differentiation, and reproduction (Sijacic et al., 2011; Beall et al., 2007; Jiang et al., 2007). Despite some initial characterization in various organisms, much remains to be learned about the mechanisms of these transcriptional regulators and their biological roles. Soybean CCP1 binds the *Gmlbc3* promoter and represses its expression, which is in agreement with the notion that CCP1 is part of a chromatin-related repression complex.

Further evidence for the key role of chromatin-based regulation in the massive gene activation during symbiotic cell differentiation in *M. truncatula* was obtained by analyzing the chromatin features of nodule-specific genes during symbiotic cell differentiation. This differentiation is driven by endoreduplication whereby post-meristematic nodule cells switch their cell cycle to an endocycle and multiply their genome from a 2C/4C DNA content in meristem cells, over 8C and 16C in differentiating cells, to 32C and 64C in maturing and mature symbiotic cells (Vinardell et al., 2003). Gene expression analysis of purified nodule nuclei of increasing ploidy levels revealed a correlation between the ploidy level and the sequential activation of nodule-specific genes in these nuclei, suggesting that endoreduplication is involved in this process (Nagyimihály et al., 2017). The expression levels of nodule-specific genes in nodule nuclei with different ploidy levels correlated with the opening of chromatin as well as with reduced H3K27me3 (trimethylation of Lys 27 in histone H3) levels in a subset of tested genes (Nagyimihály et al., 2017). Similarly, the above-mentioned SRIs showed reduced H3K27me3 levels in nodules versus roots (Pecrix et al., 2018). H3K27me3 is a repressive chromatin mark introduced by the Polycomb complex, a regulatory complex involved in the epigenetic silencing of chromatin. Thus, the activation of nodule-specific genes could rely on the opening of the chromatin and the release of H3K27me3-mediated repression during symbiotic cell differentiation.

DNA methylation represents another level of epigenetic control of the expression of symbiotic nodule cell-specific genes. DNA methylation of promoters usually leads to gene silencing. A comparison of DNA methylation levels in whole roots versus nodules or in nodule nuclei of different ploidy levels showed that many symbiotic nodule cell-specific genes, including genes encoding NCRs, calmodulin-like proteins, leghemoglobin, GRPs, and many others, as well as many transposons, become demethylated, which correlates with the activation of their expression (Satgé et al., 2016; Nagyimihály et al., 2017). Remarkably, SRIs show differential methylation patterns in roots versus nodules, and the borders of the SRIs sharply define the limits of differentially methylated regions, with the flanking regions showing very low and nondifferential methylation (Pecrix et al., 2018). The DNA demethylation of nodule-specific genes is largely regulated by the activation of the DNA demethylase gene *DEMETER* (*DME*) during symbiotic nodule cell differentiation (Satgé et al., 2016). Remarkably, *DME* activity also results in the loss of DNA methylation and the expression of *NCR*-flanking transposons during symbiotic cell differentiation (Guefrachi et al., 2014; Satgé et al., 2016). Since transposons are epigenetically silenced, the activation of these transposons in symbiotic cells confirms the notion that the restructuring of chromatin occurs in these loci during symbiotic cell differentiation.

Thus, epigenetic control via DNA methylation and histone modifications appears to maintain the nodule-specific genes and SRIs silenced. During the endoreduplication-driven differentiation of symbiotic cells, epigenetic reprogramming of these loci renders them accessible and competent for transcription. However, the current data only establish correlations between gene expression, chromatin modifications, and endoreduplication. A future challenge will be to establish causal relationships between these processes.

The Regulation of Gene Expression by Alternative Splicing, Small RNAs, and Small Peptides

Additional mechanisms that control symbiotic cell-specific expression have been reported. One of these mechanisms is the alternative splicing of the *Syntaxin 132* (*SYP132* or *SYP131I*) gene (Pan et al., 2016; Huisman et al., 2016). *SYP132* encodes a t-SNARE protein that functions in the secretory pathway controlling the fusion of secretory vesicles with the membrane of the target organelle. In *M. truncatula*, *SYP132* is transcribed into two different transcripts due to alternative splicing. The *SYP132 β* (or *SYP132C*) transcript is ubiquitously expressed, while the *SYP132 α* (or *SYP132A*) transcript is preferentially expressed in symbiotic nodule cells and in arbuscular mycorrhiza cells. The two transcripts differ in their 13th terminal exon, which in the *SYP132 β* transcript is located immediately downstream of the previous exon. By contrast, in the *SYP132 α* transcript, this canonical 13th exon is skipped and replaced by an exon located ~2000 base pairs downstream. The two encoded proteins are therefore almost identical, differing only in 60 C-terminal amino acids. RNA interference experiments specifically targeting one of these transcripts revealed that *SYP132 β* is essential for plant growth, whereas nodule- and mycorrhiza-specific *SYP132 α* is essential for arbuscule formation and bacteroid differentiation. The *SYP132*

proteins are transported to the symbiosomes in symbiotic nodule cells. Downregulating the *SYP132 α* transcript arrests bacteroid differentiation, presumably due to a blockage in the transport of nodule-specific secretory proteins, including NCRs, to the symbiosomes. It is not yet understood how this alternative splicing is regulated and if this is a common mechanism for nodule-specific gene expression or unique to *SYP132*. However, the finding that alternative splicing of *SYP132* is conserved in most dicots including plants not in the NFC and is absent in dicot species that have lost mycorrhizal symbiosis suggests that this mechanism is related to mycorrhiza formation and has been co-opted for nodule symbiosis.

miRNAs are small regulatory RNA molecules that control the expression of specific target genes by binding to their mRNAs and affecting mRNA translation or stability. miRNAs control early stages of nodulation before the establishment of symbiotic cells (reviewed in Couzigou and Combier, 2016). For example, the TF NF-YA1 in *M. truncatula*, which controls nodule meristem function, is negatively regulated by the miRNA miR169 (Combier et al., 2006). In addition, the class-III homeodomain-leucine zipper TFs MtHB8, MtCNA1, and MtCNA2, which are involved in root and nodule meristem formation as well as vascular bundle differentiation, are negatively regulated by miR166 (Boualem et al., 2008). In *L. japonicus* miR2111 undergoes shoot-to-root translocation and controls the autoregulation of nodulation response in roots by negatively regulating the expression of *TML*, encoding an F-box protein that suppresses symbiosis (Tsikou et al., 2018). In soybean, a whole array of miRNAs affect the early stages of nodule induction and therefore nodule number (Yan et al., 2015; Wang et al., 2014; Wang et al., 2015; Hossain et al., 2019). In agreement with this role of miRNAs, Argonaute5, a small RNA-guided silencing protein involved in mediating miRNA activity, is required for nodule induction (Reyero-Saavedra et al., 2017).

miRNAs are processed from primary transcripts resulting from the transcription of miRNA genes by RNA polymerase II. These primary miRNA transcripts often encode small peptides called miPEPs. miPEPs enhance the transcription of their cognate primary transcript and thereby the accumulation of the corresponding miRNA (Laressergues et al., 2015). In soybean, miR172c targets the APETALA 2 TF NNC1, which negatively regulates nodule number (Wang et al., 2014; Wang et al., 2019). The miPEP172c peptide is encoded by the primary transcript of miR172c. This peptide stimulates the expression of this primary transcript, thereby positively regulating nodulation (Couzigou and Combier, 2016). Regulatory small peptides can also be encoded by protein encoding mRNAs. An alternatively spliced mRNA of the TF NF-YA1 encodes such a peptide, called uORF1p (Combier et al., 2008b). The alternatively spliced mRNA and uORF1p are expressed in nodule tissues adjacent to the meristem. uORF1p negatively regulates the expression of *NF-YA1* and thus, in combination with the above-mentioned miR169, restricts the expression of this gene to the nodule meristem.

Thus, the initial stages of nodule formation are regulated by miRNAs and small peptides that control the transcript levels of their target TFs. Later stages of the nodule development are also regulated by miRNAs in *L. japonicus*, soybean, and *P. vulgaris*. In *L. japonicus*, miR397 is systemically induced only by the presence of functional, nitrogen-fixing nodules and it downregulates

a member of the laccase protein family (De Luis et al., 2012). miR397 is conserved in many plants and is involved in regulating copper homeostasis. However, whether this regulation takes place in symbiotic cells or in other nodule cells remains to be confirmed. The *P. vulgaris* miRNA homolog of the above-mentioned miR172c plays a regulatory role in a late stage of nodulation. miR172c represses the expression of the APETALA2 TF gene *AP2-1* (Nova-Franco et al., 2015). An analysis of the roles of AP2-1 and miR172c suggested that AP2-1 activates senescence-related genes in symbiotic cells and that miR172c is required to downregulate *AP2-1* in active nodules. By contrast, in nonactive nodules infected with a nonfunctional rhizobium strain or induced by nitrate application, miR172c expression is reduced, leading to the induction of *AP2-1* and senescence. Corresponding miRNAs in soybean and *L. japonicus* are expressed in symbiotic cells of mature nodules; these miRNA172s and their AP2 targets could play similar roles to those in *P. vulgaris* (Wang et al., 2014; Holt et al., 2015). Finally, miR393j-3p targets and represses the expression of *ENOD93* in soybean; this gene is specifically expressed in nodules and symbiotic cells (Yan et al., 2015).

Small RNAs derived from endonuclease-mediated tRNA cleavage (tRNA fragments or tRFs) are a recently described class of small RNAs that associate with Argonaute proteins and regulate gene expression (Kumar et al., 2016). A recent analysis of small RNA transcriptomes from soybean nodules and cultures of its symbiotic partner *Bradyrhizobium* revealed numerous *Bradyrhizobium*-derived tRFs, most with higher abundance in nodules than in culture (Ren et al., 2019). Twenty-five of these tRFs were predicted to target 52 different soybean genes, suggesting that they contribute to the posttranscriptional regulation of these genes. Three selected rhizobial tRFs were indeed shown to be loaded on soybean Argonaute protein 1 and to downregulate their predicted target genes *GmRHD3a/GmRHD3b*, *GmHAM4a/GmHAM4b*, and *GmLRX5*. The silencing of these target genes by the rhizobial tRFs occurred during the early stages of infection and enhanced nodulation. These findings point to a new level of interspecies communication between the rhizobia and their legume hosts. The finding that additional *Bradyrhizobium* tRFs are predicted to target other soybean genes and that a different set of rhizobial tRFs potentially targeting a different set of host genes was identified in a small RNA transcriptome from the legume *P. vulgaris* suggests that this type of signaling is not restricted to these three reported examples. Although it is too early to fully understand the implications of this new signaling mechanism, it is tempting to speculate that it introduces yet another level of control to host-specificity to match the right symbiont with the right plant. Another intriguing question is how these rhizobial small RNAs reach their target genes in the plant cell. An exciting possibility is that they might be delivered via outer membrane vesicles produced by the rhizobia, a mechanism similar to that used by the human pathogen *Pseudomonas aeruginosa*, which delivers a tRF via outer membrane vesicles to epithelial cells in the human airway (Koeppen et al., 2016).

Taken together, these pioneering studies have made it clear that miRNAs and other small RNAs play a role in coordinating and fine-tuning gene expression in symbiotic nodule cells. Accordingly, downregulating *Argonaute5* expression in *P. vulgaris* and soybean not only reduces nodule initiation but also affects the formation of

symbiotic cells (Reyero-Saavedra et al., 2017). Dedicated transcriptome analyses focusing on small RNAs expressed in nodules and infected root hairs have been performed for several legumes, including *M. truncatula* (Lelandais-Brière et al., 2009; Formey et al., 2014; Pecrix et al., 2018), soybean (Subramanian et al., 2008; Joshi et al., 2010; Yan et al., 2015; Yan et al., 2016), *P. vulgaris* (Peláez et al., 2012; Formey et al., 2016), and *L. japonicus* (Holt et al., 2015). The corresponding databases represent promising sources for the future discovery of novel RNAs that control gene expression in nodules.

Finally, long noncoding and antisense RNAs (lncRNAs) might also be involved regulating nodule-specific gene expression. LCM RNA-seq analysis of *M. truncatula* nodules (SYMBIMICS data set) has revealed a surprisingly large number (over 1000) of lncRNAs expressed in nodules. Most of these are active in differentiating or mature symbiotic cells. Many are located in SRIs, and the expression profiles of many of these lncRNAs were correlated with the expression of the closest mRNA-encoding genes in the genome, suggesting these lncRNAs play regulatory roles (Pecrix et al. 2018). For example, the TF *EFD* is expressed in young symbiotic cells in the distal and proximal infection zones of the nodule. Strikingly, an antisense transcript of *EFD* showed a complementary expression profile, that is, in older symbiotic cells in the interzone and fixation zones. This finding suggests that *EFD* is downregulated by antisense transcripts in order to promote the final stages of symbiotic cell differentiation. Indeed, a large set of the above-mentioned nodule-specific genes in *M. truncatula* (e.g. genes encoding NCRs, GRPs, ENOD8, ENOD16, ENOD20, calmodulin-like proteins, leghemoglobin, nodulin 25 IPD3, NIN, SymCRK, DME, DNF1, RSD, and ERN2) as well as many others have associated lncRNAs or antisense transcripts. The biological meaning of this will certainly be worth deciphering.

CONCLUDING REMARKS

Transcriptomic analysis of nodule formation has been instrumental in the discovery of new genes that were subsequently shown by reverse genetics to be important for nodule formation. Thus, transcriptomics have played a complementary role (and will probably continue to do so) to the many forward genetic screens that have been successfully conducted by several laboratories to identify key regulators of nodulation (reviewed by Roy et al., 2019). However, transcriptomics analysis also provides broader insights into the nodulation process. Here, we argue that nodule formation is a unique developmental process that is supported by drastic transcriptional reprogramming. This reprogramming involves the activation of many hundreds of genes, many of which are exclusively expressed in these organs and are novel genes that have been shaped by evolution since the origins of nodulation. A remarkable number of nodule-specific genes have signal peptides and are therefore transported to bacteroid-containing symbiosomes, suggesting that a major adaptation of nodule-forming plants is the massive production of a large variety of proteins that interact with endosymbionts. We have some knowledge about the functions of some of these proteins (e.g. the NCRs), but at the moment, we have no clue about most of their roles.

Even though the CSSP is the upstream regulatory apparatus that activates both nodule formation and mycorrhiza formation, the nodule-specific and mycorrhiza-specific transcriptomes are dramatically different. How the different read-outs of this conserved module are achieved is likely one of the most important questions to be solved in the field. Perhaps the components of the CSSP are arranged in multiprotein signaling modules anchored in the cell membrane through receptor complexes located in nanoclusters organized by microdomain-associated proteins, including remorins, flotillins, and FW2.2-like proteins (Haney and Long, 2010; Haney et al., 2011; Lefebvre et al., 2010; Ott, 2017; Qiao et al., 2017; Liang et al., 2018). Distinct nodulation signaling modules and mycorrhization signaling modules could activate the appropriate responses.

The regulatory mechanisms that underlie symbiotic cell formation and control the transcriptome switch are also poorly understood. With the recent identification of candidate TFs, the discovery of SRIs, the description of epigenetic modifications, and the documentation of transcriptional fine-tuning mechanisms by alternative splicing, miRNAs, miPEPs, tRFs, and antisense RNAs, we are only starting to decipher this process, and we can hope that the key regulators will soon be discovered. An intriguing observation from nodule transcriptomes is that the TF gene *IPD3* is most strongly expressed in symbiotic cells. It would be interesting to determine its specific role in these cells and how this role relates to its transcriptional activity during the very early stages of the nodulation process.

An important implication of the current analysis is that nodules are unique organs, not only from the perspective of organogenesis and histological organization but also from a genetic perspective. All attempts to transfer nodule formation to non-legume crops that are currently underway in different laboratories should take into account the finding that “naturally” evolved nodule formation relies on the concerted action of hundreds of genes, many of which appear to function specifically in nodules. It will be important in future efforts to understand in great detail not only the regulatory mechanisms of nodule-specific genes, but also their function as well as their evolution. A key issue for the future will be to understand how adaptive the nodule transcriptome response is. Gene regulation is usually considered to be an adaptive response of the organism, organ, or cell to a specific environmental or developmental condition so that genes are expressed when and where they are required. The validity of this presumption has not been tested thoroughly in plants. In bacteria, however, the picture has emerged that transcriptional responses are in large part suboptimal or even maladaptive (Price et al., 2013). Striking examples come from the rhizobium bacteroids in nodules themselves. Similar to their host cells, bacteroids undergo a major transcriptomic switch during their formation, with the induction of many hundred of genes and the downregulation of the rest of the genome. Some of these upregulated genes are essential for symbiosis, such as genes encoding nitrogenase and accessory functions. However, disappointingly from a researcher’s perspective, in various studies, the mutation of many selected genes that were strongly upregulated in mature bacteroids resulted in no symbiotic phenotype for any of the tested genes (Karunakaran et al., 2009; Lamouche et al., 2018). Similarly, in the yeast *Saccharomyces cerevisiae*,

only a small portion of genes upregulated under a particular growth condition contributed to the fitness of the yeast under that condition (Giaever et al., 2002). Thus, genes may be upregulated in a specific context only as a side-effect without having a clear functional contribution. Whether this is the case in nodules remains to be analyzed, as it is possible that part of the nodule transcriptome contributes little to symbiotic nitrogen fixation.

Clearly, it is not possible, at least with currently available technology, to engineer something as complex as nodules of the NFC. Thus, it will be important to understand to what extent the complexity present in current day nodules is required for nitrogen fixation, even at levels below the optimal level reached in legume nodules. Much attention has been paid to the early stages of nodule formation, but we would like to advocate for more intense efforts toward understanding symbiotic nodule cell formation. These unique cell types are at the heart of symbiosis and determine the high efficiency of the nitrogen-fixation process.

Supplemental Data

Supplemental Figure. Expression of *M. truncatula* TFs.

ACKNOWLEDGMENTS

The work was supported by the Agence Nationale de la Recherche (grants ANR-17-CE20-0011-02 and ANR-16-CE20-0013-03 to P.M.); the LabEx Saclay Plant Sciences-SPS (ANR-10-LABX-0040-SPS to P.M.); the Hungarian National Office for Research, Development and Innovation (grants OTKA 120120/119652 to A.K. and GINOP 2.3.2-15-2016-00014 Evomer and GINOP 2.3.2-15-2016-00015 I-KOM to E.K.); and a Balzan research grant (to E.K.). Collaboration between the two institutions is supported by the cooperative Balaton program of Campus France (39661TA) and the Hungarian National Office for Research, Development and Innovation (2017-2.2.5-TÉT-FR-2017-00002).

Received July 1, 2019; accepted November 8, 2019; published November 11, 2019.

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.
- Alves-Carvalho, S., et al. (2015). Full-length *de novo* assembly of RNA-seq data in pea (*Pisum sativum* L.) provides a gene expression atlas and gives insights into root nodulation in this species. *Plant J.* **84**: 1–19.
- Andriankaja, A., Boisson-Dernier, A., Frances, L., Sauviac, L., Jauneau, A., Barker, D.G., and de Carvalho-Niebel, F. (2007). AP2-ERF transcription factors mediate Nod factor dependent Mt *ENOD11* activation in root hairs via a novel cis-regulatory motif. *Plant Cell* **19**: 2866–2885.
- 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.
- Asamizu, E., Shimoda, Y., Kouchi, H., Tabata, S., and Sato, S. (2008). A positive regulatory role for LjERF1 in the nodulation process is revealed by systematic analysis of nodule-associated transcription factors of *Lotus japonicus*. *Plant Physiol.* **147**: 2030–2040.
- Azarakhsh, M., Kirienko, A.N., Zhukov, V.A., Lebedeva, M.A., Dolgikh, E.A., and Lutova, L.A. (2015). KNOTTED1-LIKE HOMEBOX 3: A new regulator of symbiotic nodule development. *J. Exp. Bot.* **66**: 7181–7195.
- Battenberg, K., Potter, D., Tabuloc, C.A., Chiu, J.C., and Berry, A.M. (2018). 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. *Front. Plant Sci.* **9**: 1256.
- Beall, E.L., Lewis, P.W., Bell, M., Rocha, M., Jones, D.L., and Botchan, M.R. (2007). Discovery of tMAC: A *Drosophila* testis-specific meiotic arrest complex paralogous to Myb-Muv B. *Genes Dev.* **21**: 904–919.
- Behm, J.E., Geurts, R., and Kiers, E.T. (2014). *Parasponia*: A novel system for studying mutualism stability. *Trends Plant Sci.* **19**: 757–763.
- Benedito, V.A., et al. (2008). A gene expression atlas of the model legume *Medicago truncatula*. *Plant J.* **55**: 504–513.
- Berrabah, F., Bourcy, M., Eschstruth, A., Cayrel, A., Guefrachi, I., Mergaert, P., Wen, J., Jean, V., Mysore, K.S., Gourion, B., and Ratet, P. (2014). A nonRD receptor-like kinase prevents nodule early senescence and defense-like reactions during symbiosis. *New Phytol.* **203**: 1305–1314.
- Boualem, A., Laporte, P., Jovanovic, M., Laffont, C., Plet, J., Combiér, J.P., Niebel, A., Crespi, M., and Frugier, F. (2008). MicroRNA166 controls root and nodule development in *Medicago truncatula*. *Plant J.* **54**: 876–887.
- Bourcy, M., Brocard, L., Pislariu, C.I., Cosson, V., Mergaert, P., Tadege, M., Mysore, K.S., Udvardi, M.K., Gourion, B., and Ratet, P. (2013). *Medicago truncatula* DNF2 is a PI-PLC-XD-containing protein required for bacteroid persistence and prevention of nodule early senescence and defense-like reactions. *New Phytol.* **197**: 1250–1261.
- Branca, A., et al. (2011). Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume *Medicago truncatula*. *Proc. Natl. Acad. Sci. USA* **108**: E864–E870.
- Breakspear, A., Liu, C., Roy, S., Stacey, N., Rogers, C., Trick, M., Morieri, G., Mysore, K.S., Wen, J., Oldroyd, G.E.D., Downie, J.A., and Murray, J.D. (2014). The root hair “infectome” of *Medicago truncatula* uncovers changes in cell cycle genes and reveals a requirement for Auxin signaling in rhizobial infection. *Plant Cell* **26**: 4680–4701.
- Brechenmacher, L., et al. (2008). Transcription profiling of soybean nodulation by *Bradyrhizobium japonicum*. *Mol. Plant Microbe Interact.* **21**: 631–645.
- Buhian, W.P., and Bensmihen, S. (2018). Nod factor regulation of phytohormone signaling and homeostasis during rhizobia-legume symbiosis. *Front. Plant Sci.* **9**: 1247.
- Burghardt, L.T., Guhlin, J., Chun, C.L., Liu, J., Sadowsky, M.J., Stupar, R.M., Young, N.D., and Tiffin, P. (2017). Transcriptomic basis of genome by genome variation in a legume-rhizobia mutualism. *Mol. Ecol.* **26**: 6122–6135.
- Cabeza, R., Koester, B., Liese, R., Lingner, A., Baumgarten, V., Dirks, J., Salinas-Riester, G., Pommerenke, C., Dittert, K., and Schulze, J. (2014a). An RNA sequencing transcriptome analysis reveals novel insights into molecular aspects of the nitrate impact

- on the nodule activity of *Medicago truncatula*. *Plant Physiol.* **164**: 400–411.
- Cabeza, R.A., Liese, R., Lingner, A., von Stieglitz, I., Neumann, J., Salinas-Riester, G., Pommerenke, C., Dittert, K., and Schulze, J.** (2014b). RNA-seq transcriptome profiling reveals that *Medicago truncatula* nodules acclimate N_2 fixation before emerging P deficiency reaches the nodules. *J. Exp. Bot.* **65**: 6035–6048.
- Capoen, W., Den Herder, J., Rombauts, S., De Gussem, J., De Keyser, A., Holsters, M., and Goormachtig, S.** (2007). Comparative transcriptome analysis reveals common and specific tags for root hair and crack-entry invasion in *Sesbania rostrata*. *Plant Physiol.* **144**: 1878–1889.
- Capoen, W., Goormachtig, S., De Rycke, R., Schroevers, K., and Holsters, M.** (2005). SrSymRK, a plant receptor essential for symbiosome formation. *Proc. Natl. Acad. Sci. USA* **102**: 10369–10374.
- Cerri, M.R., Frances, L., Laloum, T., Auriac, M.-C., Niebel, A., Oldroyd, G.E.D., Barker, D.G., Fournier, J., and de Carvalho-Niebel, F.** (2012). *Medicago truncatula* ERN transcription factors: Regulatory interplay with NSP1/NSP2 GRAS factors and expression dynamics throughout rhizobial infection. *Plant Physiol.* **160**: 2155–2172.
- Chen, L., et al.** (2019). A nodule-localized phosphate transporter GmPT7 plays an important role in enhancing symbiotic N_2 fixation and yield in soybean. *New Phytol.* **221**: 2013–2025.
- Chiasson, D.M., et al.** (2014). Soybean SAT1 (Symbiotic Ammonium Transporter 1) encodes a bHLH transcription factor involved in nodule growth and NH_4^+ transport. *Proc. Natl. Acad. Sci. USA* **111**: 4814–4819.
- Cissoko, M., et al.** (2018). Actinorhizal signaling molecules: *Frankia* root hair deforming factor shares properties with NIN inducing factor. *Front. Plant Sci.* **9**: 1494.
- Clavijo, F., 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.
- Clevenger, J., Chu, Y., Scheffler, B., and Ozias-Akins, P.** (2016). A developmental transcriptome map for allotetraploid *Arachis hypogaea*. *Front. Plant Sci.* **7**: 1446.
- Colebatch, G., Desbrosses, G., Ott, T., Krusell, L., Montanari, O., Kloska, S., Kopka, J., and Udvardi, M.K.** (2004). Global changes in transcription orchestrate metabolic differentiation during symbiotic nitrogen fixation in *Lotus japonicus*. *Plant J.* **39**: 487–512.
- Combiér, J.-P., Küster, H., Journet, E.-P., Hohnjec, N., Gamas, P., and Niebel, A.** (2008a). 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.
- Combiér, J.-P., de Billy, F., Gamas, P., Niebel, A., and Rivas, S.** (2008b). Trans-regulation of the expression of the transcription factor *MtHAP2-1* by a uORF controls root nodule development. *Genes Dev.* **22**: 1549–1559.
- Combiér, J.-P., Frugier, F., de Billy, F., Boualem, A., El-Yahyaoui, F., Moreau, S., Vernié, T., Ott, T., Gamas, P., Crespi, M., and Niebel, A.** (2006). *MtHAP2-1* is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in *Medicago truncatula*. *Genes Dev.* **20**: 3084–3088.
- Cope, K.R., et al.** (2019). The ectomycorrhizal fungus *Laccaria bicolor* produces lipochitooligosaccharides and uses the common symbiosis pathway to colonize *Populus* roots. *Plant Cell* **31**: 2386–2410.
- Couzigou, J.-M., et al.** (2012). *NODULE ROOT* and *COCHLEATA* maintain nodule development and are legume orthologs of *Arabidopsis BLADE-ON-PETIOLE* genes. *Plant Cell* **24**: 4498–4510.
- Couzigou, J.-M., and Combiér, J.-P.** (2016). Plant microRNAs: Key regulators of root architecture and biotic interactions. *New Phytol.* **212**: 22–35.
- Cvitanich, C., Pallisgaard, N., Nielsen, K.A., Hansen, A.C., Larsen, K., Pihakaski-Maunsbach, K., Marcker, K.A., and Jensen, E.** (2000). CPP1, a DNA-binding protein involved in the expression of a soybean *leghemoglobin c3* gene. *Proc. Natl. Acad. Sci. USA* **97**: 8163–8168.
- Czaja, L.F., Hogeckamp, C., Lamm, P., Maillet, F., Martínez, E.A., Samain, E., Dénarié, J., Küster, H., and Hohnjec, N.** (2012). Transcriptional responses toward diffusible signals from symbiotic microbes reveal MtNFP- and MtDMI3-dependent reprogramming of host gene expression by arbuscular mycorrhizal fungal lipochitooligosaccharides. *Plant Physiol.* **159**: 1671–1685.
- Czernic, P., 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.
- Damiani, I., et al.** (2016). Nod factor effects on root hair-specific transcriptome of *Medicago truncatula*: Focus on plasma membrane transport systems and reactive oxygen species networks. *Front. Plant Sci.* **7**: 794.
- de Bang, T.C., et al.** (2017). Genome-wide identification of *Medicago* peptides involved in macronutrient responses and nodulation. *Plant Physiol.* **175**: 1669–1689.
- De Luis, A., Markmann, K., Cognat, V., Holt, D.B., Charpentier, M., Parniske, M., Stougaard, J., and Voinnet, O.** (2012). Two microRNAs linked to nodule infection and nitrogen-fixing ability in the legume *Lotus japonicus*. *Plant Physiol.* **160**: 2137–2154.
- Demina, I.V., Persson, T., Santos, P., Plaszczycza, 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.
- Deng, J., Zhu, F., Liu, J., Zhao, Y., Wen, J., Wang, T., and Dong, J.** (2019). Transcription factor bHLH2 represses *CYSTEINE PROTEASE 77* to negatively regulate nodule senescence. *Plant Physiol.* **181**: 1683–1703.
- Di Giacomo, E., Laffont, C., Sciarra, F., Iannelli, M.A., Frugier, F., and Frugis, G.** (2017). KNAT3/4/5-like class 2 KNOX transcription factors are involved in *Medicago truncatula* symbiotic nodule organ development. *New Phytol.* **213**: 822–837.
- Dickstein, R., Hu, X., Yang, J., Ba, L., Coque, L., Kim, D.-J., Cook, D.R., and Yeung, A.T.** (2002). Differential expression of tandemly duplicated *Enod8* genes in *Medicago*. *Plant Sci.* **163**: 333–343.
- Domonkos, Á., et al.** (2017). NAD1 controls defense-like responses in *Medicago truncatula* symbiotic nitrogen fixing nodules following rhizobial colonization in a BacA-independent manner. *Genes (Basel)* **8**: 387.
- Downie, J.A.** (2014). Legume nodulation. *Curr. Biol.* **24**: R184–R190.
- Doyle, J.J.** (2016). Chasing unicorns: Nodulation origins and the paradox of novelty. *Am. J. Bot.* **103**: 1865–1868.
- Edgar, B.A., Zielke, N., and Gutierrez, C.** (2014). Endocycles: A recurrent evolutionary innovation for post-mitotic cell growth. *Nat. Rev. Mol. Cell Biol.* **15**: 197–210.
- El Yahyaoui, F., Küster, H., Ben Amor, B., Hohnjec, N., Pühler, A., Becker, A., Gouzy, J., Vernié, T., Gough, C., Niebel, A., Godiard, L., and Gamas, P.** (2004). Expression profiling in *Medicago truncatula* identifies more than 750 genes differentially expressed during nodulation, including many potential regulators of the symbiotic program. *Plant Physiol.* **136**: 3159–3176.
- Fonouni-Farde, C., Tan, S., Baudin, M., Brault, M., Wen, J., Mysore, K.S., Niebel, A., Frugier, F., and Diet, A.** (2016). DELLA-mediated gibberellin signalling regulates Nod factor signalling and rhizobial infection. *Nat. Commun.* **7**: 12636.
- Formey, D., Martín-Rodríguez, J., Leija, A., Santana, O., Quinto, C., Cárdenas, L., and Hernández, G.** (2016). Regulation of small RNAs and corresponding targets in Nod factor-induced *Phaseolus vulgaris* root hair cells. *Int. J. Mol. Sci.* **17**: 887.

- Formey, D., et al.** (2014). The small RNA diversity from *Medicago truncatula* roots under biotic interactions evidences the environmental plasticity of the miRNAome. *Genome Biol.* **15**: 457.
- Foucher, F., and Kondorosi, E.** (2000). Cell cycle regulation in the course of nodule organogenesis in *Medicago*. *Plant Mol. Biol.* **43**: 773–786.
- Franck, C.M., Westermann, J., and Boisson-Dernier, A.** (2018). Plant malectin-like receptor kinases: From cell wall integrity to immunity and beyond. *Annu. Rev. Plant Biol.* **69**: 301–328.
- Franssen, H.J., Xiao, T.T., Kulikova, O., Wan, X., Bisseling, T., Scheres, B., and Heidstra, R.** (2015). Root developmental programs shape the *Medicago truncatula* nodule meristem. *Development* **142**: 2941–2950.
- Frare, R., Ayub, N., Alleva, K., and Soto, G.** (2018). The ammonium channel NOD26 is the evolutionary innovation that drives the emergence, consolidation, and dissemination of nitrogen-fixing symbiosis in angiosperms. *J. Mol. Evol.* **86**: 554–565.
- Frugier, F., Poirier, S., Satiat-Jeunemaitre, B., Kondorosi, A., and Crespi, M.** (2000). A Krüppel-like zinc finger protein is involved in nitrogen-fixing root nodule organogenesis. *Genes Dev.* **14**: 475–482.
- Frühling, M., Schröder, G., Hohnjec, N., Pühler, A., Perlick, A.M., and Küster, H.** (2000). The promoter of the *Vicia faba* L. gene VEnod12 encoding an early nodulin is active in cortical cells and nodule primordia of transgenic hairy roots of *Vicia hirsuta* as well as in the prefixing zone II of mature transgenic *V. hirsuta* root nodules. *Plant Sci.* **160**: 67–75.
- Galindo-González, L., Mhiri, C., Deyholos, M.K., and Grandbastien, M.A.** (2017). LTR-retrotransposons in plants: Engines of evolution. *Gene* **626**: 14–25.
- 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.
- Gauthier-Coles, C., White, R.G., and Mathesius, U.** (2019). Nodulating legumes are distinguished by a sensitivity to cytokinin in the root cortex leading to pseudonodule development. *Front. Plant Sci.* **9**: 1901.
- Ge, Z., et al.** (2017). *Arabidopsis* pollen tube integrity and sperm release are regulated by RALF-mediated signaling. *Science* **358**: 1596–1600.
- Genre, A., Chabaud, M., Faccio, A., Barker, D.G., and Bonfante, P.** (2008). Prepenetration apparatus assembly precedes and predicts the colonization patterns of arbuscular mycorrhizal fungi within the root cortex of both *Medicago truncatula* and *Daucus carota*. *Plant Cell* **20**: 1407–1420.
- Gherbi, H., Markmann, K., Svistoonoff, S., Estevan, J., Autran, D., Giczey, G., Auguy, F., Péret, B., Laplaze, L., Franche, C., Parniske, M., and Bogusz, D.** (2008). SymRK defines a common genetic basis for plant root endosymbioses with arbuscular mycorrhizal fungi, rhizobia, and *Frankia* bacteria. *Proc. Natl. Acad. Sci. USA* **105**: 4928–4932.
- Giaeffer, G., et al.** (2002). Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**: 387–391.
- Giraud, E., et al.** (2007). Legumes symbioses: Absence of *Nod* genes in photosynthetic bradyrhizobia. *Science* **316**: 1307–1312.
- Godiard, L., Lepage, A., Moreau, S., Laporte, D., Verdenaud, M., Timmers, T., and Gamas, P.** (2011). MtbHLH1, a BHLH transcription factor involved in *Medicago truncatula* nodule vascular patterning and nodule to plant metabolic exchanges. *New Phytol.* **191**: 391–404.
- González-Sama, A., de la Peña, T.C., Kevei, Z., Mergaert, P., Lucas, M.M., de Felipe, M.R., Kondorosi, E., and Pueyo, J.J.** (2006). Nuclear DNA endoreduplication and expression of the mitotic inhibitor Ccs52 associated to determinate and lupinoid nodule organogenesis. *Mol. Plant Microbe Interact.* **19**: 173–180.
- Gourion, B., Berrabah, F., Ratet, P., and Stacey, G.** (2015). Rhizobium-legume symbioses: The crucial role of plant immunity. *Trends Plant Sci.* **20**: 186–194.
- Graham, M.A., Silverstein, K.A.T., Cannon, S.B., and VandenBosch, K.A.** (2004). Computational identification and characterization of novel genes from legumes. *Plant Physiol.* **135**: 1179–1197.
- Granqvist, E., Sun, J., Op den Camp, R., Pujic, P., Hill, L., Normand, P., Morris, R.J., Downie, J.A., Geurts, R., and Oldroyd, G.E.D.** (2015). Bacterial-induced calcium oscillations are common to nitrogen-fixing associations of nodulating legumes and non-legumes. *New Phytol.* **207**: 551–558.
- Griesmann, M., et al.** (2018). Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. *Science* **361**: eaat1743.
- Guefrachi, I., Nagymihaly, M., Pislariu, C.I., Van de Velde, W., Ratet, P., Mars, M., Udvardi, M.K., Kondorosi, E., Mergaert, P., and Alunni, B.** (2014). Extreme specificity of NCR gene expression in *Medicago truncatula*. *BMC Genomics* **15**: 712.
- Gully, D., et al.** (2018). Transcriptome profiles of Nod factor-independent symbiosis in the tropical legume *Aeschynomene evenia*. *Sci. Rep.* **8**: 10934.
- Haag, A.F., Arnold, M.F.F., Myka, K.K., Kerscher, B., Dall'Angelo, S., Zanda, M., Mergaert, P., and Ferguson, G.P.** (2013). Molecular insights into bacteroid development during Rhizobium-legume symbiosis. *FEMS Microbiol. Rev.* **37**: 364–383.
- Hakoyama, T., et al.** (2009). Host plant genome overcomes the lack of a bacterial gene for symbiotic nitrogen fixation. *Nature* **462**: 514–517.
- Haney, C.H., and Long, S.R.** (2010). Plant flotillins are required for infection by nitrogen-fixing bacteria. *Proc. Natl. Acad. Sci. USA* **107**: 478–483.
- Haney, C.H., Riely, B.K., Tricoli, D.M., Cook, D.R., Ehrhardt, D.W., and Long, S.R.** (2011). Symbiotic rhizobia bacteria trigger a change in localization and dynamics of the *Medicago truncatula* receptor kinase LYK3. *Plant Cell* **23**: 2774–2787.
- Harrison, M.J., and Ivanov, S.** (2017). Exocytosis for endosymbiosis: Membrane trafficking pathways for development of symbiotic membrane compartments. *Curr. Opin. Plant Biol.* **38**: 101–108.
- 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**: 408–411.
- Heard, J., Caspi, M., and Dunn, K.** (1997). Evolutionary diversity of symbiotically induced nodule MADS box genes: Characterization of *nmhC5*, a member of a novel subfamily. *Mol. Plant Microbe Interact.* **10**: 665–676.
- Heard, J., and Dunn, K.** (1995). Symbiotic induction of a MADS-box gene during development of alfalfa root nodules. *Proc. Natl. Acad. Sci. USA* **92**: 5273–5277.
- Heath, K.D., Burke, P.V., and Stinchcombe, J.R.** (2012). Co-evolutionary genetic variation in the legume-rhizobium transcriptome. *Mol. Ecol.* **21**: 4735–4747.
- Hochoer, V., et al.** (2011). Transcriptomics of actinorhizal symbioses reveals homologs of the whole common symbiotic signaling cascade. *Plant Physiol.* **156**: 700–711.
- Hogekamp, C., Arndt, D., Pereira, P.A., Becker, J.D., Hohnjec, N., and Küster, H.** (2011). Laser microdissection unravels cell-type-specific transcription in arbuscular mycorrhizal roots, including CAAT-box transcription factor gene expression correlating with fungal contact and spread. *Plant Physiol.* **157**: 2023–2043.

- Högslund, N., et al. (2009). Dissection of symbiosis and organ development by integrated transcriptome analysis of *Lotus japonicus* mutant and wild-type plants. *PLoS One* **4**: e6556.
- Holt, D.B., Gupta, V., Meyer, D., Abel, N.B., Andersen, S.U., Stougaard, J., and Markmann, K. (2015). micro RNA 172 (miR172) signals epidermal infection and is expressed in cells primed for bacterial invasion in *Lotus japonicus* roots and nodules. *New Phytol.* **208**: 241–256.
- Horváth, B., et al. (2011). *Medicago truncatula* IPD3 is a member of the common symbiotic signaling pathway required for rhizobial and mycorrhizal symbioses. *Mol. Plant Microbe Interact.* **24**: 1345–1358.
- Hossain, M.S., Hoang, N.T., Yan, Z., Tóth, K., Meyers, B.C., and Stacey, G. (2019). Characterization of the spatial and temporal expression of two soybean miRNAs identifies SCL6 as a novel regulator of soybean nodulation. *Front. Plant Sci.* **10**: 475.
- Huisman, R., Hontelez, J., Mysore, K.S., Wen, J., Bisseling, T., and Limpens, E. (2016). A symbiosis-dedicated SYNTAXIN OF PLANTS 13ll isoform controls the formation of a stable host-microbe interface in symbiosis. *New Phytol.* **211**: 1338–1351.
- Jardinaud, M.-F., et al. (2016). A laser dissection-RNAseq analysis highlights the activation of cytokinin pathways by Nod factors in the *Medicago truncatula* root epidermis. *Plant Physiol.* **171**: 2256–2276.
- Jensen, E., Marcker, K.A., Schell, J., and Bruijn, F.J. (1988). Interaction of a nodule specific, trans-acting factor with distinct DNA elements in the soybean leghaemoglobin lbc(3) 5' upstream region. *EMBO J.* **7**: 1265–1271.
- Jiang, J., Benson, E., Bausek, N., Doggett, K., and White-Cooper, H. (2007). Tombola, a tesmin/TSO1-family protein, regulates transcriptional activation in the *Drosophila* male germline and physically interacts with always early. *Development* **134**: 1549–1559.
- Jin, Y., Chen, Z., Yang, J., Mysore, K.S., Wen, J., Huang, J., Yu, N., and Wang, E. (2018). IPD3 and IPD3L function redundantly in rhizobial and mycorrhizal symbioses. *Front. Plant Sci.* **9**: 267.
- Jin, Y., Liu, H., Luo, D., Yu, N., Dong, W., Wang, C., Zhang, X., Dai, H., Yang, J., and Wang, E. (2016). DELLA proteins are common components of symbiotic rhizobial and mycorrhizal signalling pathways. *Nat. Commun.* **7**: 12433.
- Jørgensen, J.E., Grønlund, M., Pallisgaard, N., Larsen, K., Marcker, K.A., and Jensen, E. (1999). A new class of plant homeobox genes is expressed in specific regions of determinate symbiotic root nodules. *Plant Mol. Biol.* **40**: 65–77.
- Joshi, T., Yan, Z., Libault, M., Jeong, D.H., Park, S., Green, P.J., Sherrier, D.J., Farmer, A., May, G., Meyers, B.C., Xu, D., and Stacey, G. (2010). Prediction of novel miRNAs and associated target genes in *Glycine max*. *BMC Bioinformatics* **11** (Suppl 1): S14.
- Kaló, P., et al. (2005). Nodulation signaling in legumes requires NSP2, a member of the GRAS family of transcriptional regulators. *Science* **308**: 1786–1789.
- Kant, C., Pradhan, S., and Bhatia, S. (2016). Dissecting the root nodule transcriptome of chickpea (*Cicer arietinum* L.). *PLoS One* **11**: e0157908.
- Karmakar, K., Kundu, A., Rizvi, A.Z., Dubois, E., Severac, D., Czernic, P., Cartieaux, F., and DasGupta, M. (2019). Transcriptomic analysis with the progress of symbiosis in 'crack-entry' legume *Arachis hypogaea* highlights its contrast with 'infection thread' adapted legumes. *Mol. Plant Microbe Interact.* **32**: 271–285.
- Karunakaran, R., Ramachandran, V.K., Seaman, J.C., East, A.K., Mouhsine, B., Mauchline, T.H., Prell, J., Skeffington, A., and Poole, P.S. (2009). Transcriptomic analysis of *Rhizobium leguminosarum* biovar *viciae* in symbiosis with host plants *Pisum sativum* and *Vicia cracca*. *J. Bacteriol.* **191**: 4002–4014.
- Kazmierczak, T., Nagymihály, M., Lamouche, F., Barrière, Q., Guefrachi, I., Alunni, B., Ouadghiri, M., Ibjibijen, J., Kondorosi, É., Mergaert, P., and Gruber, V. (2017). Specific host-responsive associations between *Medicago truncatula* accessions and *Sinorhizobium* strains. *Mol. Plant Microbe Interact.* **30**: 399–409.
- Ke, D., Fang, Q., Chen, C., Zhu, H., Chen, T., Chang, X., Yuan, S., Kang, H., Ma, L., Hong, Z., and Zhang, Z. (2012). The small GTPase ROP6 interacts with NFR5 and is involved in nodule formation in *Lotus japonicus*. *Plant Physiol.* **159**: 131–143.
- Keller, J., Imperial, J., Ruiz-Argüeso, T., Privet, K., Lima, O., Michon-Coudouel, S., Biget, M., Salmon, A., Ainouche, A., and Cabello-Hurtado, F. (2018). RNA sequencing and analysis of three *Lupinus* nodulomes provide new insights into specific host-symbiont relationships with compatible and incompatible *Bradyrhizobium* strains. *Plant Sci.* **266**: 102–116.
- Kelly, S., Mun, T., Stougaard, J., Ben, C., and Andersen, S.U. (2018). Distinct *Lotus japonicus* transcriptomic responses to a spectrum of bacteria ranging from symbiotic to pathogenic. *Front. Plant Sci.* **9**: 1218.
- 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.
- Kiirika, L.M., Bergmann, H.F., Schikowsky, C., Wimmer, D., Korte, J., Schmitz, U., Niehaus, K., and Colditz, F. (2012). Silencing of the Rac1 GTPase *MtROP9* in *Medicago truncatula* stimulates early mycorrhizal and oomycete root colonizations but negatively affects rhizobial infection. *Plant Physiol.* **159**: 501–516.
- Koepfen, K., Hampton, T.H., Jarek, M., Scharfe, M., Gerber, S.A., Mielcarz, D.W., Demers, E.G., Dolben, E.L., Hammond, J.H., Hogan, D.A., and Stanton, B.A. (2016). A novel mechanism of host-pathogen interaction through sRNA in bacterial outer membrane vesicles. *PLoS Pathog.* **12**: e1005672.
- 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.
- Kudapa, H., Garg, V., Chitkineni, A., and Varshney, R.K. (2018). The RNA-Seq-based high resolution gene expression atlas of chickpea (*Cicer arietinum* L.) reveals dynamic spatio-temporal changes associated with growth and development. *Plant Cell Environ.* **41**: 2209–2225.
- Kumar, P., Kuscu, C., and Dutta, A. (2016). Biogenesis and function of transfer RNA-related fragments (tRFs). *Trends Biochem. Sci.* **41**: 679–689.
- Lamouche, F., et al. (2018). Transcriptomic dissection of *Bradyrhizobium* sp. strain ORS285 in symbiosis with *Aeschynomene* spp. inducing different bacteroid morphotypes with contrasted symbiotic efficiency. *Environ. Microbiol.* **21**: 3244–3258.
- Lang, C., and Long, S.R. (2015). Transcriptome analysis of *Sinorhizobium meliloti* and *Medicago truncatula* symbiosis using nitrogen fixation-deficient nodules. *Mol. Plant Microbe Interact.* **28**: 856–868.
- Laporte, P., Satiat-Jeunemaître, B., Velasco, I., Csorba, T., Van de Velde, W., Campalans, A., Burgyan, J., Arevalo-Rodríguez, M., and Crespi, M. (2010). A novel RNA-binding peptide regulates the establishment of the *Medicago truncatula*-*Sinorhizobium meliloti* nitrogen-fixing symbiosis. *Plant J.* **62**: 24–38.
- Larrainzar, E., 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.
- Lauressergues, D., Couzigou, J.-M., Clemente, H.S., Martinez, Y., Dunand, C., Bécard, G., and Combier, J.-P. (2015). Primary

- transcripts of microRNAs encode regulatory peptides. *Nature* **520**: 90–93.
- Lefebvre, B., et al.** (2010). A remorin protein interacts with symbiotic receptors and regulates bacterial infection. *Proc. Natl. Acad. Sci. USA* **107**: 2343–2348.
- Legocki, R.P., and Verma, D.P.S.** (1980). Identification of “Nodule-Specific” host proteins (nodulins) involved in the development of symbiosis. *Cell* **20**: 153–163.
- Lelandais-Brière, C., Naya, L., Sallet, E., Calenge, F., Frugier, F., Hartmann, C., Gouzy, J., and Crespi, M.** (2009). Genome-wide *Medicago truncatula* small RNA analysis revealed novel microRNAs and isoforms differentially regulated in roots and nodules. *Plant Cell* **21**: 2780–2796.
- Li, H.L., Wang, W., Mortimer, P.E., Li, R.Q., Li, D.Z., Hyde, K.D., Xu, J.C., Soltis, D.E., and Chen, Z.D.** (2015). Large-scale phylogenetic analyses reveal multiple gains of actinorhizal nitrogen-fixing symbioses in angiosperms associated with climate change. *Sci. Rep.* **5**: 14023.
- Liang, Y., Cao, Y., Tanaka, K., Thibivilliers, S., Wan, J., Choi, J., Kang, C., Qiu, J., and Stacey, G.** (2013). Nonlegumes respond to rhizobial Nod factors by suppressing the innate immune response. *Science* **341**: 1384–1387.
- Liang, P., Stratil, T.F., Popp, C., Marín, M., Folgmann, J., Mysore, K.S., Wen, J., and Ott, T.** (2018). Symbiotic root infections in *Medicago truncatula* require remorin-mediated receptor stabilization in membrane nanodomains. *Proc. Natl. Acad. Sci. USA* **115**: 5289–5294.
- Libault, M.** (2018). Transcriptional reprogramming of legume genomes: Perspective and challenges associated with single-cell and single cell-type approaches during nodule development. *Front. Plant Sci.* **9**: 1600.
- Libault, M., Joshi, T., Benedito, V.A., Xu, D., Udvardi, M.K., and Stacey, G.** (2009b). Legume transcription factor genes: What makes legumes so special? *Plant Physiol.* **151**: 991–1001.
- Libault, M., Joshi, T., Takahashi, K., Hurley-Sommer, A., Puricelli, K., Blake, S., Finger, R.E., Taylor, C.G., Xu, D., Nguyen, H.T., and Stacey, G.** (2009a). Large-scale analysis of putative soybean regulatory gene expression identifies a *Myb* gene involved in soybean nodule development. *Plant Physiol.* **151**: 1207–1220.
- Libault, M., Farmer, A., Brechenmacher, L., Drnevich, J., Langley, R.J., Bilgin, D.D., Radwan, O., Neece, D.J., Clough, S.J., May, G.D., and Stacey, G.** (2010). Complete transcriptome of the soybean root hair cell, a single-cell model, and its alteration in response to *Bradyrhizobium japonicum* infection. *Plant Physiol.* **152**: 541–552.
- Liese, R., Schulze, J., and Cabeza, R.A.** (2017). Nitrate application or P deficiency induce a decline in *Medicago truncatula* N₂-fixation by similar changes in the nodule transcriptome. *Sci. Rep.* **7**: 46264.
- Limpens, E., Ivanov, S., van Esse, W., Voets, G., Fedorova, E., and Bisseling, T.** (2009). *Medicago* N₂-fixing symbiosomes acquire the endocytic identity marker Rab7 but delay the acquisition of vacuolar identity. *Plant Cell* **21**: 2811–2828.
- Limpens, E., Mirabella, R., Fedorova, E., Franken, C., Franssen, H., Bisseling, T., and Geurts, R.** (2005). Formation of organelle-like N₂-fixing symbiosomes in legume root nodules is controlled by *DMI2*. *Proc. Natl. Acad. Sci. USA* **102**: 10375–10380.
- Limpens, E., Moling, S., Hooiveld, G., Pereira, P.A., Bisseling, T., Becker, J.D., and Küster, H.** (2013). Cell- and tissue-specific transcriptome analyses of *Medicago truncatula* root nodules. *PLoS One* **8**: e64377.
- Liu, C.W., Breakspear, A., Guan, D., Cerri, M.R., Jackson, K., Jiang, S., Robson, F., Radhakrishnan, G.V., Roy, S., Bone, C., Stacey, N., and Rogers, C., et al.** (2019a). NIN acts as a network hub controlling a growth module required for rhizobial infection. *Plant Physiol.* **179**: 1704–1722.
- Liu, H., Zhang, C., Yang, J., Yu, N., and Wang, E.** (2018). Hormone modulation of legume-rhizobial symbiosis. *J. Integr. Plant Biol.* **60**: 632–648.
- Liu, J., Miller, S.S., Graham, M., Bucciarelli, B., Catalano, C.M., Sherrier, D.J., Samac, D.A., Ivashuta, S., Fedorova, M., Matsumoto, P., Gantt, J.S., and Vance, C.P.** (2006). Recruitment of novel calcium-binding proteins for root nodule symbiosis in *Medicago truncatula*. *Plant Physiol.* **141**: 167–177.
- Liu, J., Rutten, L., Limpens, E., van der Molen, T., van Velzen, R., Chen, R., Chen, Y., Geurts, R., Kohlen, W., Kulikova, O., and Bisseling, T.** (2019b). A remote cis-regulatory region is required for *NIN* expression in the pericycle to initiate nodule primordium formation in *Medicago truncatula*. *Plant Cell* **31**: 68–83.
- Magne, K., George, J., Berbel Tornero, A., Broquet, B., Madueño, F., Andersen, S.U., and Ratet, P.** (2018). *Lotus japonicus* *NOOT-BOP-COCH-LIKE1* is essential for nodule, nectary, leaf and flower development. *Plant J.* **94**: 880–894.
- 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.
- Markmann, K., and Parniske, M.** (2009). Evolution of root endosymbiosis with bacteria: How novel are nodules? *Trends Plant Sci.* **14**: 77–86.
- Marsh, J.F., Rakocevic, A., Mitra, R.M., Brocard, L., Sun, J., Eschstruth, A., Long, S.R., Schultze, M., Ratet, P., and Oldroyd, G.E.D.** (2007). *Medicago truncatula* *NIN* is essential for rhizobial-independent nodule organogenesis induced by autoactive calcium/calmodulin-dependent protein kinase. *Plant Physiol.* **144**: 324–335.
- Martínez-Hidalgo, P., and Hirsch, A.M.** (2017). The nodule microbiome: N₂-fixing rhizobia do not live alone. *Phytobiomes J.* **1**: 70–82.
- Maunoury, N., et al.** (2010). Differentiation of symbiotic cells and endosymbionts are coupled to two transcriptome-switches in *Medicago truncatula* nodulation. *PLoS One* **5**: e9519.
- Mecchia, M.A., Santos-Fernandez, G., Duss, N.N., Somoza, S.C., Boisson-Dernier, A., Gagliardini, V., Martínez-Bernardini, A., Fabrice, T.N., Ringli, C., Muschietti, J.P., and Grossniklaus, U.** (2017). RALF4/19 peptides interact with LRX proteins to control pollen tube growth in *Arabidopsis*. *Science* **358**: 1600–1603.
- Mergaert, P.** (2018). Role of antimicrobial peptides in controlling symbiotic bacterial populations. *Nat. Prod. Rep.* **35**: 336–356.
- Mergaert, P., Nikovics, K., Kelemen, Z., Maunoury, N., Vaubert, D., Kondorosi, A., and Kondorosi, E.** (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.
- Messinese, E., Mun, J.H., Yeun, L.H., Jayaraman, D., Rougé, P., Barre, A., Lougnon, G., Schornack, S., Bono, J.J., Cook, D.R., and Ané, J.M.** (2007). A novel nuclear protein interacts with the symbiotic DMI3 calcium- and calmodulin-dependent protein kinase of *Medicago truncatula*. *Mol. Plant Microbe Interact.* **20**: 912–921.
- Metz, B.A., Welters, P., Hoffmann, H.J., Jensen, E., Schell, J., and de Bruijn, F.J.** (1988). Primary structure and promoter analysis of leghemoglobin genes of the stem-nodulated tropical legume *Sesbania rostrata*: Conserved coding sequences, cis-elements and trans-acting factors. *Mol. Gen. Genet.* **214**: 181–191.
- Middleton, P.H., et al.** (2007). An ERF transcription factor in *Medicago truncatula* that is essential for Nod factor signal transduction. *Plant Cell* **19**: 1221–1234.

- Miller, S.S., Driscoll, B.T., Gregerson, R.G., Gantt, J.S., and Vance, C.P. (1998). Alfalfa malate dehydrogenase (MDH): Molecular cloning and characterization of five different forms reveals a unique nodule-enhanced MDH. *Plant J.* **15**: 173–184.
- Moling, S., Pietraszewska-Bogiel, A., Postma, M., Fedorova, E., Hink, M.A., Limpens, E., Gadella, T.W., and Bisseling, T. (2014). Nod factor receptors form heteromeric complexes and are essential for intracellular infection in *Medicago* nodules. *Plant Cell* **26**: 4188–4199.
- Montiel, J., Downie, J.A., Farkas, A., Bihari, P., Herczeg, R., Bálint, B., Mergaert, P., Kereszt, A., and Kondorosi, É. (2017). Morphotype of bacteroids in different legumes correlates with the number and type of symbiotic NCR peptides. *Proc. Natl. Acad. Sci. USA* **114**: 5041–5046.
- Montiel, J., Nava, N., Cárdenas, L., Sánchez-López, R., Arthikala, M.-K., Santana, O., Sánchez, F., and Quinto, C. (2012). A *Phaseolus vulgaris* NADPH oxidase gene is required for root infection by Rhizobia. *Plant Cell Physiol.* **53**: 1751–1767.
- Moreau, S., Verdenaud, M., Ott, T., Letort, S., de Billy, F., Niebel, A., Gouzy, J., de Carvalho-Niebel, F., and Gamas, P. (2011). Transcription reprogramming during root nodule development in *Medicago truncatula*. *PLoS One* **6**: e16463.
- Morieri, G., Martinez, E.A., Jarynowski, A., Driguez, H., Morris, R., Oldroyd, G.E., and Downie, J.A. (2013). Host-specific Nod-factors associated with *Medicago truncatula* nodule infection differentially induce calcium influx and calcium spiking in root hairs. *New Phytol.* **200**: 656–662.
- Murphy, E., and De Smet, I. (2014). Understanding the RALF family: A tale of many species. *Trends Plant Sci.* **19**: 664–671.
- Nagy Mihály, M., Veluchamy, A., Györgypál, Z., Ariel, F., Jégu, T., Benhamed, M., Szűcs, A., Kereszt, A., Mergaert, P., and Kondorosi, É. (2017). Ploidy-dependent changes in the epigenome of symbiotic cells correlate with specific patterns of gene expression. *Proc. Natl. Acad. Sci. USA* **114**: 4543–4548.
- Nallu, S., Silverstein, K.A.T., Samac, D.A., Bucciarelli, B., Vance, C.P., and VandenBosch, K.A. (2013). Regulatory patterns of a large family of defensin-like genes expressed in nodules of *Medicago truncatula*. *PLoS One* **8**: e60355.
- Nallu, S., Silverstein, K.A.T., Zhou, P., Young, N.D., and Vandenbosch, K.A. (2014). Patterns of divergence of a large family of nodule cysteine-rich peptides in accessions of *Medicago truncatula*. *Plant J.* **78**: 697–705.
- Nelson, M.S., and Sadowsky, M.J. (2015). Secretion systems and signal exchange between nitrogen-fixing rhizobia and legumes. *Front. Plant Sci.* **6**: 491.
- Nouwen, N., Arrighi, J.F., Cartieaux, F., Chaintreuil, C., Gully, D., Klopp, C., and Giraud, E. (2017). The role of rhizobial (NifV) and plant (FEN1) homocitrate synthases in *Aeschynomene*/photosynthetic *Bradyrhizobium* symbiosis. *Sci. Rep.* **7**: 448.
- Nova-Franco, B., Íñiguez, L.P., Valdés-López, O., Alvarado-Affantranger, X., Leija, A., Fuentes, S.I., Ramírez, M., Paul, S., Reyes, J.L., Girard, L., and Hernández, G. (2015). The microRNA72c-APETALA2-1 node as a key regulator of the common bean-Rhizobium etli nitrogen fixation symbiosis. *Plant Physiol.* **168**: 273–291.
- O'Rourke, J.A., Iniguez, L.P., Fu, F., Bucciarelli, B., Miller, S.S., Jackson, S.A., McClean, P.E., Li, J., Dai, X., Zhao, P.X., Hernandez, G., and Vance, C.P. (2014). An RNA-Seq based gene expression atlas of the common bean. *BMC Genomics* **15**: 866.
- O'Rourke, J.A., et al. (2015). The *Medicago sativa* gene index 1.2: A web-accessible gene expression atlas for investigating expression differences between *Medicago sativa* subspecies. *BMC Genomics* **16**: 502.
- Oldroyd, G.E.D. (2013). Speak, friend, and enter: Signalling systems that promote beneficial symbiotic associations in plants. *Nat. Rev. Microbiol.* **11**: 252–263.
- Oldroyd, G.E.D., and Dixon, R. (2014). Biotechnological solutions to the nitrogen problem. *Curr. Opin. Biotechnol.* **26**: 19–24.
- Ott, T. (2017). Membrane nanodomains and microdomains in plant-microbe interactions. *Curr. Opin. Plant Biol.* **40**: 82–88.
- Ovchinnikova, E., et al. (2011). IPD3 controls the formation of nitrogen-fixing symbiosomes in pea and *Medicago* Spp. *Mol. Plant Microbe Interact.* **24**: 1333–1344.
- Paez Valencia, J., Goodman, K., and Otegui, M.S. (2016). Endocytosis and endosomal trafficking in plants. *Annu. Rev. Plant Biol.* **67**: 309–335.
- Pan, H., Oztas, O., Zhang, X., Wu, X., Stonoha, C., Wang, E., Wang, B., and Wang, D. (2016). A symbiotic SNARE protein generated by alternative termination of transcription. *Nat. Plants* **2**: 15197.
- Pan, H., Stonoha-Arther, C., and Wang, D. (2018). *Medicago* plants control nodulation by regulating proteolysis of the receptor-like kinase DMI2. *Plant Physiol.* **177**: 792–802.
- 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.
- Pecrix, Y., et al. (2018). Whole-genome landscape of *Medicago truncatula* symbiotic genes. *Nat. Plants* **4**: 1017–1025.
- Peláez, P., Trejo, M.S., Íñiguez, L.P., Estrada-Navarrete, G., Covarrubias, A.A., Reyes, J.L., and Sanchez, F. (2012). Identification and characterization of microRNAs in *Phaseolus vulgaris* by high-throughput sequencing. *BMC Genomics* **13**: 83.
- Peng, Z., Liu, F., Wang, L., Zhou, H., Paudel, D., Tan, L., Maku, J., Gallo, M., and Wang, J. (2017). Transcriptome profiles reveal gene regulation of peanut (*Arachis hypogaea* L.) nodulation. *Sci. Rep.* **7**: 40066.
- Persson, T., et al. (2015). *Candidatus* Frankia datiscaae Dg1, the actinobacterial microsymbiont of *Datisca glomerata*, expresses the canonical *nod* genes *nodABC* in symbiosis with its host plant. *PLoS One* **10**: e0127630.
- Pierre, O., Hopkins, J., Comber, M., Baldacci, F., Engler, G., Brouquisse, R., Hérouart, D., and Boncompagni, E. (2014). Involvement of papain and legumain proteinase in the senescence process of *Medicago truncatula* nodules. *New Phytol.* **202**: 849–863.
- Pislariu, C.I., Sinharoy, S., Torres-Jerez, I., Nakashima, J., Blancaflor, E.B., and Udvardi, M.K. (2019). The nodule-specific PLAT domain protein NPD1 is required for nitrogen-fixing symbiosis. *Plant Physiol.* **180**: 1480–1497.
- Poole, P., Ramachandran, V., and Terpolilli, J. (2018). Rhizobia: From saprophytes to endosymbionts. *Nat. Rev. Microbiol.* **16**: 291–303.
- Price, M.N., Deutschbauer, A.M., Skerker, J.M., Wetmore, K.M., Ruths, T., Mar, J.S., Kuehl, J.V., Shao, W., and Arkin, A.P. (2013). Indirect and suboptimal control of gene expression is widespread in bacteria. *Mol. Syst. Biol.* **9**: 660.
- Qiao, Z., Brechenmacher, L., Smith, B., Strout, G.W., Mangin, W., Taylor, C., Russell, S.D., Stacey, G., and Libault, M. (2017). The GmFWL1 (FW2-2-like) nodulation gene encodes a plasma membrane microdomain-associated protein. *Plant Cell Environ.* **40**: 1442–1455.
- Ren, B., Wang, X., Duan, J., and Ma, J. (2019). Rhizobial tRNA-derived small RNAs are signal molecules regulating plant nodulation. *Science* **365**: 919–922.
- Reyero-Saavedra, M.D.R., Qiao, Z., Sánchez-Correa, M.D.S., Díaz-Pineda, M.E., Reyes, J.L., Covarrubias, A.A., Libault, M., and Valdés-López, O. (2017). Gene silencing of *Argonaute5* negatively

- affects the establishment of the legume-rhizobia symbiosis. *Genes* (Basel) **8**: 352.
- Richter, H.E., Sandal, N.N., Marcker, K.A., and Sengupta-Gopalan, C.** (1991). Characterization and genomic organization of a highly expressed late nodulin gene subfamily in soybeans. *Mol. Gen. Genet.* **229**: 445–452.
- Rodrigues, S.G., Stickels, R.R., Goeva, A., Martin, C.A., Murray, E., Vanderburg, C.R., Welch, J., Chen, L.M., Chen, F., and Macosko, E.Z.** (2019). Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**: 1463–1467.
- Roux, B., 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.
- Roy, S., Liu, W., Nandety, R.S., Crook, A.D., Mysore, K.S., Pislariu, C.I., Frugoli, J.A., Dickstein, R., and Udvardi, M.K.** (2019). Celebrating 20 years of genetic discoveries in legume nodulation and symbiotic nitrogen fixation. *Plant Cell tpc.00279.2019*.
- Ruffel, S., et al.** (2008). Systemic signaling of the plant nitrogen status triggers specific transcriptome responses depending on the nitrogen source in *Medicago truncatula*. *Plant Physiol.* **146**: 2020–2035.
- Salgado, M.G., van Velzen, R., Nguyen, T.V., Battenberg, K., Berry, A.M., Lundin, D., and Pawlowski, K.** (2018). Comparative analysis of the nodule transcriptomes of *Ceanothus thyrsiflorus* (Rhamnaceae, Rosales) and *Datisca glomerata* (Datiscaceae, Cucurbitales). *Front. Plant Sci.* **9**: 1629.
- Sandal, N.N., Bojsen, K., and Marcker, K.A.** (1987). A small family of nodule specific genes from soybean. *Nucleic Acids Res.* **15**: 1507–1519.
- Satgé, C., Moreau, S., Sallet, E., Lefort, G., Auriac, M.-C., Remblière, C., Cottret, L., Gallardo, K., Noirot, C., Jardinaud, M.-F., and Gamas, P.** (2016). Reprogramming of DNA methylation is critical for nodule development in *Medicago truncatula*. *Nat. Plants* **2**: 16166.
- Schauser, L., Roussis, A., Stiller, J., and Stougaard, J.** (1999). A plant regulator controlling development of symbiotic root nodules. *Nature* **402**: 191–195.
- Schiessl, K., et al.** (2019). *NODULE INCEPTION* recruits the lateral root developmental program for symbiotic nodule organogenesis in *Medicago truncatula*. *Curr. Biol.* **29**: 3657–3668.e5.
- Seabra, A.R., Pereira, P.A., Becker, J.D., and Carvalho, H.G.** (2012). Inhibition of glutamine synthetase by phosphinothricin leads to transcriptome reprogramming in root nodules of *Medicago truncatula*. *Mol. Plant Microbe Interact.* **25**: 976–992.
- Senovilla, M., Castro-Rodríguez, R., Abreu, I., Escudero, V., Kryvoruchko, I., Udvardi, M.K., Imperial, J., and González-Guerrero, M.** (2018). *Medicago truncatula* copper transporter 1 (MtCOPT1) delivers copper for symbiotic nitrogen fixation. *New Phytol.* **218**: 696–709.
- Severin, A.J., et al.** (2010). RNA-Seq Atlas of Glycine max: A guide to the soybean transcriptome. *BMC Plant Biol.* **10**: 160.
- Shubin, N., Tabin, C., and Carroll, S.** (2009). Deep homology and the origins of evolutionary novelty. *Nature* **457**: 818–823.
- Sijacic, P., Wang, W., and Liu, Z.** (2011). Recessive antimorphic alleles overcome functionally redundant loci to reveal TSO1 function in *Arabidopsis* flowers and meristems. *PLoS Genet.* **7**: e1002352.
- Singh, S., Katzer, K., Lambert, J., Cerri, M., and Parniske, M.** (2014). CYCLOPS, a DNA-binding transcriptional activator, orchestrates symbiotic root nodule development. *Cell Host Microbe* **15**: 139–152.
- Sinharoy, S., Torres-Jerez, I., Bandyopadhyay, K., Kereszt, A., Pislariu, C.I., Nakashima, J., Benedito, V.A., Kondorosi, E., and Udvardi, M.K.** (2013). The C2H2 transcription factor regulator of symbiosome differentiation represses transcription of the secretory pathway gene VAMP721a and promotes symbiosome development in *Medicago truncatula*. *Plant Cell* **25**: 3584–3601.
- Smit, P., Raedts, J., Portyanko, V., Debellé, F., Gough, C., Bisseling, T., and Geurts, R.** (2005). NSP1 of the GRAS protein family is essential for rhizobial Nod factor-induced transcription. *Science* **308**: 1789–1791.
- Soltis, D.E., Soltis, P.S., Morgan, D.R., Swensen, S.M., Mullin, B.C., Dowd, J.M., and Martin, P.G.** (1995). Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. *Proc. Natl. Acad. Sci. USA* **92**: 2647–2651.
- Soyano, T., and Hayashi, M.** (2014). Transcriptional networks leading to symbiotic nodule organogenesis. *Curr. Opin. Plant Biol.* **20**: 146–154.
- 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.
- Stähl, P.L., et al.** (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**: 78–82.
- Starker, C.G., Parra-Colmenares, A.L., Smith, L., Mitra, R.M., and Long, S.R.** (2006). Nitrogen fixation mutants of *Medicago truncatula* fail to support plant and bacterial symbiotic gene expression. *Plant Physiol.* **140**: 671–680.
- Stokstad, E.** (2016). The nitrogen fix. *Science* **353**: 1225–1227.
- Stonoha-Arther, C., and Wang, D.** (2018). Tough love: Accommodating intracellular bacteria through directed secretion of antimicrobial peptides during the nitrogen-fixing symbiosis. *Curr. Opin. Plant Biol.* **44**: 155–163.
- Stougaard, J., Sandal, N.N., Grøn, A., Kühle, A., and Marcker, K.A.** (1987). 5' Analysis of the soybean leghaemoglobin lbc(3) gene: Regulatory elements required for promoter activity and organ specificity. *EMBO J.* **6**: 3565–3569.
- 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.
- Subramanian, S., Fu, Y., Sunkar, R., Barbazuk, W.B., Zhu, J.K., and Yu, O.** (2008). Novel and nodulation-regulated microRNAs in soybean roots. *BMC Genomics* **9**: 160.
- Sugawara, M., et al.** (2013). Comparative genomics of the core and accessory genomes of 48 *Sinorhizobium* strains comprising five genospecies. *Genome Biol.* **14**: R17.
- Sugiyama, A., Saida, Y., Yoshimizu, M., Takanashi, K., Sosso, D., Frommer, W.B., and Yazaki, K.** (2017). Molecular characterization of LjSWEET3, a sugar transporter in nodules of *Lotus japonicus*. *Plant Cell Physiol.* **58**: 298–306.
- Sullivan, D., Brisson, N., Goodchild, B., and Verma, D.P.S.** (1981). Molecular cloning and organization of two leghaemoglobin genomic sequences of soybean. *Nature* **289**: 516–518.
- Suzaki, T., Ito, M., Yoro, E., Sato, S., Hirakawa, H., Takeda, N., and Kawaguchi, M.** (2014). Endoreduplication-mediated initiation of symbiotic organ development in *Lotus japonicus*. *Development* **141**: 2441–2445.
- Svistoonoff, 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.
- Szabados, L., Ratet, P., Grunenberg, B., and de Bruijn, F.J.** (1990). Functional analysis of the *Sesbania rostrata* leghemoglobin *glb3* gene 5'-upstream region in transgenic *Lotus corniculatus* and *Nicotiana tabacum* plants. *Plant Cell* **2**: 973–986.
- Szczygłowski, K., Szabados, L., Fujimoto, S.Y., Silver, D., and de Bruijn, F.J.** (1994). Site-specific mutagenesis of the nodule-infected

- cell expression (NICE) element and the AT-rich element ATRE-BS2* of the *Sesbania rostrata* leghemoglobin *glb3* promoter. *Plant Cell* **6**: 317–332.
- Takanashi, K., Takahashi, H., Sakurai, N., Sugiyama, A., Suzuki, H., Shibata, D., Nakazono, M., and Yazaki, K.** (2012). Tissue-specific transcriptome analysis in nodules of *Lotus japonicus*. *Mol. Plant Microbe Interact.* **25**: 869–876.
- 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.
- Tautz, D., and Domazet-Lošo, T.** (2011). The evolutionary origin of orphan genes. *Nat. Rev. Genet.* **12**: 692–702.
- Temple, S.J., Vance, C.P., and Gantt, J.S.** (1998). Glutamate synthase and nitrogen assimilation. *Trends Plant Sci.* **3**: 51–56.
- Tesfaye, M., et al.** (2013). Spatio-temporal expression patterns of *Arabidopsis thaliana* and *Medicago truncatula* defensin-like genes. *PLoS One* **8**: e58992.
- Teulet, A., et al.** (2019). The rhizobial type III effector ErnA confers the ability to form nodules in legumes. *Proc. Natl. Acad. Sci. USA* **116**: 21758–21768.
- 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.
- Trujillo, D.I., Silverstein, K.A., and Young, N.D.** (2014). Genomic characterization of the LEED.PEEDs, a gene family unique to the *medicago* lineage. *G3 (Bethesda)* **4**: 2003–2012.
- Trujillo, D.I., Silverstein, K.A.T., and Young, N.D.** (2019). Nodule-specific PLAT domain proteins are expanded in the *Medicago* lineage and required for nodulation. *New Phytol.* **222**: 1538–1550.
- Tsikou, D., Yan, Z., Holt, D.B., Abel, N.B., Reid, D.E., Madsen, L.H., Bhasin, H., Sexauer, M., Stougaard, J., and Markmann, K.** (2018). Systemic control of legume susceptibility to rhizobial infection by a mobile microRNA. *Science* **362**: 233–236.
- Udvardi, M., and Poole, P.S.** (2013). Transport and metabolism in legume-rhizobia symbioses. *Annu. Rev. Plant Biol.* **64**: 781–805.
- van Brussel, A.A.N., Bakhuizen, R., van Spronsen, P.C., Spaink, H.P., Tak, T., Lugtenberg, B.J.J., and Kijne, J.W.** (1992). Induction of pre-infection thread structures in the leguminous host plant by mitogenic lipo-oligosaccharides of *Rhizobium*. *Science* **257**: 70–72.
- Van de Velde, W., Guerra, J.C., De Keyser, A., De Rycke, R., Rombauts, S., Maunoury, N., Mergaert, P., Kondorosi, E., Holsters, M., and Goormachtig, S.** (2006). Aging in legume symbiosis. A molecular view on nodule senescence in *Medicago truncatula*. *Plant Physiol.* **141**: 711–720.
- 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.
- van Velzen, R., et al.** (2018). Comparative genomics of the nonlegume *Parasponia* reveals insights into evolution of nitrogen-fixing rhizobia symbioses. *Proc. Natl. Acad. Sci. USA* **115**: E4700–E4709.
- van Zeijl, A., Op den Camp, R.H.M., Deinum, E.E., Charnikhova, T., Franssen, H., Op den Camp, H.J.M., Bouwmeester, H., Kohlen, W., Bisseling, T., and Geurts, R.** (2015). *Rhizobium* lipochitooligosaccharide signaling triggers accumulation of cytokinins in *Medicago truncatula* roots. *Mol. Plant* **8**: 1213–1226.
- van Zeijl, A., Wardhani, T.A.K., Seifi Kalhor, M., Rutten, L., Bu, F., Hartog, M., Linders, S., Fedorova, E.E., Bisseling, T., Kohlen, W., and Geurts, R.** (2018). CRISPR/Cas9-mediated mutagenesis of four putative symbiosis genes of the tropical tree *Parasponia andersonii* reveals novel phenotypes. *Front. Plant Sci.* **9**: 284.
- Verdier, J., et al.** (2013). A regulatory network-based approach dissects late maturation processes related to the acquisition of desiccation tolerance and longevity of *Medicago truncatula* seeds. *Plant Physiol.* **163**: 757–774.
- Vernié, T., Kim, J., Frances, L., Ding, Y., Sun, J., Guan, D., Niebel, A., Gifford, M.L., de Carvalho-Niebel, F., and Oldroyd, G.E.D.** (2015). The NIN transcription factor coordinates diverse nodulation programs in different tissues of the *Medicago truncatula* root. *Plant Cell* **27**: 3410–3424.
- Vernié, T., Moreau, S., de Billy, F., Plet, J., Combier, J.-P., Rogers, C., Oldroyd, G., Frugier, F., Niebel, A., and Gamas, P.** (2008). EFD is an ERF transcription factor involved in the control of nodule number and differentiation in *Medicago truncatula*. *Plant Cell* **20**: 2696–2713.
- Vinardell, J.M., Fedorova, E., Cebolla, A., Kevei, Z., Horváth, G., Kelemen, Z., Tarayre, S., Roudier, F., Mergaert, P., Kondorosi, A., and Kondorosi, E.** (2003). Endoreduplication mediated by the anaphase-promoting complex activator CCS52A is required for symbiotic cell differentiation in *Medicago truncatula* nodules. *Plant Cell* **15**: 2093–2105.
- Wang, C., et al.** (2016). *NODULES WITH ACTIVATED DEFENSE 1* is required for maintenance of rhizobial endosymbiosis in *Medicago truncatula*. *New Phytol.* **212**: 176–191.
- Wang, D., Griffiths, J., Starker, C., Fedorova, E., Limpens, E., Ivanov, S., Bisseling, T., and Long, S.** (2010). A nodule-specific protein secretory pathway required for nitrogen-fixing symbiosis. *Science* **327**: 1126–1129.
- Wang, H., Moore, M.J., Soltis, P.S., Bell, C.D., Brockington, S.F., Alexandre, R., Davis, C.C., Latvis, M., Manchester, S.R., and Soltis, D.E.** (2009). Rosid radiation and the rapid rise of angiosperm-dominated forests. *Proc. Natl. Acad. Sci. USA* **106**: 3853–3858.
- Wang, L., Sun, Z., Su, C., Wang, Y., Yan, Q., Chen, J., Ott, T., and Li, X.** (2019). A GmNINa-miR172c-NNC1 regulatory network coordinates the nodulation and autoregulation of nodulation pathways in soybean. *Mol. Plant* **12**: 1211–1226.
- Wang, Y., Li, K., Chen, L., Zou, Y., Liu, H., Tian, Y., Li, D., Wang, R., Zhao, F., Ferguson, B.J., Gresshoff, P.M., and Li, X.** (2015). MicroRNA167-directed regulation of the auxin response factors GmARF8a and GmARF8b is required for soybean nodulation and lateral root development. *Plant Physiol.* **168**: 984–999.
- Wang, Y., Wang, L., Zou, Y., Chen, L., Cai, Z., Zhang, S., Zhao, F., Tian, Y., Jiang, Q., Ferguson, B.J., Gresshoff, P.M., and Li, X.** (2014). Soybean miR172c targets the repressive AP2 transcription factor NNC1 to activate *ENOD40* expression and regulate nodule initiation. *Plant Cell* **26**: 4782–4801.
- Werner, G.D.A., Cornwell, W.K., Sprent, J.I., Kattge, J., and Kiers, E.T.** (2014). A single evolutionary innovation drives the deep evolution of symbiotic N₂-fixation in angiosperms. *Nat. Commun.* **5**: 4087.
- Wildermuth, M.C., Steinwand, M.A., McRae, A.G., Jaenisch, J., and Chandran, D.** (2017). Adapted biotroph manipulation of plant cell ploidy. *Annu. Rev. Phytopathol.* **55**: 537–564.
- Wu, Z., Wang, M., Yang, S., Chen, S., Chen, X., Liu, C., Wang, S., Wang, H., Zhang, B., Liu, H., Qin, R., and Wang, X.** (2019). A global coexpression network of soybean genes gives insights into the evolution of nodulation in nonlegumes and legumes. *New Phytol.* **223**: 2104–2119.
- Xiao, T.T., Schilderink, S., Moling, S., Deinum, E.E., Kondorosi, E., Franssen, H., Kulikova, O., Niebel, A., and Bisseling, T.** (2014). Fate map of *Medicago truncatula* root nodules. *Development* **141**: 3517–3528.

- Yan, Z., Hossain, M.S., Arikiti, S., Valdés-López, O., Zhai, J., Wang, J., Libault, M., Ji, T., Qiu, L., Meyers, B.C., and Stacey, G.** (2015). Identification of microRNAs and their mRNA targets during soybean nodule development: Functional analysis of the role of miR393j-3p in soybean nodulation. *New Phytol.* **207**: 748–759.
- Yamaya-Ito, H., Shimoda, Y., Hakoyama, T., Sato, S., Kaneko, T., Hossain, M.S., Shibata, S., Kawaguchi, M., Hayashi, M., Kouchi, H., and Umehara, Y.** (2018). Loss-of-function of ASPARTIC PEPTIDASE NODULE-INDUCED 1 (APN1) in *Lotus japonicus* restricts efficient nitrogen-fixing symbiosis with specific *Mesorhizobium loti* strains. *Plant J.* **93**: 5–16.
- Yan, Z., et al.** (2016). Identification and functional characterization of soybean root hair microRNAs expressed in response to *Bradyrhizobium japonicum* infection. *Plant Biotechnol. J.* **14**: 332–341.
- Yang, W.-C., de Blank, C., Meskiene, I., Hirt, H., Bakker, J., van Kammen, A., Franssen, H., and Bisseling, T.** (1994). *Rhizobium* nod factors reactivate the cell cycle during infection and nodule primordium formation, but the cycle is only completed in primordium formation. *Plant Cell* **6**: 1415–1426.
- Yang, Z.** (2008). Cell polarity signaling in Arabidopsis. *Annu. Rev. Cell Dev. Biol.* **24**: 551–575.
- Yano, K., et al.** (2008). CYCLOPS, a mediator of symbiotic intracellular accommodation. *Proc. Natl. Acad. Sci. USA* **105**: 20540–20545.
- Young, N.D., et al.** (2011). The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature* **480**: 520–524.
- Yu, H., Xiao, A., Dong, R., Fan, Y., Zhang, X., Liu, C., Wang, C., Zhu, H., Duanmu, D., Cao, Y., and Zhang, Z.** (2018). Suppression of innate immunity mediated by the CDPK-Rboh complex is required for rhizobial colonization in *Medicago truncatula* nodules. *New Phytol.* **220**: 425–434.
- Yuan, S.L., et al.** (2017). RNA-Seq analysis of nodule development at five different developmental stages of soybean (*Glycine max*) inoculated with *Bradyrhizobium japonicum* strain 113-2. *Sci. Rep.* **7**: 42248.
- Zhou, P., et al.** (2017). Exploring structural variation and gene family architecture with *De Novo* assemblies of 15 *Medicago* genomes. *BMC Genomics* **18**: 261.