



## MtEFD and MtEFD2: Two transcription factors with distinct neofunctionalization in symbiotic nodule development

Marie-Françoise Jardinaud, Justine Fromentin, Marie-Christine Auriac, Sandra Moreau, Yann Pecrix, Ludivine Taconnat, Ludovic Cottret, Grégoire Aubert, Sandrine Balzergue, Judith Burstin, et al.

### ► To cite this version:

Marie-Françoise Jardinaud, Justine Fromentin, Marie-Christine Auriac, Sandra Moreau, Yann Pecrix, et al.. MtEFD and MtEFD2: Two transcription factors with distinct neofunctionalization in symbiotic nodule development. *Plant Physiology*, 2022, 189 (3), pp.1587 - 1607. 10.1093/plphys/kiac177 . hal-03728310

**HAL Id: hal-03728310**

**<https://hal.science/hal-03728310>**

Submitted on 27 Jul 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

Short title: *MtEFD* and *MtEFD2* in nodule development

Corresponding Author: Pascal Gamas; [Pascal.Gamas@inrae.fr](mailto:Pascal.Gamas@inrae.fr)

*MtEFD* and *MtEFD2*: two transcription factors with distinct neofunctionalization in symbiotic nodule development

Marie-Françoise Jardinaud<sup>\*1</sup>, Justine Fromentin<sup>\*1</sup>, Marie-Christine Auriac<sup>1</sup>, Sandra Moreau<sup>1</sup>, Yann Pecrix<sup>1</sup>, Ludivine Taconnat<sup>2,3</sup>, Ludovic Cottret<sup>1</sup>, Grégoire Aubert<sup>4</sup>, Sandrine Balzergue<sup>2,3</sup>, Judith Burstin<sup>4</sup>, Sébastien Carrère<sup>1</sup>, Pascal Gamas<sup>1</sup>

\*: co first authors

1. LIPME, Université de Toulouse, INRAE, CNRS, Castanet-Tolosan, France

2. Université Paris-Saclay, CNRS, INRAE, Université Evry, Institute of Plant Sciences Paris-Saclay (IPS2), 91405, Orsay, France

3. Université de Paris, CNRS, INRAE, Institute of Plant Sciences Paris-Saclay (IPS2), 91405, Orsay, France

4. Agroécologie, AgroSup Dijon, INRAE, Université Bourgogne Franche-Comté, Dijon, France

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors

(<https://academic.oup.com/plphys/pages/General-Instructions>) is Marie-Françoise Jardinaud.

One sentence summary

The nitrogen-fixing symbiosis has recruited two paralogous root transcription factors for nodule development, through evolution of their promoter and coding sequence, respectively. Footnotes:

List of author contributions:

M-F.J., J.F. and P.G. conceived the research plans; M-F.J. and J.F. performed most of the experiments in Mt; G.A. performed experiments in pea; Y.P. designed the CRISPR-Cas9 constructs; L.T., S.C., and L.C. performed microarray experiments, bioinformatics analyses, and phylogenomic analyses, respectively; S.M. and M-C.A. provided technical assistance to M-F.J. and P.G.; S.B. supervised microarray experiments; P.G. and J.B. supervised the experiments in Mt and pea, respectively; M-F.J., J.F. and P.G. analyzed the data; P.G. conceived the project and wrote the article with contributions of

M-F.J., J.F., S.B., L.C.; P.G. agrees to serve as the author responsible for contact and ensures communication.

Present addresses:

Sandrine Balzergue: IRHS, INRAE, AGROCAMPUS-Ouest, Université d'Angers, France

Yann Pecrix: PVBMT, CIRAD, Université de la Réunion, France

Ludivine Taconnat: GEVES, SNES, Beaucouzé, France

## Abstract

Rhizobium-legume nitrogen-fixing symbiosis involves the formation of a specific organ, the root nodule, which provides bacteria with the proper cellular environment for atmospheric nitrogen fixation. Coordinated differentiation of plant and bacterial cells is an essential step of nodule development, for which few transcriptional regulators have been characterized. *MtEFD* (*Medicago truncatula* ETHYLENE RESPONSE FACTOR REQUIRED FOR NODULE DIFFERENTIATION) encodes an AP2/ERF (APETALA2/ETHYLENE RESPONSIVE FACTOR) transcription factor, the mutation of which leads to both hypernodulation and severe defects in nodule development. *MtEFD* positively controls a negative regulator of cytokinin signaling, the *MtRR4* (RESPONSE REGULATOR 4) gene. Here we showed that the *Mtefd-1* mutation affects both plant and bacterial endoreduplication in nodules, as well as the expression of hundreds of genes in young and mature nodules, upstream of known regulators of symbiotic differentiation. *MtRR4* expressed with the *MtEFD* promoter complemented *Mtefd-1* hypernodulation but not the nodule differentiation phenotype. Unexpectedly, a non-legume homologue of *MtEFD*, AtERF003 in *Arabidopsis* (*Arabidopsis thaliana*), could efficiently complement both phenotypes of *Mtefd-1*, in contrast to the *MtEFD* paralogue *MtEFD2* expressed in the root and nodule meristematic zone. A domain swap experiment showed that *MtEFD2* differs from *MtEFD* by its C-terminal fraction outside the DNA binding domain. Furthermore, CRISPR-Cas9 mutagenesis of *MtEFD2* led to a reduction in the number of nodules formed in *Mtefd-1*, with down-regulation of a set of genes, including notably *MtNF-YA1* (NUCLEAR FACTOR-YA1) and *MtNF-YB16*, which are essential for nodule meristem establishment. We therefore conclude that nitrogen-fixing symbiosis recruited two proteins originally expressed in roots, *MtEFD* and *MtEFD2*, with distinct functions and neofunctionalization processes for each of them.

## Introduction

The rhizobium-legume symbiosis provides agro-ecosystems with considerable amounts of assimilable nitrogen (N), a critical factor for plant growth, thanks to the capacity of the bacterial nitrogenase to fix atmospheric nitrogen. The symbiotic N fixation (SNF) takes place in dedicated root organs, called nodules, that provide rhizobia with carbon sources and a micro-oxic environment, essential for the nitrogenase to be functional. Nodules can be of indeterminate (e.g., in pea (*Pisum sativum*) and *Medicago* sp.) or determinate (e.g., in soybean (*Glycine max*) and *Lotus* sp.) type, i.e., with or without a permanent (apical) meristem, respectively. Two main steps can be distinguished in nodule organogenesis: the first one is nodule inception in root tissues, which follows a signaling process triggered by the perception by the plant of specific rhizobial chito-oligosaccharidic molecules, the Nod factors. This leads to cell divisions in the root cortex and rhizobial infections in the root epidermis. Both processes involve a series of symbiotic transcription factors (TFs), with NODULE INCEPTION (NIN) as a master regulator (Schauser et al., 1999; Marsh et al., 2007), as well as phytohormones, among which cytokinins (CK) and auxins are major players (Lin et al., 2020). CK signaling in the root cortex leads to *NIN* induction, itself activating NF-Y and LATERAL ORGAN BOUNDARIES-DOMAIN PROTEIN 16 (LBD16) TF expression, contributing to local auxin accumulation and consequently cell division (Schiessl et al., 2019; Soyano et al., 2019). LBD16 being also involved in lateral root organogenesis, this reveals an important overlap between the genetic programs respectively leading to nodule and lateral root inception, also evidenced by the importance of the SHORTROOT-SCARECROW module in both developmental programs (Dong et al., 2021). The second step is cell differentiation associated with nodule development, which allows symbiotic traits to be acquired by plant and bacterial cells. At this stage, a large number of genes are strongly and specifically up-regulated in developing nodules compared to the roots (Mergaert et al., 2020). In *Medicago truncatula*, the symbiotic differentiation process first takes place in cells generated by divisions of inner (C4-C5) root cortical cells in the nodule primordium, and later in cells produced by an apical nodule meristem established from the C3 cortical cell layer (Xiao et al., 2014).

The meristematic region of a *M. truncatula* nodule, termed Zone I (ZI), involves distinct division centers, associated with peripheral vascular bundles (VB) and central tissues respectively. The nodule vasculature is ontologically related to roots, based on the ontogeny of the VB meristems and the expression of key regulators of the root meristem [WUSCHEL-RELATED HOMEODOMAIN (MtWOX5) and PLETHORA (PLT1 and PLT2) TFs] (Osipova et al., 2012; Franssen et al., 2015; Magne et al., 2018). Two genes, with distinct but overlapping expression profiles in the nodule meristem, are essential to regulate the meristematic subdomains and control nodule identity: *MtNODULE ROOT 1* (*MtNOOT1*)

and *MtNOOT2* (Couzigou et al., 2012; Magne et al., 2018). A *Mtnoot1 Mtnoot2* double mutant shows a complete nodule to root conversion, with ectopic roots developing from VB meristems.

The symbiotic differentiation of indeterminate nodules involves cells generated by the central nodule meristem, along a longitudinal gradient from apical to proximal nodule zones. Once meristematic cells exit the division cycles in the sub-meristematic zone (termed distal ZII), they rapidly enlarge and begin to endoreduplicate. This endoreduplication is controlled by *MtCCS52A*, a cell cycle switch gene activated at the final stage of nodule primordium formation (Vinardell et al., 2003). Knocked down lines of *MtCCS52A* exhibit a reduction of the endoreduplication level (predominantly 8C, vs 32C nuclei in control nodules) and produce small fix- nodules, thereby demonstrating the importance of endoreduplication for nodule development. *MtCCS52A* is expressed in the whole nodule ZII (or infection zone) (Vinardell et al., 2003; Roux et al., 2014), where rhizobia are released from infection threads (tubular structures of plant origin). Rhizobia are surrounded by a plant membrane showing specific identity markers (Limpens et al., 2009), thereby forming the so-called symbiosome. In the same time, hundreds of plant genes are massively up-regulated in successive waves (Maunoury et al., 2010; Roux et al., 2014; Pecrix et al., 2018). This allows the cell environment required for SNF to be gradually set up, leading to the formation of the interzone II-III (IZ) and nitrogen-fixing ZIII.

A hallmark of nodule differentiation is the accumulation of leghemoglobin, essential to keep a low free oxygen concentration while maintaining respiration. Differentiation also involves the upregulation of various transporters and important modifications of the cytoskeleton organization and membrane trafficking in the plant cell [e.g. (Gavrin et al., 2014; Zhang et al., 2019)]. A C<sub>2</sub>H<sub>2</sub> TF termed RSD (REGULATOR OF SYMBIOTIC DIFFERENTIATION) is thought to promote symbiosome development by regulating the plant secretory pathway (Sinharoy et al., 2013). Furthermore, in *Medicago* and other legumes of the Inverted Repeat Lacking Clade, a gene family strongly up-regulated in the infection / differentiation zones encodes so-called Nodule-specific Cysteine-Rich peptides (NCRs) (Pan and Wang, 2017; Mergaert, 2018; Stonoha-Arther and Wang, 2018), processed and targeted to the symbiosomes by a signal peptidase complex (SPC) including DOES NOT FIX NITROGEN 1 (DNF1) (Wang et al., 2010). More than 600 NCR genes have been identified in *M. truncatula* (de Bang et al., 2017; Montiel et al., 2017). While the role of most of them remains elusive, several have been demonstrated to be essential factors of bacteroid differentiation and / or survival (Van de Velde et al., 2010; Farkas et al., 2014; Horváth et al., 2015; Kim et al., 2015). Interestingly, the DNF1 SPC was recently shown to process NIN, to produce a smaller NIN form that controls late nodule development and numerous associated genes (Feng et al., 2021).

As with plant cells, rhizobia undergo several rounds of endoreduplication, along with strong morphological changes and modifications in gene expression triggered by the micro-oxic environment and NCR peptides, resulting in terminally differentiated bacteroids. Importantly, in spite of their presence in huge numbers within plant cells, rhizobia do not trigger plant immune responses, unless the interaction is impaired by mutations. Thus, several fix- *M. truncatula* mutants exhibit defense reactions which affect the maintenance of rhizobia: *dnf2*, *nad1* (*nodules with activated defense 1*), *symcrk* (*symbiotic cys-rich receptor kinase*), *npd1* (*nodule-specific Plat-domain 1*) or *npd1*, 2, 3, 4, 5 (Bourcy et al., 2013; Berrabah et al., 2015; Wang et al., 2016; Yu et al., 2018; Pislariu et al., 2019).

Not many transcriptional regulators controlling symbiotic differentiation *per se* have been demonstrated in *M. truncatula*, besides MtRSD and MtNIN. MtNF-YA1, MtNOOT1 and MtNOOT2 are essential for the establishment and the maintenance of the central nodule meristem respectively, but their mutation does not prevent the symbiotic differentiation (Couzigou et al., 2012; Xiao et al., 2014; Magne et al., 2018). Not considering TFs expressed in nodule VBs, three TFs appear to be more directly involved in nodule differentiation / development. MtEFD (ETHYLENE RESPONSE FACTOR INVOLVED IN NODULE DIFFERENTIATION) is expressed in ZII (Vernié et al., 2008), under the direct or indirect control of INTERACTING PROTEIN OF DMI3 (IPD3) transcriptional regulator (Ovchinnikova et al., 2011). The two others are NIN (Feng et al., 2021; Liu et al., 2021) and MtNLP2, a NIN-like protein expressed in the IZ and the ZIII, where it notably controls the expression of leghemoglobins, together with NIN (Jiang et al., 2021). Furthermore, epigenetic regulators are also important, likely by modulating the access of TFs to chromatin regions. Thus a knocked-down mutant of DEMETER (DME), a DNA demethylase strongly up-regulated in proximal ZII and IZ, is fix- and impaired in nodule differentiation (Satgé et al., 2016). In addition, gene expression in the nodule differentiation zone often coincides with reduced levels of repressive histone marks (e.g., H3K27me3) and increased levels of active histone marks (H3K9ac) (Nagymihaly et al., 2017; Pecrix et al., 2018).

MtEFD belongs to the large family of AP2/ERF TFs, more specifically the group V, as MtERN1 (ERF REQUIRED FOR NODULATION 1) and MtERN2, two key regulators of NF signaling, nodule inception and rhizobium infection (Andrianakaja et al., 2007; Middleton et al., 2007; Cerri et al., 2016). MtEFD, however, differs from MtERN1 and MtERN2, both in its AP2/ERF DNA binding domain and in its expression profile, since *MtEFD* is neither induced by NF nor associated with infection threads, in contrast to *MtERN1* and *MtERN2* (Vernié et al., 2008). *Mtefd-1*, a deletion mutant of *MtEFD*, is fix- and exhibits strong alterations in the nodule ZII and III (Vernié et al., 2008). In addition, *Mtefd-1* exhibits a hypernodulation phenotype and frequent multilobed nodules. Transcriptomic analyses supported by a transactivation assay indicated that MtEFD positively controls a negative regulator of

the CK pathway, the type A response regulator *MtRR4* gene. Because CKs play a key role in nodule inception, it was hypothesized that the downregulation of *MtRR4* might explain the hypernodulation phenotype of *Mtefd-1*. It was also speculated that MtEFD might regulate a gradient of CK activity within the nodule and thereby impact nodule differentiation (Vernié et al., 2008). The importance of CKs for nodule development / differentiation has later been established by the nodule phenotype of a *Mtcre1/chk1* CK receptor mutant, which exhibits multiple lobes and incomplete differentiation (Plet et al., 2011). A multilobed nodule phenotype corresponding with a decreased *MtEFD* and *MtRR4* expression was also observed in *M. truncatula* roots where three KNOX homeodomain TF genes, termed *MtKNAT3/4/5-like* (or *MtKNOX3*, *MtKNOX5* and *MtKNOX9*), were simultaneously knocked-down (Di Giacomo et al., 2017). These TFs, expressed in nodule primordia and different nodule zones, belong to a KNOX homeodomain subclass promoting the differentiation of aerial organs in *Arabidopsis* (*Arabidopsis thaliana*).

Here we revisited the impact of the *Mtefd-1* mutation on nodule development, with analyses of the endoreduplication levels and affected genes, notably *MtRR4*. We also investigated *MtEFD* evolution, by testing the capacity of a close *MtEFD* paralog, *MtEFD2*, and of the *A. thaliana* *EFD* orthologue, AtERF003, to complement the *Mtefd-1* mutation. Finally, we assessed whether *MtEFD* and *MtEFD2*, which exhibit distinct but overlapping expression profiles, could both be involved in the control of nodulation.

## Results

### The *Mtefd-1* mutant exhibits an endoreduplication phenotype

To further investigate the impact of the *Mtefd-1* mutation on nodule differentiation, we analyzed nodule endoreduplication. In wild type (wt) 13-day-old nodules, the DNA content of plant cells ranged from 2C to 64C (five cycles of endoreduplication), with 32C and 64C nuclei representing on average 9.4% and 0.35% of total nuclei respectively (in attapulgite growth condition, supplemental Table S1). In *Mtefd-1* nodules, the relative abundance of 32C and 64C nuclei was strongly decreased (0.8% and 0.08% of the population on average, respectively), whereas the frequency of 16C nuclei was not statistically different from wt (3.5% vs. 4.2%) (Fig1. A; supplemental Table S1). This was observed in different growth conditions (aeroponic or attapulgite condition) and at different time points, from 13 to 30 days post inoculation (supplemental Table S1), therefore ruling out a mere delay of endoreduplication. This phenotype differed from the one observed with five other fix- *M. truncatula* mutants, namely TR3, TR36, TRV36, TR183 and TRV43 (Maunoury et al., 2010), for which corresponding genes have not yet been identified, except for TR3, allelic to *ipd3* (Ovchinnikova et al.,



2011). These five mutants showed either a low proportion (TRV36 and TRV43) or a total absence (TR3, TR183 and TR36) of both 16C and 32C cells (Fig. 1 C-F; supplemental Table S1), as previously reported (Maunoury et al., 2010). While the *rsd-1* (Sinharoy et al., 2013) and *dme* (Satsgé et al., 2016) fix- mutants are also affected in the level of both 16C and 32C cells, *Mtefd-1* appears to be a plant fix- mutant specifically affected at the 16C-32C transition.

The endoreduplication of bacterial cells (*Sinorhizobium meliloti* 2011) was also clearly reduced in *Mtefd-1* nodules (Fig. 1B). In addition, another hallmark of bacteroid differentiation, the expression of NifH, a nitrogenase subunit, was severely affected, as seen with a *NifH:GFP* fusion. While a strong *NifH:GFP* signal was observed in the IZ/ZIII of wt nodules, only a weak signal was observed in a few scattered cells of *Mtefd-1* nodules at 21 dpi (Fig. 1G-H). The weakness of the signal was not due to cell death, which was found by a live/dead assay to be very limited (Fig. 1I-J). The impact of *Mtefd-1* on nodule differentiation is therefore not due to a deregulated plant immune response, in contrast to several other fix- plant mutants (*dnf2*, *symcrk*, *nad-1*, *npd1*), which was further confirmed by transcriptomic data (see below).

The *Mtefd-1* mutation impacts a large proportion of genes expressed during late nodule development

One of the first generations of *M. truncatula* microarrays (Mt16KOLIPlus) was used for the original transcriptomic analyses of the *Mtefd-1* mutant, which revealed 223 and 34 genes respectively down- and up-regulated at least two fold (adj. p-value <0.05) vs wt nodules at 10 dpi, and none at 4 dpi (Vernié et al., 2008). Here, to get a more complete view of genes affected by *Mtefd-1*, we successively used whole-genome Nimblegen microarrays (Verdier et al., 2013) and RNAseq to compare wt and *Mtefd-1* non-inoculated roots, immature (4-day-old) and mature (10-day-old) isolated nodules. The two approaches gave consistent results (supplemental Table S2; supplemental Table S3), but we focus here on RNAseq results (supplemental Fig. S1), which are more sensitive and take advantage of the last (Mt5.0) *M. truncatula* genome release (Pecrix et al., 2018).

The analysis of wt samples showed a large set of genes strongly differentially regulated between nodules and roots, with 2,140 and 4,311 genes >4-fold up-regulated (FDR<0.01) in 4 dpi and 10 dpi nodules respectively (supplemental Fig. S2A-B). Furthermore, the comparison of mature vs immature wt nodules revealed 4,189 genes up-regulated during late nodule development (fold change= FC>4, FDR<0.01) (supplemental Fig. S2C). To determine where they are expressed in the nodule, we made use of the 16 expression patterns previously defined from RNAseq analysis of laser-dissected nodule zones (Roux et al., 2014; Pecrix et al., 2018) (see legend of Fig. 2A). As expected, they were mostly

expressed in the differentiation and nitrogen-fixation nodule zones (expression patterns 6 to 11 and 15-16) (Fig. 2A-C).

The analysis of *Mtefd-1* transcriptomes revealed 246 and 1,719 strongly down-regulated genes compared to wt in 4 and 10 dpi nodules respectively ( $FC < -4$ ,  $FDR < 0.01$ ) (Fig. 2D; supplemental Fig. S2D). This represented 8.6% and 37.6% of the genes up-regulated in wt nodules vs roots, respectively. By contrast, only 14 genes were down-regulated in *Mtefd-1* roots vs wt roots (supplemental Table S3). The analysis of 4 dpi nodules was particularly interesting, because those are less impacted by indirect, developmental effects of *Mtefd-1*. At this stage, the induction of several genes was completely abolished in *Mtefd-1*, including notably *MtRR4* (Table 1). A >4-fold reduction of expression was detected for 18 early (patterns 4 to 6) NCR genes, with notably 9 out of 12 NCR genes in pattern 4 and 5 affected, while the nodule-specific thioredoxin MtTRX S1, regulating NCR activity and required for bacteroid differentiation (Ribeiro et al., 2017) was 65-fold down-regulated in *Mtefd-1* vs wt 4 dpi nodules. Two other families proposed to be involved in nodule differentiation were also affected, namely nodule-specific Glycine-rich proteins (Alunni et al., 2007) and the Medicago lineage-specific SNARP/LEED..PEED family (Trujillo et al., 2014), with *MtNODGRP36* and *MtLP9/MtLP10* down-regulated (153-, 31- and 405-fold respectively) in 4 dpi nodules and additional members of both families affected at 10 dpi (Table 1; supplemental Table S3). In addition, the expression of various genes encoding other secreted Cys-rich peptides of unknown function was severely impaired (50- to 220-fold downregulation), including a cluster of eight leginsulin genes (de Bang et al., 2017) and three knottin-domain nodulins (MtN1, MtN1b and MtN15) (Table 1 and supplemental Table S3). Of note, *MtCCS52a* was moderately but significantly down-regulated ( $FC = -2.3$  and  $-1.6$  at 4 and 10 dpi, respectively), which is interesting considering the importance of this gene for nodule endoreduplication. Finally, the GO analysis revealed a downregulation of defense-related genes in *Mtefd-1* nodules (GO:0006952; p-value =  $7.5E-04$ ; supplemental Table S3).

At 10 dpi, many genes down-regulated in *Mtefd-1* vs wt nodules corresponded to the late differentiation and nitrogen-fixing zones (Fig. 2E), with 48.1% of pattern 6-11 genes down-regulated in *Mtefd-1* (1,890 genes  $FC < -2$ ,  $FDR < 0.01$ ). 455 NCR genes were down-regulated in *Mtefd-1* ( $FC < -4$ ), along with a series of leghemoglobin (GO:0015671) and transporter genes (GO:0006810) (Table 1 and supplemental Table S3). This reflects a major developmental impact of the *Mtefd-1* mutation, with likely many indirect effects. As an illustration of possible cascade effects, the *MtRSD* (pattern 7) and *MtNCR169* (corresponding to the *dnf7* mutant; pattern 9) genes were down-regulated 4- and 9-fold, respectively, in 10 dpi *Mtefd-1* nodules (Table 1). Transcriptomic analyses of *rsd* and *dnf7* mutant nodules have previously identified 163 and 93 down-regulated genes ( $FC \leq -2$ ) compared to wt controls, respectively (Sinharoy et al., 2013; Horváth et al., 2015), all expressed in the late nodule

differentiation zone (pattern 7-11 and 15-16; supplemental Fig. S2F-G). We found that a majority of them were down-regulated in *Mtefd-1* (78.4% for *dnf7* and 79.5% for *rsd-1*; supplemental Fig. S2E), while only 12 genes were shared between *dnf7* and *rsd-1* affected genes. Likewise, *MtDME* was down-regulated 3.4-fold in *Mtefd-1*, similar to the downregulation level leading to fix- nodules in *MtDMEi* knock-down mutants (Satgé et al., 2016), while *MtNLP2* (Jiang et al., 2021) was 10-fold down-regulated. In summary, *Mtefd-1* affects several key regulators of late nodule development.

The *dnf7* mutant nodules exhibit early senescence, with a high induction of *MtCP2* and *MtCP6* cysteine protease genes, which are markers of nodule developmental senescence (Horváth et al., 2015). By contrast, *MtCP2* and *MtCP6* were not differentially expressed in *Mtefd-1* nodules as compared to wt nodules (Table 1), possibly because of the residual level of *NCR169* expression in *Mtefd-1*. Furthermore, genes preventing plant immune responses against rhizobium in nodules (*MtDNF2*, *MtSYMCRK*, *MtNAD1*) were only modestly impacted by the *efd-1* mutation (around 5-fold down-regulation), by comparison with their considerable level of induction in wt nodules (1760-, 756- and 477-fold respectively in 10 dpi nodules vs roots) (Table 1 and supplemental Table S3). Consistently, the defense response genes highly activated in *dnf2*, *symcrk* or *nad1* mutants were either not [*MtVSP* (VACUOLAR SORTING PROTEIN); *MtBGL* (BETA-1,3-GLUCANASE); *MtNDR1* (NON-RACE SPECIFIC DISEASE RESISTANCE PROTEIN 1); *MtPR3* (PATHOGENESIS RELATED PROTEIN 3); *MtPAL* (PHENYLALANINE AMMONIA LYASE); *MtPR10-1*, *10-2*, *10-3*, *10-4*, *10-5*) or only moderately up-regulated (*MtPR10-6*; *MtVPE* (VACUOLE PROCESSING ENZYME))] in 10 dpi *Mtefd-1* vs wt nodules. Consistent with the live/dead assay (Fig. 1J), these transcriptomic data thus confirmed that *MtEFD* controls the differentiation process and not immune responses.

In addition to down-regulated genes, we found a number of up-regulated genes in the *Mtefd-1* mutant compared to wt samples (42 in non-inoculated roots, 718 in 4 dpi nodules and 380 in 10 dpi nodules; FC>4, FDR<0.01; Fig. 2D, supplemental Table S3, supplemental Fig. S2D). At 10 dpi, a fraction of up-regulated genes likely reflects the absence of ZIII development in *efd-1* nodules, mechanically leading to a relative enrichment in ZI and II transcripts compared to wt nodules (Fig. 2A-C). We also observed the upregulation at 4 dpi of auxin-related genes, including *MtLBD11* (Schiessl et al., 2019) and *MtARF13* (AUXIN RESPONSE FACTOR 13) transcriptional activator genes, as well as *MtGH3* (GRETCHEN HAGEN 3) and *MtSAUR1* (SMALL AUXIN UP RNA 1) auxin-responsive genes, potentially related to increased cortical cell divisions (ccd) that we observed in *Mtefd-1* (see below). The 10-fold upregulation at 4 dpi of *MtCEP11* (C-TERMINALLY ENCODED PEPTIDE 11) (Table 1), related to the positive regulator of nodulation *MtCEP1* (Gautrat et al., 2020), might also be linked to the *Mtefd-1* hypernodulation phenotype. More intriguing was the observation that 44 genes normally expressed in the late differentiation zone (patterns 8-11, 15), were up-regulated at 4 dpi

while down-regulated at 10 dpi (FC>4, FDR<0.01) (Table 1 and supplemental Table S3). Those included 21 late NCR genes (notably *MtNCR169*), a cluster of nodule-specific calmodulin-like genes, whose products are targeted to the symbiosome (Liu et al., 2006), as well as a redox control gene, *MtRbohA* (*RESPIRATORY BURST OXIDASE HOMOLOG A*), normally expressed in the zone III (Marino et al., 2011). These genes are thus prematurely induced in *Mtefd-1* nodules. In other words, MtEFD seems to both positively regulate differentiation genes and prevent early expression of a set of late symbiotic genes.

*MtRR4* can complement the hypernodulation but not the nodule differentiation phenotype of the *Mtefd-1* mutant

In line with our original observations indicating that MtEFD positively controls *MtRR4* expression (Vernié et al., 2008), we found here that *MtRR4* and *MtEFD* have a similar expression profile in nodules (Fig. 2A bottom) and that *MtRR4* induction is abolished in *Mtefd-1* nodules (Table 1 and supplemental Fig. S3A for a validation by RT-qPCR analysis). *MtRR4* is by far the most highly expressed member of the *MtRRA* gene family (ten genes) in mature wt nodules (supplemental Table S3). Using RNAseq data from an independent time course analysis (Schiessl et al., 2019), we observed two waves of *MtRR4* induction taking place during nodulation, the first one in the first 24h following rhizobium infection, and the second one about three days later, i.e., at the stage of nodule primordium growth and young nodule emergence from the root (supplemental Fig. S3B). Since *MtEFD* is not induced before about 2 dpi (Vernié et al., 2008; Schiessl et al., 2019), it is likely that only the second wave of *MtRR4* induction is impaired in *Mtefd-1*.

Since *MtRR4* is thought to negatively regulate cytokinin signaling, we decided to assess the impact of the *Mtefd-1* mutation on the expression of a reporter of CK signaling, *TCSn:GUS* (Zürcher et al., 2013; Jardinaud et al., 2016), using hairy root transformation. In wt and *Mtefd-1* *S. meliloti*-inoculated roots, *TCSn:GUS* was expressed in nodule primordia at all developmental stages, from stage I to stage VI as defined by (Xiao et al., 2014) (Fig. 3A-D). In the *Mtefd-1* mutant, we observed a more intense GUS signal in individual cells within nodule primordia (Fig. 3B and D) as well as frequent unorganized cortical cell divisions, associated or not with infection threads (Fig. 3E). In mature (10 dpi) wt nodules, the *TCSn:GUS* signal was observed in a narrow, meristematic or submeristematic region (Fig. 3F, supplemental Fig. S4). By contrast, in *Mtefd-1* nodules, *TCSn:GUS* was expressed in a broader region including the infection / differentiation zones (Fig. 3G, supplemental Fig. S4). Yet, a marker of the central and vascular nodule meristems, the *pWOX5* (*WUSCHEL-RELATED HOMEODOMAIN-5*):*GUS* construct (Osipova et al., 2012; Roux et al., 2014), remained expressed in highly localized apical cells in *Mtefd-1*, as in wt nodules (Fig. 3H and I). Of note, *Agrobacterium rhizogenes*-mediated

transformation may change the hormonal balance but this would also be the case in the control wt background. Overall, this suggested an alteration of CK signaling in the infection/differentiation zones of *Mtefd-1* nodules, while the nodule meristematic centers are not affected. By contrast, auxin signaling appeared to be minimally affected in mature *Mtefd-1* nodules, based on the expression profile of a *DR5:GUS* auxin-responsive reporter (Ulmasov et al, 1995) (supplemental Fig. S4).

Bearing in mind the importance of CK for both early and late stages of nodule development (Plet et al., 2011), we then asked whether *MtRR4* could complement the different nodulation phenotypes of *Mtefd-1*. *MtRR4* was expressed under the control of *MtEFD* promoter to ensure a proper localization of its transcripts and transformed in *Mtefd-1* using *A. rhizogenes*. To assess the complementation efficiency in a sensitive way, nodules were induced using *S. meliloti* carrying the *NifH:GFP* construct. As expected, transformation of *Mtefd-1* with *pMtEFD:MtEFD* produced pink nodules, with numerous cells expressing *NifH:GFP*, as in wt control nodules (Fig. 4A-C). By contrast, nodules produced in *pMtEFD:MtRR4*-transformed roots were very similar to empty vector-transformed nodules, i.e., small, white and with very few GFP-expressing cells (Fig. 4D) even though *MtRR4* was well expressed (as verified by RT-qPCR; supplemental Fig. S5). Furthermore, an analysis of nodule cell endoreduplication showed that the *pMtEFD:MtRR4* construct was unable to restore the 32C peak in *Mtefd-1* nodules, in contrast to *pMtEFD:MtEFD* (Fig. 4E). We then tested whether *pMtEFD:MtRR4* could complement the *Mtefd-1* hypernodulation phenotype. In contrast to the nodule development phenotype, we found that *pMtEFD:MtRR4* was as efficient as *pMtEFD:MtEFD* in restoring a normal level of nodulation (Fig. 4F). Altogether, this suggested that *MtRR4* downregulation is a major (and possibly the only) component of the *Mtefd-1* hypernodulation phenotype, in contrast to its nodule differentiation phenotype, thus likely to involve other players.

The *Arabidopsis* orthologue of *MtEFD*, *AtERF003*, complements the nodule differentiation phenotype of *Mtefd-1*, in contrast to a close *MtEFD* paralogue, *MtEFD2*

We identified proteins closely related to *MtEFD* in a variety of legume and non-legume plant species. Those included a *MtEFD* paralogue in *M. truncatula* dubbed *MtEFD2* (MtrunA17\_Chr3g0137174; 67.7% amino acid identity). The top 30 hits of Blastp searches in the ncbi non-redundant (nr) protein database with *MtEFD* were only legume ERFs (expect values <1e-82). By contrast, 14 non-legume ERFs were found amongst the 30 top hits for *MtEFD2* (expect values <1e-92). Multiple alignment analyses indicated that the AP2/ERF DNA binding domain of *MtEFD*/*MtEFD2* subgroups and *AtERF003*, the only *Arabidopsis thaliana* ERF closely related to *MtEFD*, are highly similar, with a perfect conservation of amino acids involved in the direct interaction with DNA (Fig. 5A, supplemental Fig. S6). A phylogenetic tree was generated with the protein fraction following the

AP2/ERF domain, using AtERF003, the 15 best hits for MtEFD and MtEFD2, and MtERN1 as an outgroup (subgroup Vb, vs subgroup Va for MtEFD). It showed that MtEFD2 and AtERF003 belong to a same subgroup, clearly distinct from the MtEFD subgroup (supplemental Fig. S7), with a further separation among MtEFD2-related proteins between legume and non-legume proteins. In the protein fraction following the AP2/ERF domain, two motifs described for the group V ERFs, termed CMV-1 and CMV-2 (Nakano et al., 2006; Cerri et al., 2015) are also conserved amongst EFD/EFD2-related proteins, albeit with some differences between the two subgroups (Fig. 5A; supplemental Fig. S6). The remaining region of about sixty amino acids, located between CMV-1 and the C-terminal CMV-2, is more variable and poorly structured, according to AlphaFold modeling (supplemental Fig. S8). Yet an analysis of a set of 39 accessions of Pea (a species genetically closely related to *M. truncatula*), representing a wide range of diversity (Siol et al., 2017), showed that this region is well conserved in PsEFD (supplemental Fig. S9), suggesting a possible role in the structure and /or function of EFD.

We found that *pMtEFD2:GUS* had a different expression pattern than *pMtEFD:GUS* in *A. rhizogenes*-transformed roots and nodules. In non-inoculated roots, *pMtEFD2:GUS* was expressed in the meristematic zone of the root tip, while *pMtEFD:GUS* was expressed in both zones flanking the root meristem, notably in the transition zone (where root cells elongate and begin to differentiate) (Fig. 5B-C). Consistently, independent transcriptome analyses revealed that *MtEFD2* expression is positively regulated by auxin, which is not the case for *MtEFD* (Herrbach et al., 2017). Following *S. meliloti* inoculation, the expression of both genes was detected in nodule primordia (Fig. 5D-E), including at very early stages, consistent with RNAseq analyses of spot-inoculated roots (Schiessl et al., 2019) (supplemental Fig. S3B). However, from the stage III primordia (establishment of the nodule meristem (Xiao et al., 2014)), *pMtEFD:GUS* seemed more expressed at the base of the primordium (Fig. 5D), possibly corresponding to cells undergoing a differentiation process, which was not the case with *pMtEFD2:GUS*. In 4 dpi nodules, *MtEFD* and *MtEFD2* were up-regulated 133-fold and seven-fold vs roots, respectively, with *MtEFD* transcripts being about nine times more abundant than *MtEFD2* transcripts (supplemental Table S3). In mature wt nodules, the maximal expression of *MtEFD* and *MtEFD2* was in ZII and in ZI, respectively (laser capture microdissection-RNAseq and promoter:GUS analyses; Fig. 2A and Fig. 5F-G), with some overlapping expression in distal ZII. In the *Mtefd-1* mutant, *pMtEFD2:GUS* was weakly expressed except for a few scattered cells in the nodule primordia (Fig. 5H), consistent with a 2.6-fold decrease of *MtEFD2* expression detected by RNAseq in *Mtefd-1* vs wt 4 dpi nodules (supplemental Table S3).

Since MtEFD and MtEFD2 proteins are closely related, we then wondered whether *MtEFD2* could complement the *Mtefd-1* nodule phenotype if expressed from the *MtEFD* promoter. To further

explore possible neofunctionalization processes, we also tested a non-legume protein from the MtEFD2 subgroup, AtERF003. Following hairy root transformation of *Mtefd-1* and inoculation with *S. meliloti* *NifH::GFP*, we found that *pMtEFD:MtEFD2*-transformed roots were unable to develop pink nodules at three weeks post-inoculation (wpi), in contrast to *pMtEFD:MtEFD* –transformed roots, while a few pink nodules were observed at six wpi (Table 2, Fig. 5I). In contrast, surprisingly, *pMtEFD:AtERF003* efficiently complemented *Mtefd-1* at three wpi, with the production of pink nodules harboring numerous *pnifH::GFP* expressing cells, and plant and bacteroid endoreduplication levels similar to the *pMtEFD:MtEFD* control (Table 2, Fig. 5I). Furthermore, *pMtEFD:AtERF003* also complemented the hypernodulated phenotype of *Mtefd-1*, in contrast to *pMtEFD:MtEFD2* (Fig. 5J). We identified one non-conservative modification (P57 replaced by T57) in the AP2/ERF DNA binding site of MtEFD2, not found in MtEFD nor AtERF003 (Fig. 5A). However, the reciprocal substitution of this amino acid residue in MtEFD and MtEFD2 did not modify the outcome of the complementation assay (Table 2). We then performed a reciprocal domain swap between MtEFD and MtEFD2, joining the N-terminal part, including the AP2/ERF DNA binding site, of one protein (amino acids 1-64) to the remaining C-terminal part of the other. We found that a MtEFD2:MtEFD chimeric protein (MtEFD2 N-terminal part joined to MtEFD C-terminal part) was fully able to complement *Mtefd-1*, whereas the reciprocal MtEFD:MtEFD2 chimeric protein was totally inefficient (Table 2). This confirmed that the AP2/ERF domains of MtEFD and MtEFD2 are functionally identical, and showed the importance and neofunctionalization of the remaining part of the proteins.

#### *Mtefd2* mutations decrease the nodule number in *Mtefd-1*

Since *MtEFD2* is up-regulated seven-fold in wt 4 dpi nodules, we tested whether *Mtefd2* mutations could impact nodulation. The analysis of a mutant with a TnT1 insertion in the DNA binding site (Noble Foundation population, R108 background, line NF-9808) revealed no nodulation phenotype. Then, to investigate possible interactions between *Mtefd-1* and *Mtefd2* mutations, we set up a multi-guide CRISPR-CAS9 mutagenesis approach based on *A. rhizogenes*-mediated transformation, as previously described for *MtNLP2* (Jiang et al., 2021). The four-guide construct that we used had a very high editing rate of *MtEFD2* (>70%) (supplemental Fig. S10), while no off-target edits were detected in *MtEFD* (n=36). As for insertion mutants, no significant difference (*P*-value=0.26, Kruskal-Wallis test) was observed in nodule number for these *Mtefd2* mutants in a wt A17 background. In contrast, in the *Mtefd-1* background, a significant decrease in nodule number was observed for bi-allelic mutants of *Mtefd2* at 17 dpi (Fig. 5K). The nodules formed in the double mutant were similar in terms of development to those formed in *Mtefd-1* (supplemental Fig. S11).

RNAseq analysis of nodulated root fragments at 4 dpi revealed a set of 122 genes that were down-regulated in the *Mtefd2* mutants, in both the wt A17 and *Mtefd-1* backgrounds, whereas they were up-regulated in the *Mtefd-1* single mutant (Table 3; supplemental table S4). A majority of them (70.5%) were induced in wt nodules 4 dpi (FC>4). These genes represent possible candidates for mediating the impact of *Mtefd2* on nodulation. They included notably five *MtNodGRP* (Alunni et al., 2007) (mean downregulation of ~30-fold in *Mtefd2*), the symbiotic remorin gene *MtSYMREM1* (Lefebvre et al., 2010) (downregulated ~20-fold), as well as *MtNF-YA1* and *MtNF-YB16*, whose products interact in the nodule (Baudin et al., 2015) (downregulated >4-fold and 2.5-fold in *Mtefd2*, respectively). Consistently, the six *STY* (*SHORT INTERNODES/STYLISH*) genes and one *YUCCA* gene belonging to the NF-YA1 regulation cascade involved in auxin signaling and nodule emergence (Shrestha et al., 2021) were also downregulated in *Mtefd-2* at 4 dpi (table 3 and supplemental table S4).

## Discussion

While substantial progress has been made in recent years in understanding nodule initiation, the control of later stages of nodule development remains relatively undocumented. Our further characterization of the EFD transcription factor supports its key role in the control of symbiotic differentiation of *M. truncatula* nodules.

Thus, we found that endoreduplication of both rhizobium and plant cells, a hallmark of differentiation, is impaired in the *Mtefd-1* KO mutant, with a pattern not yet described in other fix-mutants, namely the absence of 32C plant nuclei (64C nuclei being barely detectable in wt nodules in our conditions). Importantly, no concomitant cell death and strong defense reactions were observed in *Mtefd-1* nodules, supporting the primary defect of *Mtefd-1* being differentiation and not the control of the plant immune response toward rhizobium, in contrast to other reported plant fix-mutants. The absence of 32C nuclei in *Mtefd-1* might be related to the down-regulation of *MtCCS52a* expression. Indeed *MtCCS52a* antisense lines with a ~2.5 fold decrease in *MtCCS52a* transcript accumulation (similar to the decrease of *MtCCS52a* expression in 4 dpi *Mtefd-1* nodules) also exhibit a strong reduction of 32C nuclei (Vinardell et al., 2003). Furthermore, since transcription waves coincide with growing ploidy levels (Nagyimihaly et al., 2017), there might be an EFD-dependent developmental transition corresponding to the last cycle(s) of endoreduplication, possibly taking place in the nodule proximal zone II, where *MtCCS52a* is maximally expressed (Roux et al., 2014). In agreement with this, our RNAseq analysis indicated that the expression of a strong proportion of genes specifically expressed in wt IZ is impaired in *Mtefd-1* nodules (notably 46.7% of pattern 8 and 9 genes, i.e., 616 genes, with a >4-fold down-regulation compared to wt nodules). Those include two



important symbiotic regulator genes, namely *MtRSD* (Sinharoy et al., 2013) and *MtNCR169* (Horváth et al., 2015), expressed at a later stage of development than *MtEFD* (Fig. 2A). Consistent with *MtRSD* and *MtNCR169* acting downstream *MtEFD*, we found that a large fraction (~75%) of genes affected in *rsd-1* and *ncr169* mutants were also down-regulated in *Mtefd-1*.

Another advance provided by this work was the identification of hundreds of genes affected in very young (4 dpi) *Mtefd-1* nodules, less prone to indirect developmental effects. In addition to the previously identified EFD target gene, *MtRR4* (Vernié et al., 2008), we found a series of genes whose upregulation is abolished or highly impaired in *Mtefd-1*. Those include a set of genes encoding small secreted, Cys-rich peptides, that is 24 early NCR, 8 leginsulin genes, *MtN1*-related and *MtN15* nodulin genes, 4 LTP and defensin genes, all massively induced during wt nodulation. Some of them may contribute to rhizobium differentiation, as NCR peptides or nodule-specific Glycine-rich proteins (Alunni et al., 2007) and SNARP/LEED..PEED peptides (Laporte et al., 2010; Trujillo et al., 2014), also affected in *Mtefd-1*. Expression of the nodule-specific, symbiosome-localized, redox regulator *MtTRX S1* (Ribeiro et al., 2017) was also strongly inhibited in *Mtefd-1* (>60-fold at both 4 and 10 dpi), suggesting that not only the expression but also the activity of symbiosome-targeted peptides is likely affected. Of note, RNAi-mediated downregulation of *MtTRX S1* is sufficient to alter bacteroid differentiation and nodule activity (Ribeiro et al., 2017).

We also found out that many (718) genes are up-regulated in 4dpi *Mtefd-1* nodules, compared to wt nodules. In addition to auxin-related genes, hypothesized to relate to increased Rhizobium-induced ccd in *Mtefd-1* (see below), 44 genes, normally expressed in the late differentiation zone of wt nodules (patterns 8, 9, 10, 11, 15), were up-regulated at 4 dpi but down-regulated in 10 dpi *Mtefd-1* nodules. This suggested an altered timing of transcriptional activation, with *MtEFD* controlling both the repression (in young nodules) and activation (in mature nodules) of genes associated with late stages of nodule development, notably encoding: NCR peptides [including *NCR169* (Horváth et al., 2015)], nodule-specific symbiosome-targeted calmodulin-like proteins (Liu et al., 2006), and a zone III-associated redox regulator, *MtRbohA* (Marino et al., 2011). Since the expression of these genes is normally finely regulated in successive waves (Maunoury et al., 2010; Guefrachi et al., 2014; Roux et al., 2014), it is possible that their premature expression is deleterious to the differentiation process. Thus, NCR peptides are diverse in terms of biochemical properties, localization in the symbiosome, and probably interactors and functions (Mergaert, 2018). Modifying the timing of expression and the balance between different types of NCRs may therefore strongly modify the outcome of their action.

In summary, both the down and upregulation of multiple nodule-associated genes in *Mtefd-1* could be responsible for *Mtefd-1* defects in plant and bacteroid differentiation. From a mechanistic point of

view, EFD and all related proteins carry a so-called conserved motif V-2, highly enriched in acidic and apolar amino-acid residues. This motif is strongly reminiscent of the EDLL motif found in the subclass IX of ERFs and demonstrated to be a transcriptional activation domain (Tiwari et al., 2012). By contrast, EFD does not seem to carry a repressive domain, such as the EAR motif found in several ERFs (Nakano et al., 2006). Consequently, the EFD repressive activity is probably either indirect or mediated by interactions with repressive factors. This was observed with AP2 proteins which act mainly as activators, but are able to interact with the TOPLESS repressor in other protein complexes (Horstman et al., 2014).

Among the set of target genes (direct or indirect) of MtEFD, we paid particular attention to *MtRR4* because of the key role played by CK for nodule inception and development. *MtRR4* is the most highly expressed type A response regulator gene in mature wt nodules, and its expression is abolished in *Mtefd-1* nodules. The CK signaling reporter, *TCSn:GUS*, was found to be expressed in a narrow submeristematic region in wt nodules, but in a much larger zone in *Mtefd-1* nodules (Fig. 3F-G). This might be due to the absence of MtRR4, thought to be a negative regulator of CK signaling. Considering the links between CK and the cell cycle control (Schaller et al., 2014), including endoreduplication, the *Mtefd-1* mutant would probably be a useful tool to investigate this issue in nodules. Our attempts to complement *Mtefd-1* using *MtRR4* expressed from the *MtEFD* promoter were successful for the hypernodulation but not the nodule differentiation phenotype. This suggests that MtRR4 plays a major role in a MtEFD-dependent negative feedback loop for nodulation, while additional or alternative players are involved in the control of nodule differentiation. Examination of RNAseq data from a detailed time course study (Schiessl et al., 2019) revealed two waves of *MtRR4* transcriptional activation during early nodulation stages (supplemental Fig. S3). We propose that the first wave is induced by the burst of CK signaling in the cortex leading to ccd and nodule primordium formation, since *MtRR4* is a primary CK response gene (Plet et al., 2011). The second one might be consecutive to *MtEFD* transcriptional activation, and contribute to the negative feedback regulation of the CK pathway (see model in Fig. 6). Considering the numerous additional and disorganized ccd that we observed in *Mtefd-1* roots (Fig. 3E and Fig. 5H), it is tempting to hypothesize that MtEFD may act locally via MtRR4 to restrict ccd and the number of nodules. In the same time, MtEFD would contribute to the differentiation of cells resulting from ccd and nodule meristem activity (most likely once infected). Whether the MtEFD-dependent control of nodulation relates to the systemic autoregulation of nodulation (AON) mechanisms remains an open question. It was proposed that *MtEFD* expression could be activated by the CLE/SUNN (CLAVATA3/EMBRYO-SURROUNDING REGION / SUPERNUMERIC NODULES) AON pathway because *MtCLE12* overexpression leads both to *MtEFD* induction (albeit moderate: 2.3 fold) and repression of nodulation, in a *SUNN*-dependent way (Saur

et al., 2011). However, the impact of *MtCLE12* overexpression on nodulation could not be tested in the *Mtefd-1* mutant (Saur et al., 2011), preventing evaluation of the role of *MtEFD* in the CLE/SUNN pathway.

During nodule evolution, *MtEFD* has probably been recruited from a root developmental program, because it is only expressed in roots and nodules, in both cases in the differentiation zone. We could not however identify a robust root phenotype co-segregating with the *Mtefd-1* mutation, possibly because of functional redundancy issues or technical difficulties due to root plasticity. As a first approach to investigate the evolution of *MtEFD*, we asked whether two *MtEFD* closely related genes, from *M. truncatula* (*MtEFD2*) and from *A. thaliana* (*AtERF003*), expressed under the control of the *MtEFD* promoter, could complement the nodule differentiation phenotype of *Mtefd-1*.

Unexpectedly, we found that *AtERF003* but not *MtEFD2* efficiently complemented *Mtefd-1*. This strongly suggests that the main innovation for the symbiotic function of *MtEFD* resides in its promoter, even though *MtEFD* regulates a large set of nodule-specific genes that do not exist in *A. thaliana* (which is unable to establish endosymbiotic interactions, whether with nitrogen-fixing bacteria or mycorrhizal fungi). This is in line with recent studies which revealed that a major mechanism for the emergence of nitrogen-fixing symbioses was the evolution of *cis* regulatory elements in promoter regions (Liu and Bisseling, 2020; Mergaert et al., 2020; Dong et al., 2021).

Still, how *AtERF003* can activate nodule-specific genes is an intriguing question. A first element of response is that the DNA binding sites (DBS) of *MtEFD* and *AtERF003* are highly conserved. Furthermore, *MtEFD* does not activate its nodule-specific target genes in the root, nor following its induction by the bacterial pathogen *Ralstonia solanacearum* (Moreau et al., 2014). Therefore, another layer of regulation must exist in the nodule, to which *AtERF003* is also responsive. This could be the presence of nodule-specific interactors in transcriptional complexes and / or the presence of a peculiar epigenetic landscape. Indeed many nodule differentiation genes are found in genomic regions (notably the so-called symbiotic islands) strongly enriched in epigenetic marks which are highly differential in nodules vs roots or between nodule zones (Satgé et al., 2016; Nagymihaly et al., 2017; Pecrix et al., 2018). DNA demethylation and replacement of repressive by active histone marks could thus be important factors to give *MtEFD/AtERF003* access to a set of promoter regions. In support of this hypothesis, DNA affinity purification sequencing (DAP-seq) showed that the binding to DNA of many AP2/ERFs is sensitive to cytosine methylation (O'Malley et al., 2016).

The DBS of *MtEFD2* is also very similar to *MtEFD* DBS, which it could efficiently replace in a chimeric *EFD2:EFD* protein, in contrast to the rest (C-terminal part) of *MtEFD2* protein. This suggests a neofunctionalization of the *MtEFD2* protein involving modifications in its C-terminal part, in addition

to a distinct expression pattern (*MtEFD2* being maximally expressed in/very close to the root and nodule division zones). Precise information is lacking regarding the role of the C-terminal part of MtEFD-related proteins, but it is likely involved in protein-protein interactions. Although we were unable to detect a nodulation phenotype of single *Mtefd2* mutants, *Mtefd-1 Mtefd2* double mutants were affected in the number of nodules produced. This suggests that MtEFD2 plays a role in nodule inception, possibly partially overlapping with MtEFD role, or perhaps only in the particular environment resulting from the *Mtefd-1* mutation. It is possible, for example, that the downregulation of *MtNF-YA1* and *MtNF-B16* resulting from the *Mtefd2* mutations does not affect the normal level of nodulation but becomes a limiting factor for hypernodulation in the *Mtefd-1* background. Transcriptome analyses suggested that MtEFD2 contributes to the activation of MtNF-YA1/MtNF-B16 and their regulatory cascade involved in auxin biosynthesis six *STY* genes and one *YUCCA* gene. Since *STY* genes are required for nodule emergence (Shrestha et al., 2021), it is tempting to speculate that this contributes to decreased nodule formation in *Mtefd-1 Mtefd2*. In any case, the *MtEFD/MtEFD2* case is reminiscent of several other examples of paralogous genes playing overlapping or synergistic roles during nodulation, such as *MtNOOT1/MtNOOT2* (Magne et al., 2018) *MtERN1/MtERN2* (Cerri et al., 2016), *MtNF-YA1/MtNF-YA2* (Laloum et al., 2014).

## Conclusions

Our study revealed the recruitment to the nitrogen-fixing symbiosis of two proteins probably originally operating in roots, MtEFD and MtEFD2, expressed in the differentiation and meristematic zones of both roots and nodules, respectively, and with distinct neofunctionalization processes for each of them. We propose that, following the evolution of its promoter region leading to a strong induction in nodules, *MtEFD* has become a major regulator of nodule symbiotic differentiation, contributing directly or indirectly to the activation of almost half of the genes specifically expressed in the late differentiation zone, while preventing the premature activation of some of them. In addition, MtEFD is involved in the control of root cortical cell divisions and nodule number, probably through MtRR4 action on CK signaling. For a better understanding of the MtEFD-associated upstream and downstream regulation network, it would now be necessary to identify the *cis* and *trans* elements that control *MtEFD* expression, and to identify the DNA sites bound by MtEFD, keeping in mind that these may be affected by DNA methylation. Identification of protein interactors and possible post-translational modifications would also be valuable to better understand the differences in the mode of action of MtEFD and MtEFD2, and more generally to document the functional importance of the C-terminal part of AP2/ERF proteins.

## Materials and Methods

### Plant genotypes

The *Medicago truncatula* A17 was used as the wt reference. The *Mtefd-1* null deletion mutant (backcrossed twice) was previously described (Vernié et al., 2008). The TR36, TRV36, TRV43, TR183 and TR3 mutant seeds (Maunoury et al. 2010) were provided by Peter Mergaert (I2BC, Gif-sur-Yvette). The *Mtefd2* mutant was generated by multi-guide CRISPR-Cas9 mutagenesis of *MtEFD2* using *M. truncatula* A17 hairy root transformation. It was therefore a mix of mutations, with >95% of large deletions or frame-shift mutations, as described in supplemental Fig. S9. The *Mtefd-1 Mtefd2* double mutant was similarly generated by CRISPR-Cas9 mutagenesis of *MtEFD2* using *Mtefd-1* hairy roots.

### Plant growth and hairy root transformation

Seeds were sterilized and placed at 4°C during 5 days then germinated one night at 25°C. Germinated seeds were grown on agar plates containing Fahraeus medium supplemented with 0.5 mM NH<sub>4</sub>NO<sub>3</sub> and covered with a filter paper (at 25°C; light/dark photoperiod of 16h/8h), in attapulgit (at 25°C) or in aeroponic caissons (at 22°C; 75% hygrometry; light intensity: 200 µE.m<sup>-2</sup>.s<sup>-1</sup>; light-dark photoperiod: 16 h-8 h). In aeroponic caissons, plants were grown for about seven days in aeroponic medium supplemented with 5 mM NH<sub>4</sub>NO<sub>3</sub> (Barker et al., 2007) and then in nitrogen-free aeroponic medium three days before inoculation with *Sinorhizobium meliloti* 2011.

Hairy root transformations (analyses of all GUS constructs, *Mtefd-1* complementation experiments and CRISPR-Cas9 mutagenesis of *MtEFD2*) were performed using ARqua1 *Agrobacterium rhizogenes* as described (Boisson-Dernier et al., 2001). Composite plants were selected on kanamycin 25 mg/L (20°C; day/night, 12h) and systematically checked for the expression of the *DsRED* marker present on the T-DNA construct (stereomicroscope MZFLIII, Leica). They were transferred in aeroponic caissons following two weeks of growth on plates at 20°C (Fahraeus medium with 5 mM NH<sub>4</sub>NO<sub>3</sub>) and grown for seven days in the presence of 5 mM NH<sub>4</sub>NO<sub>3</sub> before nitrogen starvation and inoculation with *S. meliloti*.

Under aeroponic conditions, nodules started fixing nitrogen at around 7 dpi and 10 dpi for non-transformed plants and transformed composite plants, respectively.

### Light and fluorescence microscopy

Bacterial viability (LIVE/DEAD® BacLight Viability Kit -Thermoscientific) and *NifH::GFP* expression were analyzed using 80 µm sections (microtome LEICA VT 1000S) of nodules embedded in 8% low-melting agarose (NuSieve™ GTG™ Agarose Lonza). They observed with an Axioplan 2 microscope (Zeiss). For GFP and SYTO™ 9, the fluorescence observations were made using an excitation wavelength of 475nm, a barrier filter of 394 nm and an emission filter of 397nm. For propidium iodide and DsRed,

the set up was an excitation wavelength of 546 nm, a barrier filter of 580 nm and an emission filter of 590 nm. Images were taken with an Axiocam MRc camera (Zeiss). In complementation assays, sections were stained with 0.5 µg/ml DAPI (4', 6-diamidino-2-phenylindole) (excitation wavelength of 365 nm, barrier filter 395 nm and emission filter 397 nm).

For analysis of promoter:GUS fusions, roots or nodules were pre fixed in paraformaldehyde 0.4% (v/v) for 1h, stained for  $\beta$ -glucuronidase (GUS) activity for 4h at 37°C and fixed in glutaraldehyde 2% (v/v). Longitudinal sections were made from roots embedded in 8% (w/v) low-melting agarose, using a vibrating microtome (Leica VT 1000S), and observed with a Zeiss Axiophot light microscope.

For analyses of thinner sections (*TCSn:GUS*-transformed *S.meliloti*-inoculated roots), roots fragments were fixed under vacuum in 0.1 M sodium phosphate buffer containing 0.3% (v/v) paraformaldehyde, rinsed in Phosphate buffer (three times) and incubated in GUS staining buffer [0.1 M NaPO<sub>4</sub> pH7.4, 400 µM K<sub>3</sub> Fe(CN)<sub>6</sub>, 400 µM K<sub>4</sub>Fe(CN)<sub>6</sub>, 0.625% (w/v) X-Gluc] for 6 hours at 37°C. Samples were rinsed and fixed again in 0.1 M Phosphate buffer containing 2% (v/v) glutaraldehyde. Roots were then dehydrated by sequential treatments in 30, 50, 70, 90 and 100% (v/v) ethanol and embedded in glycolmethacrylate (Technovit 7100; Haereus-Kulzer) according to the manufacturer's instructions. Images of 10 µm sections were acquired using a Nanozoomer HT (Hamamatsu).

#### RNA extraction and transcriptome analyses

Total RNA was extracted from ~1cm-long root tips, isolated root nodules or nodulated root fragments, using the Qiagen rNeasy Plant kit according to the supplier's instructions. These root tips represent good reference samples as they contain the region of *MtEFD* and *MtEFD2* expression and part of the nodulation competence zone. Microarray analyses were carried out at URGV (Evry, France), using microarrays based on the Roche-NimbleGen technology. High density microarray slides contained twelve chambers, each with 249,087 oligonucleotides representing 83,029 *M. truncatula* A17 loci (each oligonucleotide in triplicate). Four independent biological replicates were analyzed, with  $\geq 12$  aeroponically grown plants per replicate. For each comparison, one technical replicate per biological replicate with fluorochrome reversal was performed (i.e., four dye-switch hybridizations per comparison). The labeling of cRNAs with Cy3-dUTP or Cy5-dUTP (Perkin-Elmer-NEN Life Science Products) and the hybridization to the slides were performed as described in (Lurin et al., 2004). Two micron scanning was performed with InnoScan900 scanner (Innopsys<sup>R</sup>, Carbonne, FRANCE) and raw data were extracted using Mapix<sup>R</sup> software (Innopsys<sup>R</sup>, Carbonne, FRANCE). For each array, the raw data comprised the logarithm of median feature pixel intensity at wavelengths 635 nm (red) and 532 nm (green); a global intensity-dependent normalization using the Loess procedure (Yang et al., 2002) was performed to correct the dye bias. Differential analysis was based

on log-ratios averaged over duplicate probes and technical replicates. Hence the number of available data for each gene equaled the number of biological replicates and was used to calculate the moderated t-test (Smyth, 2004). Since no evidence that specific variances vary between probes was found by Limma, moderated t-statistic was assumed to follow a standard normal distribution. Adjusted p-values were calculated using the optimized FDR approach (Storey and Tibshirani, 2003) with R software (R development Core Team, 2005). SqueezeVar function of the library Limma was used to smooth the specific variances by computing empirical Bayes posterior means. The library kerfdr was used to calculate the adjusted p-values.

RNAseq analyses were performed with three biological replicates, each representing pools of >12 independent plants or transformed roots. Oriented RNAseq was carried out at the GeT-PlaGe core facility, INRAE Toulouse, France (<http://www.get.genotoul.fr>). RNAseq libraries were prepared according to Illumina's protocols using the TruSeq Stranded mRNA sample prep kit (Illumina) to analyze mRNA. Briefly, mRNA was selected using poly-T beads. Then, RNA was fragmented to generate double-stranded cDNA, and adaptors were ligated to be sequenced. Libraries were amplified by 11 cycles of PCR. Library quality was assessed using a Fragment Analyzer (Advanced Analytical Technologies, Inc.) and libraries were quantified by qPCR using a Library Quantification Kit (Kapa). RNA-seq experiments were performed using a NovaSeq technology, with a paired-end read length of 2 × 150 bp.

Expression measure was performed using *M. truncatula* annotation version 5.1.8. We used nf-core/rnaseq pipeline version 3.0 (doi:10.5281/zenodo.4323183) with the following parameters " --skip\_alignment --pseudo\_aligner salmon " that performs adapter and quality trimming with Trim Galore software version 0.6.6. followed by transcript assignation and quantification with salmon tool (version 1.4.0). DEGs were detected with EdgeR Bioconductor package version 3.34.0 (Robinson and Smyth, 2008). Genes with no counts across all libraries were discarded prior to further analysis. Normalization was performed using trimmed mean of *M* values method (Robinson and Oshlack, 2010). Heatmaps showing correlation between replicates were generated using the package pheatmap version 1.0.12 (<https://CRAN.R-project.org/package=pheatmap>) with sample-to-sample Euclidean distances. DEGs were called using the GLM (Fitted generalized linear models) likelihood ratio test, with an FDR adjusted p-value (Benjamini and Yekutieli, 2001). Analyses of GO term enrichment were performed using the topGO package version 2.44.0 (Alexa et al., 2006). Correspondences between gene identifiers (Nimblegen, Affymetrix, Mt20120830, Mt4.0 and Mt5.0), as well as gene names/acronyms were from Legoo (Carrère et al., 2019; Carrere et al., 2021) and downloaded from the downloads section of Mt5.0 genome browser (<https://medicago.toulouse.inra.fr/MtrunA17r5.0-ANR/>), as of 20211025. NCR genes were annotated using (Montiel et al., 2017) and (de Bang et al., 2017), with correspondences to Mt5.0 annotated

proteins established using blastP. The published NCR number was used for hits with > 90% identity over >90% query. Additional NCR numbers (MtNCR-new1...) were used for 86 Mt5.0 proteins with the Interpro domain IPR009810:Late nodulin and not corresponding to published NCRs. In total 691 NCR genes were annotated in Mt5.0.

### Phylogenetic analysis

Sequences were aligned with MAFFT (v7.310) (Kato and Toh, 2008). Columns with more than 50% of gaps were pruned. The phylogenetic tree was built with IQ-TREE (v1.5.5) (Nguyen et al., 2015). IQ-tree used ModelFinder (Kalyaanamoorthy et al., 2017) to identify JTT+G4 as the best-fit evolutionary model according to Bayesian information criterion (BIC). Branches were tested by SH-like aLRT (Anisimova and Gascuel, 2006). Finally, the tree was drawn with Itol (Letunic and Bork, 2021).

### Statistical analysis

Endoreduplication levels were compared using one-way Anova followed by post hoc Tukey test. Normality (Shapiro's test) and homoscedasticity (Bartlett's test) were previously verified ( $P$ -value>0.05). Transcriptomic and phylogenetic data were analyzed as described above. Nodule numbers were statistically analyzed with the Kruskal-Wallis multiple comparison test or the Wilcoxon two-sample method using R software (version 4.1.0) (two-sided tests) and the "agricolae" package (version 1.3-5). Boxplots were drawn with the 'ggplot2' package (version 3.3.5).

### Construct production

The TCSn:*GUS* construct was previously described (Jardinaud et al., 2016). For *pMtWOX5:GUS*, a 2108 bp fragment upstream the ATG was PCR amplified (PrimeSTAR® HS DNA Polymerase, TaKaRa) using *MtWOX5* promoter Forward and Reverse (supplemental Table S5). The amplicon was then cloned into a modified pCAMBIA<sub>2200</sub> vector (Fliegman et al., 2013) using the *BsaI* sites to generate the *pMtWOX5:GUS* binary vector. For the transcriptional fusion between the *MtEFD* and *MtEFD2* promoters and the *GUS* gene, 960bp and 2457bp regions were amplified respectively, using *MtEFD* and *MtEFD2* promoter Forward and Reverse primers (supplemental Table S5). The *GUS* fusion was achieved using *BsaI* digest-T4 DNA ligase ligation protocol with a mix of amplicon and level 0 plasmids EC47822 (pL1V-R2), EC75111 (pL0M-SC-GUS) and EC41414 (pL0M-T-T35S-1) (Schiessl et al., 2019). Level2 assembly was performed using *BpiI* digest-T4 DNA ligase ligation protocol and the following level1 blocs: the previously generated level1 assembly, EC15529 (pL1M-R1-p35S-Kana-TNOS); pUbiq (EC15062- pAtUbi10; EC15073 -DSRed; EC41432-TOCS-1;EC47822), End Linker (EC41766 – pL1M-ELE-3) and the binary vector EC50506 (pL2V-1-50506).



For the complementation experiments, *MtEFD*, *MtEFD2*, *MtRR4* and *AtERF003* wild type coding sequences were amplified with *MtEFD*, *MtEFD2*, *MtRR4*, *AtERF003* Forward and Reverse primers respectively (supplemental Table S5). To introduce P57/T57 point mutations in *MtEFD* and *MtEFD2*, fragments were amplified using *MtEFD* and *MtEFD2* mut Forward and Reverse (supplemental Table S5). For the chimeric *MtEFD*:*MtEFD2* and *MtEFD2*:*MtEFD* constructs (domain swaps), fragments were amplified with *MtEFD*:*MtEFD2* chim Forward and Reverse, and *MtEFD2*:*MtEFD* chim Forward and Reverse, respectively (supplemental Table S5). Level1 (p*MtEFD*:coding sequence: T35S) and Level2 assemblies were generated as previously described.

For CRISPR-Cas9 mutagenesis, guide RNAs (gRNA; listed in supplemental table 5) were designed with CRISPOR (version 4.8, <http://crispor.tefor.net/>) program, with the INRA A17r5.0 r1.6 *M. truncatula* genome release (Pecrix et al., 2018) and the "20bp-NGG-Sp Cas9, SpCas9-HF1, eSpCas9 1.1" option. The *S. pyogenes* Cas9 coding DNA sequence was used, with a SV40 NLS sequence added at the C-terminus. Four gRNAs, interspaced by tRNAs, were preassembled as a synthetic polycistronic gene, as described (Xie et al., 2015). They were expressed from *M. truncatula* U6.1 (MtrunA17\_Ch3g0136831) and U6.6 (MtrunA17\_Ch7g0251721) promoters (each for two guides). They were cloned using Golden Gate cloning technology into backbone plasmids provided by the ENSA project (Schiessl et al., 2019) (Engineering Nitrogen Symbiosis for Africa ; <https://www.ensa.ac.uk> ). T-DNAs included a kanamycin resistance module (p35S:KanR:TNos) and a DsRed fluorescent reporter module (pAtUbi10:DsRed:TOcs), located respectively close to the right and left borders of the T DNA.

Genotyping was performed by nested PCR and systematic sequencing of PCR products, using individual transformed root systems, on DNA from root segments adjacent to nodules.

## Accession numbers

GSE138899 (Nimblegen analyses); SRP349933 (RNAseq of CRISPR-Cas9 mutant and control samples); SRP349926 (RNAseq of non-transformed wt and *Mtefd-1* samples)

## Supplementary Data

**Supplemental Table S1.** Percentage of 16C, 32C and 64C nuclei in nodules of wt A17, *Mtefd-1* and other fix minus mutants.

**Supplemental Table S2.** Differentially expressed genes detected using Nimblegen gene chips.

**Supplemental Table S3.** RNAseq analysis of nitrogen-starved root tips, 4 dpi and 10 dpi nodules in wt A17 and *Mtefd-1* backgrounds. Tab 2: RNAseq data of the nine *MtRR4* genes with RNAseq reads. Tab 3: Venn diagrams showing the overlap between differentially regulated genes detected using

Nimblegen chips and RNAseq data. Tab 4-7: GO term enrichment analysis for genes down- and up-regulated in *Mtefd-1* vs wt 4 and 10 dpi nodules.

**Supplemental Table S4.** RNAseq analysis of 4 dpi nodulated root fragments of *Mtefd2* mutants generated by CRISPR-Cas9 in wt A17 and *Mtefd-1* backgrounds. Tab2: GO term enrichment analysis for genes downregulated in *efd2* mutants in both wt A7 and *Mtefd-1* mutant, while upregulated in *Mtefd-1* at 4 dpi.

**Supplemental Table S5.** Primers and guides

**Supplemental Figure S1.** Clustering analysis of RNAseq data produced using non-inoculated N-starved root tips and isolated *S. meliloti* 2011-induced nodules at 4 and 10 days post-inoculation, in wt (A17) and *Mtefd-1* backgrounds.

**Supplemental Figure S2.** Genes up- and down-regulated in wt nodules, and comparison of genes downregulated in the *Mtefd-1*, *Mtrsd-1*, and *Mtdnf-7* mutants.

**Supplemental Figure S3.** Time course analysis of *MtEFD*, *MtRR4* and *MtEFD2* expression.

**Supplemental Figure S4.** Expression pattern of the auxin responsive reporter DR5:*GUS*, as compared to the cytokinin-responsive reporter TCSn:*GUS* in wt and *Mtefd-1* nodules.

**Supplemental Figure S5.** Quantitative RT-PCR analysis of gene expression in *Mtefd-1* complementation assays

**Supplemental Figure S6.** Multiple alignment analysis of *MtEFD*- and *MtEFD2*-related proteins

**Supplemental Figure S7.** Phylogenetic tree of *MtEFD*-related proteins.

**Supplemental Figure S8.** Protein structure of AtERF003 (At5g25190), *MtEFD* and *MtEFD2* predicted by AlphaFold v2.0 (Jumper et al. 2021) and visualized using Mol\* Viewer (Sehna et al., 2021).

**Supplemental Figure S9.** Multiple alignment analysis (Multalin) of *MtEFD* and *PseFD* proteins from a collection of 39 pea accessions.

**Supplemental Figure S10.** Multi-guide CRISPR-Cas9-induced mutations in *MtEFD2*, obtained following *Agrobacterium rhizogenes* transformation.

**Supplemental Figure S11.** Macroscopic development of nodules in the *Mtefd-1* mutant and the *Mtefd-1 Mtefd2* double mutant.

## Funding information

This work was supported by the Agence Nationale de la Recherche (grant "GENOPEA" ANR-09-GENM-026), by INRAE (CJS fellowship for J.F.; SPE grant "EFFOR") and the "Laboratoire d'Excellence (LABEX) "TULIP" (ANR-10-LABX-41). The POPS platform benefited from the support of the LABEX Saclay Plant Sciences-SPS (ANR-10-LABX-0040-SPS).

## Acknowledgments

We are grateful to Peter Mergaert (ISV CNRS, Gif-sur-Yvette) for providing us with Fix- Medicago mutants, Ludovic Legrand (LIPME, Toulouse) for helping us with Alphafold, Rene Geurts and Eric Limpens for providing us with the *S. meliloti* *NifH::GFP* strain, and Jean-Marie Prospéri (INRA, Montpellier) for supplying *M. truncatula* A17 seeds. We thank the ENSA project for providing plasmids for Golden Gate cloning, and the FRAIB-TRI Imaging Platform (Université de Toulouse, CNRS) for its excellent support. RT-qPCR and RNAseq experiments were carried out at the Toulouse Genopole GeT-PlaGe platform.

## Tables

**Table 1.** The *Mtefd-1* mutation leads to downregulation of a set of genes expressed in immature nodules and early differentiation zone, as well as premature activation of a set of genes expressed in the late differentiation zone. Genes associated with nodule senescence and immune response are weakly affected in *Mtefd-1*. Gene expression patterns are from (Pecrix et al., 2018). FC = fold change.

Mt5.0 r1.8 ID	Gene name	Expression pattern	FC <i>Mtefd-1</i> vs wt 4 dpi nodules	FC <i>Mtefd-1</i> vs wt 10 dpi nodules	FC wt nodules 4 dpi vs roots	FC wt nodules 10 dpi vs roots
<b>Downregulated in 4 dpi <i>Mtefd-1</i> nodules</b>						
MtrunA17_Ch4g0002631	<i>MtEFD</i>	5	-314.6	-143.2	133.0	42.7
MtrunA17_Ch5g0414931	<i>MtRR4</i>	5	-22.7	-16.9	30.5	35.2
MtrunA17_Ch3g0092781	<i>MtLegin39</i>	4	-216.1	-15.6	211.1	13.8
MtrunA17_Ch3g0092931	<i>MtLegin41</i>	5	-644.8	-214.4	3967.5	762.9
MtrunA17_Ch7g0236361	<i>MtLP9</i>	5	-405.2	-500.3	822.6	440.7
MtrunA17_Ch7g0236381	<i>MtLP10</i>	6	-31.6	-8.3	330.9	495.3
MtrunA17_Ch6g0458644	<i>MtN1</i>	NA	-142.5	-57.0	221.0	49.7
MtrunA17_Ch4g0064831	<i>MtN15</i>	5	-69.0	-44.5	1813.2	521.6
MtrunA17_Ch7g0253571	<i>MtN20</i>	5	-25.0	-28.9	3949.9	2128.5
MtrunA17_Ch2g0314811	<i>MtNCR477</i>	5	-10.3	-17.9	272.1	112.1
MtrunA17_Ch8g0353641	<i>MtNCR730</i>	5	-91.4	-37.1	89.2	32.7
MtrunA17_Ch8g0367791	<i>MtNCR737</i>	6	-13.3	-23.7	523.2	359.8
MtrunA17_Ch5g0415651	<i>MtNCR797</i>	5	-26.4	-23.0	179.5	32.2
MtrunA17_Ch5g0438331	<i>MtNodGRP36</i>	4	-153.7	-196.6	1038.0	324.7
MtrunA17_Ch2g0317721	<i>MtTRX S1</i>	5	-65.7	-65.5	3455.8	512.6
<b>Upregulated in 4 dpi <i>Mtefd-1</i> nodules</b>						
MtrunA17_Ch7g0229931	<i>MtNCR169</i>	9	13.4	-9.4	2.4	4807.4
MtrunA17_Ch2g0301771	<i>MtNCR238</i>	8	36.6	-7.5	1.7	4755.6
MtrunA17_Ch5g0431111	<i>MtNCR447</i>	9	29.5	-9.3	1.1	1130.3
MtrunA17_Ch3g0100541	<i>MtCAML3</i>	10	15.2	-9.3	1.7	11715.6
MtrunA17_Ch1g0192121	<i>MtRbohA</i>	11	9.2	-3.7	4.2	161.0
MtrunA17_Ch5g0421331	<i>MtGH3</i>	15	4.5	1.4	0.7	1.8
MtrunA17_Ch8g0385011	<i>MtSAUR1</i>	1	2.9	2.1	0.5	0.5
MtrunA17_Ch4g0030001	<i>MtLBD11</i>	12	4.1	4.52	1.7	5.1
MtrunA17_Ch8g0369751	<i>MtCEP11</i>	0	10.0	4.42	0.3	0.1
<b>Weakly regulated in <i>Mtefd-1</i> nodules</b>						
MtrunA17_Ch5g0405841	<i>MtCP2</i>	0	1.2	-1.2	2.4	2.2
MtrunA17_Ch4g0040881	<i>MtCP6</i>	11	-1.4	1.5	0.4	1021.5
MtrunA17_Ch4g0044681	<i>MtDNF2</i>	15	-4.8	-5.8	479.1	1759.6
MtrunA17_Ch3g0119041	<i>MtSYMCRK</i>	10	-1.2	-4.5	278.1	756.5

**Table 2.** *AtERF003* but not *MtEFD2* can complement the nodule developmental phenotype of *Mtefd-1*, and the difference between MtEFD and MtEFD2 is in their C-terminal part, outside the N terminal DNA-binding domain

<i>Mtefd-1</i> transformed with:	Pink nodules	<sup>32</sup> C endoreduplication	<i>NifH:GFP</i> expression
empty vector	no (n=0/54)	no (n=19)	no (n=19)
<i>pMtEFD:MtEFD</i>	yes (n=53/61)	yes (n=19)	yes (n=11)
<i>pMtEFD:MtEFD2</i>	no (n=0/41)	no (n=11)	no (n=12)
<i>pMtEFD:N-MtEFD:C-MtEFD2</i>	no (n=0/39)	no (n=12)	no (n=17)
<i>pMtEFD:N-MtEFD2:C-MtEFD</i>	yes (n=27/35)	yes (n=10)	yes (n=13)
<i>pMtEFD:AtERF003</i>	yes (n=39/44)	yes (n=14)	yes (n=12)

**Table 3.** Examples of genes downregulated in *Mtefd2* mutant 4 dpi nodules. The three last columns correspond to CRISPR-Cas9 experiments in wt *M.truncatula* A17 or *Mtefd-1* background. *MtEFD2* correspond to non-edited *MtEFD2* gene, as determined by genotyping and sequencing. FC= fold change. Except where specified, all FC values are significant (FDR<0.01)

Gene Id (Mt5.0 r1.8)	Gene name	FC wt 4 dpi nodules vs root tips (non-transformed)	FC <i>Mtefd-1</i> vs wt 4 dpi nodules	FC <i>Mtefd-2</i> vs <i>MtEFD2</i> (wt A17 background)	FC <i>Mtefd-2</i> vs <i>MtEFD2</i> ( <i>Mtefd-1</i> background)
MtrunA17_Ch1g0177091	MtNF-YA1	1964.6	2.5	-4.3	-5.3
MtrunA17_Ch4g0067091	<i>MtNF-YB16</i>	240.5	2.3	-2.5	-4.1
MtrunA17_Ch2g0296321	<i>MtNF-YB18</i>	344.9	2.2	-2.2	-3.4
MtrunA17_Ch2g0301491	<i>MtNodGRP12</i>	588.1	3.2	-6.9	-33.8
MtrunA17_Ch4g0033101	<i>MtNodGRP23</i>	6427.3	2.8	-56.9	-45.6
MtrunA17_Ch3g0094011	<i>MtNodGRP4;MtN25</i>	2896.3	2.6	-41.1	-31.8
MtrunA17_Ch8g0386521	<i>MtSYMREM1</i>	3350.1	2.6	-19.6	-15.9
MtrunA17_Ch5g0422861	<i>MtNCR025</i>	1234.7	2.3	-23.4	-35.5
MtrunA17_Ch6g0474261	<i>MtNCR109</i>	1105.1	3.1	-8.6	-17.6
MtrunA17_Ch8g0372461	<i>MtSTY2</i>	35.8	-1.6 (FDR=0.99)	-2.4	-2.2
MtrunA17_Ch8g0353111	<i>MtSTY9</i>	51.6	-1.7 (FDR=0.99)	-3.3	-2.3
MtrunA17_Ch7g0262591	<i>MtYUC8</i>	1.9	-1.8 (FDR=0.06)	-1.8	-2.2

## Figure legends

**Figure 1.** *Mtefd-1* nodules exhibits an unusual plant cell endoreduplication phenotype, and a poor *S. meliloti nifH* expression level, without significant cell death.

A, B, Comparison of plant (A) and bacterial (B) cell endoreduplication in wt and *Mtefd-1* nodules (n=19).

C, D, E, F, Levels of plant cell endoreduplication in wt (C), TRV43 (D), *Mtefd-1* (E), and TR183 (F); each graph is representative of observations from two biological repeats (n=22, 9, 19, 16 respectively).

G, H, Expression of *S. meliloti pnifH::GFP* in wild-type (G) and *Mtefd-1* (H) nodules. DNA was stained with DAPI (4',6-diamidino-2-phenylindole; blue fluorescence). I, J Live-dead assay of wild-type (I) and *Mtefd-1* (J) nodules. Cells showing a cyan signal (SYTO™ 9) are alive while the magenta signal (propidium iodide) correspond to dead cells as well as plant nuclei (particularly abundant in the nodule meristem).

Asterisks indicate the nodule meristematic region. Nodules were analyzed at 21 days post-inoculation. Bars= 250 µm.

Plants were grown under aeroponic condition for A, B, G, H, I, J, and attapulgitte condition for C, D, E, F.

**Figure 2.** Expression pattern of *MtEFD* and other regulators of nodule development, and down-regulation by the *Mtefd-1* mutation of many genes expressed during late nodule development.

A, Top: 16 expression patterns defined in wt nodules (Pecrix et al., 2018), based on the average relative expression levels (in percentage of total expression) in five laser-dissected nodule zones, namely zone I (ZI), distal zone II (dZII), proximal zone II (pZII), interzone II-III (IZ) and fixation zone III (ZIII) (RNAseq data from Roux et al., 2014, mapped on the MtV5.0 genome release and reanalyzed). P1 to P11 are mostly expressed in a single zone, while P12 to P15 are expressed in all zones with a preferential one. A, Bottom: expression profile (with pattern number in parentheses) of *MtEFD* and its paralogue *MtEFD2*, as well as *MtRR4* and three TF genes thought to regulate nodule differentiation, shown in our final model (Fig. 6).

B, Distribution of all nodule-expressed genes amongst the 16 expression patterns (wild-type nodules).

C, Number of genes upregulated in mature (10 dpi) vs immature (4 dpi) wild-type nodules (fold change FC>4; FDR<0.01), per expression pattern.

D, Cumulative number of genes up- and down-regulated at least 2-fold in *Mtefd-1* vs wt 10 dpi nodules.

E, Number of genes down-regulated in *Mtefd-1* vs wild-type 10 dpi nodules (FC<-4; FDR<0.01), per expression pattern.

**Figure 3.** Cortical cell divisions and cytokinin signaling are modified in *Mtefd-1* *S.meliloti*-inoculated roots and nodules.

A-D, Expression pattern of the CK signaling reporter *TCSn:GUS* in young (stage III: A, B) and older (stage IV-V: C, D) nodule primordia, in wild-type (A, C) and *Mtefd-1* (B, D) backgrounds, at 4 dpi with *S. meliloti* *phemA::lacZ*, using *A. rhizogenes*-transformed plants (10 µm sections), following revelation of GUS and LacZ activity (blue and violet signal, respectively). Note the stronger GUS signal in *Mtefd-1*.

E, Section (10 µm) of *TCSn:GUS* *Mtefd-1* roots at 4 dpi following GUS staining. Note the presence of numerous cortical cell divisions (asterisks) within and between nodule primordia.

F-G, Expression of *TCSn:GUS* in mature wild-type (F) and *Mtefd-1* (G) nodules (50 µm sections)

H-I, Expression of *pWOX5:GUS* in mature wild-type (H) and *Mtefd-1* (I) nodules (50 µm sections)

F to I nodules were analyzed at 21 dpi. Bars =100 µm (A-E) or 250 µm (F-I).

**Figure 4.** The type A response regulator *MtRR4* complements the hypernodulation phenotype of *Mtefd-1* but not its nodule differentiation phenotype.

A-D, *Mtefd-1* complementation assays, using *S. meliloti* *NifH:GFP*-induced nodules and *Agrobacterium rhizogenes*-mediated root transformation.

A, wild-type mature nodules;

B, *Mtefd-1* nodules are efficiently complemented by *pMtEFD:MtEFD*;

C, *pMtEFD:MtRR4* fails to restore full nodule development and *NifH:GFP* expression in *Mtefd-1*;

D, *Mtefd-1* nodule transformed with an empty vector.

Blue and green fluorescence: DAPI-staining of DNA and GFP expression, respectively (50 µm sections). DsRed expression (transformation marker) was checked for all nodules.

E, Analysis by flow cytometry of the endoreduplication level of wt, *efd-1*, and *efd-1* transformed by *MtEFD* or *MtRR4*; *MtRR4* fails to restore the 32C level of plant cell endoreduplication.

F, Analysis of the average nodule number produced in wt, *efd-1*, and *efd-1* transformed by *MtEFD* or *MtRR4*; *pMtEFD:MtRR4* complements the hypernodulation phenotype of *Mtefd-1*. Statistical significance was analyzed by Kruskal-Wallis multiple comparison test. Samples with different letters (a and b) show a statistical difference for  $P \leq 0.001$ . Root fresh weight, measured after nodule removal. Center lines, box range, whiskers, black points and red points indicate the median, lower to upper quartile, 1.5× interquartile, outliers, and individual values respectively. All nodules were analyzed at 21 dpi. Bars = 250  $\mu$ m.

**Figure 5.** MtEFD2, a MtEFD paralogue with a distinct expression pattern, is unable to complement the *Mtefd-1* nodule phenotypes, in contrast to AtERF003, and positively regulates the nodule number in the *Mtefd-1* mutant background.

A, Multiple alignment analysis of MtEFD-like proteins in *M. truncatula* (MtEFD2; 79.1% similarity) and *A. thaliana* (AtERF003; 77.9% similarity). Amino acids involved in the direct interaction with DNA (Nakano et al., 2006) are indicated by an asterisk (AP2/ERF domain). CMV-1 and CMV-2 (= conserved motifs 1 and 2 of group V ERF proteins) are as previously defined (Cerri et al., 2015). Red and blue fonts: high and low consensus value (>90% and >50%, respectively).

B, C, Expression of *pMtEFD:GUS* (B) and *pMtEFD2:GUS* (C) in wild-type root tips (cleared roots)

D, E, Expression of *pMtEFD:GUS* (D) and *pMtEFD2:GUS* (E) in wild-type nodule primordia (10  $\mu$ m sections)

F, G, Expression of *pMtEFD:GUS* (F) and *pMtEFD2:GUS* (G) in wild-type mature nodules (50  $\mu$ m sections)

H, Expression of *pMtEFD2:GUS* in a *Mtefd-1* mutant roots inoculated with *S. meliloti* 2011 (10  $\mu$ m section). Asterisks indicate unorganized cortical cell divisions adjacent to nodule primordia.

B to H: observations following revelation of GUS activity.

I, Sections (50  $\mu$ m) of *S. meliloti* *NifH::GFP*-induced nodules (17 dpi) produced on wt *M. truncatula* A17 transformed with an empty vector, and *Mtefd-1* mutant transformed with an empty vector, *MtEFD2* and *AtERF003* expressed from *MtEFD* promoter; note the presence of differentiated bacteroids around a central vacuole in wt nodules and *pEFD:AtERF003* transformed nodules only. Blue and green fluorescence: DAPI-staining of DNA and GFP expression, respectively. DsRed expression (transformation marker) was checked for all nodules.

J, Average number of nodules produced at 21 dpi by *Mtefd-1* transformed by an empty vector, *AtERF003*, *MtEFD* and *MtEFD2* expressed from *MtEFD* promoter.

K, Average number of nodules produced at 17 dpi in the *Mtefd-1* mutant and the *Mtefd-1 MtEfd2* double mutant.

b:  $P$ -value<0.001 (Kruskal-Wallis test). Root fresh weight, measured after nodule removal.

Bars = 250  $\mu$ m (B, C, F, G, I) or 100  $\mu$ m (D, E, H).

*Agrobacterium rhizogenes*-mediated root transformation for B to K.

**Figure 6.** Model of MtEFD and MtEFD2 involvement in *M. truncatula* nodule development, integrated with other known regulators.

Inspired by Xiao et al., 2014 and Shresta et al., 2020 for primordium and nodule meristem formation, by Schiessl et al., 2019 and Soyano et al., 2019 regarding NIN and auxin biosynthesis, and by Sinharoy et al., 2013, Feng et al., 2021, Jiang et al., 2021 regarding MtRSD, NIN and NLP2 in nodule development and activity. In the primordium cells, MtEFD would be first involved in the



differentiation of nodule, following the Nod Factor signaling pathway leading to induction of root cortical cell divisions by cytokinins (CK) and auxin. MtEFD would also contribute to the transcriptional activation of MtRR4, a negative regulator of CK signaling, thereby contributing to negative feedback regulation of ccd and nodulation. In the nodule, MtEFD would be involved in the differentiation of cells generated by the nodule meristem, including late endoreduplication cycles (C16 to C32 and C64). Based on their respective expression profile (Fig. 2), MtEFD would act before two other regulators of nodule development, namely MtRSD and MtNLP2. *MtNIN* is only weakly downregulated in the *Mtefd-1* mutant at 10 dpi. MtEFD also contributes to the regulation of CK signaling in the nodule (Fig. 3G), possibly through MtRR4, but MtRR4 cannot complement the *Mtefd-1* differentiation nodule phenotype, whereas it complements the *Mtefd-1* hypernodulation phenotype. MtEFD2 is proposed to contribute to MtNF-YA1 expression and action, through its regulation cascade comprising *MtSTY* genes. NF = Nod Factors; CK = cytokinins; ccd = cortical cell divisions; P= Pericycle; E= endodermis; C3, C4, C5 = cortical cell layers 3, 4, 5 respectively.

## References

- Alexa A, Rahnenfuhrer J, Lengauer T** (2006) Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* **22**: 1600-1607
- Alunni B, Kevei Z, Redondo-Nieto M, Kondorosi A, Mergaert P, 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
- Andriankaja A, Boisson-Dernier A, Frances L, Sauviac L, Jauneau A, Barker DG, 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
- Anisimova M, Gascuel O** (2006) Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol* **55**: 539-552
- Barker DG, Pfaff T, Moreau D, Groves E, Ruffel S, Lepetit M, Whitehand S, Maillet F, Nair RM** (2007) Growing *Medicago truncatula*: choice of substrates and growth conditions. ISBN 0-9754303-1-9. <http://www.noble.org/MedicagoHandbook/>
- Baudin M, Laloum T, Lepage A, Ripodas C, Ariel F, Frances L, Crespi M, Gamas PC, Blanco FA, Zanetti ME, de Carvalho-Niebel F, Niebel A** (2015) A phylogenetically conserved group of NF-Y transcription factors interact to control nodulation in legumes. *Plant Physiol* **169**: 2761-2773
- Benjamini Y, Yekutieli D** (2001) The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics* **29**: 1165–1188
- Berrabah F, Ratet P, Gourion B** (2015) Multiple steps control immunity during the intracellular accommodation of rhizobia. *J Exp Bot* **66**: 1977-1985
- Boisson-Dernier A, Chabaud M, Garcia F, Becard G, Rosenberg C, Barker DG** (2001) *Agrobacterium rhizogenes*-transformed roots of *Medicago truncatula* for the study of nitrogen-fixing and endomycorrhizal symbiotic associations. *Mol Plant Microbe Interact* **14**: 695-700
- Bourcy M, Brocard L, Pislariu CI, Cosson V, Mergaert P, Tadege M, Mysore KS, Udvardi MK, Gourion B, 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
- Carrère S, Verdenaud M, Gough C, Gouzy J, Gamas P** (2019) LeGOO: An Expertized Knowledge Database for the Model Legume *Medicago truncatula*. *Plant and Cell Physiology* **61**: 203–211
- Carrere S, Verdier J, Gamas P** (2021) MtExpress, a Comprehensive and Curated RNAseq-based Gene Expression Atlas for the Model Legume *Medicago truncatula*. *Plant Cell Physiol* **62**: 1494-1500

- Cerri M, Gamas P, de Carvalho-Niebel F** (2015) AP2/ERF transcription factors and root nodulation. In Fd Bruijn, ed, **Biological Nitrogen Fixation**. Wiley/Blackwell, pp 609-622
- Cerri MR, Frances L, Kelner A, Fournier J, Middleton PH, Auriac MC, Mysore KS, Wen J, Erard M, Barker DG, Oldroyd GE, de Carvalho-Niebel F** (2016) The Symbiosis-Related ERN Transcription Factors Act in Concert to Coordinate Rhizobial Host Root Infection. *Plant Physiol* **171**: 1037-1054
- Couzigou JM, Zhukov V, Mondy S, Abu el Heba G, Cosson V, Ellis TH, Ambrose M, Wen J, Tadege M, Tikhonovich I, Mysore KS, Putterill J, Hofer J, Borisov AY, Ratet P** (2012) NODULE ROOT and COCHLEATA maintain nodule development and are legume orthologs of Arabidopsis BLADE-ON-PETIOLE genes. *Plant Cell* **24**: 4498-4510
- de Bang TC, Lundquist PK, Dai X, Boschiero C, Zhuang Z, Pant P, Torres-Jerez I, Roy S, Nogales J, Veerappan V, Dickstein R, Udvardi MK, Zhao PX, Scheible WR** (2017) Genome-Wide Identification of Medicago Peptides Involved in Macronutrient Responses and Nodulation. *Plant Physiol* **175**: 1669-1689
- Di Giacomo E, Laffont C, Sciarra F, Iannelli MA, Frugier F, 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
- Dong W, Zhu Y, Chang H, Wang C, Yang J, Shi J, Gao J, Yang W, Lan L, Wang Y, Zhang X, Dai H, Miao Y, Xu L, He Z, Song C, Wu S, Wang D, Yu N, Wang E** (2021) An SHR-SCR module specifies legume cortical cell fate to enable nodulation. *Nature* **589**: 586-590
- Farkas A, Maróti G, Durgó H, Györgypál Z, Lima RM, Medzihradszky KF, Kereszt A, Mergaert P, Kondorosi É** (2014) *Medicago truncatula* symbiotic peptide NCR247 contributes to bacteroid differentiation through multiple mechanisms. *Proc Natl Acad Sci U S A* **111**: 5183-5188
- Feng J, Lee T, Schiessl K, Oldroyd GED** (2021) Processing of NODULE INCEPTION controls the transition to nitrogen fixation in root nodules. *Science* **374**: 629-632
- Fliegmann J, Canova S, Lachaud C, Uhlenbroich S, Gascioli V, Pichereaux C, Rossignol M, Rosenberg C, Cumener M, Pitorre D, Lefebvre B, Gough C, Samain E, Fort S, Driguez H, Vauzeilles B, Beau JM, Nurisso A, Imbert A, Cullimore J, Bono JJ** (2013) Lipo-chitooligosaccharidic symbiotic signals are recognized by LysM receptor-like kinase LYR3 in the legume *Medicago truncatula*. *ACS Chem Biol* **8**: 1900-1906
- Franssen HJ, Xiao TT, Kulikova O, Wan X, Bisseling T, Scheres B, Heidstra R** (2015) Root developmental programs shape the *Medicago truncatula* nodule meristem. *Development* **142**: 2941-2950
- Gautrat P, Laffont C, Frugier F** (2020) Compact Root Architecture 2 Promotes Root Competence for Nodulation through the miR2111 Systemic Effector. *Curr Biol* **30**: 1339-1345
- Gavrin A, Kaiser BN, Geiger D, Tyerman SD, Wen Z, Bisseling T, Fedorova EE** (2014) Adjustment of host cells for accommodation of symbiotic bacteria: vacuole defunctionalization, HOPS suppression, and TIP1g retargeting in *Medicago*. *Plant Cell* **26**: 3809-3822
- Guefrachi I, Nagymihály M, Pislariu CI, Van de Velde W, Ratet P, Mars M, Udvardi MK, Kondorosi E, Mergaert P, Alunni B** (2014) Extreme specificity of NCR gene expression in *Medicago truncatula*. *BMC Genomics* **15**: 712
- Herrbach V, Chirinos X, Rengel D, Agbevenou K, Vincent R, Pateyron S, Huguet S, Balzergue S, Pasha A, Provart N, Gough C, Bensmihen S** (2017) Nod factors potentiate auxin signaling for transcriptional regulation and lateral root formation in *Medicago truncatula*. *J Exp Bot* **68**: 569-583
- Horstman A, Willemsen V, Boutilier K, Heidstra R** (2014) AINTEGUMENTA-LIKE proteins: hubs in a plethora of networks. *Trends Plant Sci* **19**: 146-157
- Horváth B, Domonkos Á, Kereszt A, Szűcs A, Ábrahám E, Ayaydin F, Bóka K, Chen Y, Chen R, Murray JD, Udvardi MK, Kondorosi É, Kaló P** (2015) Loss of the nodule-specific cysteine rich peptide, NCR169, abolishes symbiotic nitrogen fixation in the *Medicago truncatula* *dnf7* mutant. *Proc Natl Acad Sci U S A* **112**: 15232-15237

- Jardinaud MF, Boivin S, Rodde N, Catrice O, Kisiala A, Lepage A, Moreau S, Roux B, Cottret L, Sallet E, Brault M, Emery RJ, Gouzy J, Frugier F, Gamas P** (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
- Jiang S, Jardinaud MF, Gao J, Pecrix Y, Wen J, Mysore K, Xu P, Sanchez-Canizares C, Ruan Y, Li Q, Zhu M, Li F, Wang E, Poole PS, Gamas P, Murray JD** (2021) NIN-like protein transcription factors regulate leghemoglobin genes in legume nodules. *Science* **374**: 625-628
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D** (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**: 583-589
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS** (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* **14**: 587-589
- Katoh K, Toh H** (2008) Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* **9**: 286-298
- Kim M, Chen Y, Xi J, Waters C, Chen R, Wang D** (2015) An antimicrobial peptide essential for bacterial survival in the nitrogen-fixing symbiosis. *Proc Natl Acad Sci U S A* **112**: 15238-15243
- Laloum T, Baudin M, Frances L, Lepage A, Billault-Penneteau B, Cerri MR, Ariel F, Jardinaud MF, Gamas P, de Carvalho-Niebel F, Niebel A** (2014) Two CCAAT-box-binding transcription factors redundantly regulate early steps of the legume-rhizobia endosymbiosis. *Plant J* **79**: 757-768
- Laporte P, Satiat-Jeunemaitre B, Velasco I, Csorba T, Van de Velde W, Campalans A, Burgyan J, Arevalo-Rodriguez M, 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
- Lefebvre B, Timmers T, Mbengue M, Moreau S, Herve C, Toth K, Bittencourt-Silvestre J, Klaus D, Deslandes L, Godiard L, Murray JD, Udvardi MK, Raffaele S, Mongrand S, Cullimore J, Gamas P, Niebel A, Ott T** (2010) A remorin protein interacts with symbiotic receptors and regulates bacterial infection. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 2343-2348
- Letunic I, Bork P** (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* **49**: W293-W296
- Limpens E, Ivanov S, van Esse W, Voets G, Fedorova E, Bisseling T** (2009) Medicago N<sub>2</sub>-fixing symbiosomes acquire the endocytic identity marker Rab7 but delay the acquisition of vacuolar identity. *Plant Cell* **21**: 2811-2828
- Lin J, Frank M, Reid D** (2020) No Home without Hormones: How Plant Hormones Control Legume Nodule Organogenesis. *Plant Commun* **1**: 100104
- Liu J, Bisseling T** (2020) Evolution of NIN and NIN-like Genes in Relation to Nodule Symbiosis. *Genes (Basel)* **11**: 777
- Liu J, Miller SS, Graham M, Bucciarelli B, Catalano CM, Sherrier DJ, Samac DA, Ivashuta S, Fedorova M, Matsumoto P, Gantt JS, Vance CP** (2006) Recruitment of novel calcium-binding proteins for root nodule symbiosis in *Medicago truncatula*. *Plant Physiol* **141**: 167-177
- Liu J, Rasing M, Zeng T, Klein J, Kulikova O, Bisseling T** (2021) NIN is essential for development of symbiosomes, suppression of defence and premature senescence in *Medicago truncatula* nodules. *New Phytol* **230**: 290-303
- Lurin C, Andrés C, Aubourg S, Bellaoui M, Bitton F, Bruyère C, Caboche M, Debast C, Gualberto J, Hoffmann B, Lecharny A, Le Ret M, Martin-Magniette ML, Mireau H, Peeters N, Renou JP, Szurek B, Taconnat L, Small I** (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* **16**: 2089-2103

- Magne K, Couzigou JM, Schiessl K, Liu S, George J, Zhukov V, Sahl L, Boyer F, Iantcheva A, Mysore KS, Wen J, Citerne S, Oldroyd GED, Ratet P** (2018) MtNODULE ROOT1 and MtNODULE ROOT2 Are Essential for Indeterminate Nodule Identity. *Plant Physiol* **178**: 295-316
- Marino D, Andrio E, Danchin EG, Oger E, Gucciardo S, Lambert A, Puppo A, Pauly N** (2011) A *Medicago truncatula* NADPH oxidase is involved in symbiotic nodule functioning. *New Phytol* **189**: 580-592
- Marsh JF, Rakocevic A, Mitra RM, Brocard L, Sun J, Eschstruth A, Long SR, Schultze M, Ratet P, Oldroyd GE** (2007) *Medicago truncatula* NIN is essential for rhizobial-independent nodule organogenesis induced by autoactive calcium/calmodulin-dependent protein kinase. *Plant Physiol* **144**: 324-335
- Maunoury N, Redondo-Nieto M, Bourcy M, Van de Velde W, Alunni B, Laporte P, Durand P, Agier N, Marisa L, Vaubert D, Delacroix H, Duc G, Ratet P, Aggerbeck L, Kondorosi E, Mergaert P** (2010) Differentiation of Symbiotic Cells and Endosymbionts in *Medicago truncatula* Nodulation Are Coupled to Two Transcriptome-Switches. *PLoS One* **5**: e9519
- Mergaert P** (2018) Role of antimicrobial peptides in controlling symbiotic bacterial populations. *Nat Prod Rep* **35**: 336-356
- Mergaert P, Kereszt A, Kondorosi E** (2020) Gene Expression in Nitrogen-Fixing Symbiotic Nodule Cells in *Medicago truncatula* and Other Nodulating Plants. *Plant Cell* **32**: 42-68
- Middleton PH, Jakab J, Penmetsa RV, Starker CG, Doll J, Kalo P, Prabhu R, Marsh JF, Mitra RM, Kereszt A, Dudas B, Vandenbosch K, Long SR, Cook DR, Kiss GB, Oldroyd GE** (2007) An ERF Transcription Factor in *Medicago truncatula* That Is Essential for Nod Factor Signal Transduction. *Plant Cell* **19**: 1221-1234
- Montiel J, Downie JA, Farkas A, Bihari P, Herczeg R, Balint B, Mergaert P, Kereszt A, Kondorosi E** (2017) Morphotype of bacteroids in different legumes correlates with the number and type of symbiotic NCR peptides. *Proc Natl Acad Sci U S A* **114**: 5041-5046
- Moreau S, Fromentin J, Vailleau F, Vernié T, Huguet S, Balzergue S, Frugier F, Gamas P, Jardinaud MF** (2014) The symbiotic transcription factor MteFD and cytokinins are positively acting in the *Medicago truncatula* and *Ralstonia solanacearum* pathogenic interaction. *New Phytol* **201**: 1343-1357
- Nagyimihaly M, Veluchamy A, Gyorgypal Z, Ariel F, Jegu T, Benhamed M, Szucs A, Kereszt A, Mergaert P, Kondorosi E** (2017) Ploidy-dependent changes in the epigenome of symbiotic cells correlate with specific patterns of gene expression. *Proc Natl Acad Sci U S A* **114**: 4543-4548
- Nakano T, Suzuki K, Fujimura T, Shinshi H** (2006) Genome-Wide Analysis of the ERF Gene Family in Arabidopsis and Rice. *Plant Physiol* **140**: 411-432
- Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ** (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**: 268-274
- O'Malley RC, Huang SC, Song L, Lewsey MG, Bartlett A, Nery JR, Galli M, Gallavotti A, Ecker JR** (2016) Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. *Cell* **165**: 1280-1292
- Osipova MA, Mortier V, Demchenko KN, Tsyganov VE, Tikhonovich IA, Lutova LA, Dolgikh EA, Goormachtig S** (2012) Wuschel-related homeobox5 gene expression and interaction of CLE peptides with components of the systemic control add two pieces to the puzzle of autoregulation of nodulation. *Plant Physiol* **158**: 1329-1341
- Ovchinnikova E, Journet EP, Chabaud M, Cosson V, Ratet P, Duc G, Fedorova E, Liu W, den Camp RO, Zhukov V, Tikhonovich I, Borisov A, Bisseling T, Limpens E** (2011) IPD3 controls the formation of nitrogen-fixing symbiosomes in pea and *Medicago* Spp. *Mol Plant Microbe Interact* **24**: 1333-1344
- Pan H, Wang D** (2017) Nodule cysteine-rich peptides maintain a working balance during nitrogen-fixing symbiosis. *Nat Plants* **3**: 17048
- Pecrix Y, Staton SE, Sallet E, Lelandais-Briere C, Moreau S, Carrere S, Blein T, Jardinaud MF, Latrasse D, Zouine M, Zahm M, Kreplak J, Mayjonade B, Satge C, Perez M, Cauet S, Marande W,**

- Chantry-Darmon C, Lopez-Roques C, Bouchez O, Berard A, Debelle F, Munos S, Bendahmane A, Berges H, Niebel A, Buitink J, Frugier F, Benhamed M, Crespi M, Gouzy J, Gamas P** (2018) Whole-genome landscape of *Medicago truncatula* symbiotic genes. *Nat Plants* **4**: 1017-1025
- Pislariu CI, Sinharoy S, Torres-Jerez I, Nakashima J, Blancaflor EB, Udvardi MK** (2019) The Nodule-Specific PLAT Domain Protein NPD1 Is Required for Nitrogen-Fixing Symbiosis. *Plant Physiol* **180**: 1480-1497
- Plet J, Wasson A, Ariel F, Le Signor C, Baker D, Mathesius U, Crespi M, Frugier F** (2011) MtCRE1-dependent cytokinin signaling integrates bacterial and plant cues to coordinate symbiotic nodule organogenesis in *Medicago truncatula*. *Plant J* **65**: 622-633
- Ribeiro CW, Baldacci-Cresp F, Pierre O, Larousse M, Benyamina S, Lambert A, Hopkins J, Castella C, Cazareth J, Alloing G, Boncompagni E, Couturier J, Mergaert P, Gamas P, Rouhier N, Montrichard F, Frendo P** (2017) Regulation of Differentiation of Nitrogen-Fixing Bacteria by Microsymbiont Targeting of Plant Thioredoxin s1. *Curr Biol* **27**: 250-256
- Robinson MD, Oshlack A** (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11**: R25
- Robinson MD, Smyth GK** (2008) Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics* **9**: 321-332
- Roux B, Rodde N, Jardinaud MF, Timmers T, Sauviac L, Cottret L, Carrère S, Sallet E, Courcelle E, Moreau S, Debellé F, Capela D, de Carvalho-Niebel F, Gouzy J, Bruand C, Gamas P** (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
- Satgé C, Moreau S, Sallet E, Lefort G, Auriac MC, Remblière C, Cottret L, Gallardo K, Noirot C, Jardinaud MF, Gamas P** (2016) Reprogramming of DNA methylation is critical for nodule development in *Medicago truncatula*. *Nat Plants* **2**: 16166
- Saur IM, Oakes M, Djordjevic MA, Imin N** (2011) Crosstalk between the nodulation signaling pathway and the autoregulation of nodulation in *Medicago truncatula*. *New Phytol* **190**: 865-874
- Schaller GE, Street IH, Kieber JJ** (2014) Cytokinin and the cell cycle. *Curr Opin Plant Biol* **21**: 7-15
- Schauser L, Roussis A, Stiller J, Stougaard J** (1999) A plant regulator controlling development of symbiotic root nodules. *Nature* **402**: 191-195
- Schiessl K, Lilley JLS, Lee T, Tamvakis I, Kohlen W, Bailey PC, Thomas A, Luptak J, Ramakrishnan K, Carpenter MD, Mysore KS, Wen J, Ahnert S, Grieneisen VA, Oldroyd GED** (2019) NODULE INCEPTION Recruits the Lateral Root Developmental Program for Symbiotic Nodule Organogenesis in *Medicago truncatula*. *Curr Biol* **29**: 3657-3668 e5
- Sehnal D, Bittrich S, Deshpande M, Svobodova R, Berka K, Bazgier V, Velankar S, Burley SK, Koca J, Rose AS** (2021) Mol\* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Res* **49**: W431-W437
- Shrestha A, Zhong S, Therrien J, Huebert T, Sato S, Mun T, Andersen SU, Stougaard J, Lepage A, Niebel A, Ross L, Szczyglowski K** (2021) *Lotus japonicus* Nuclear Factor YA1, a nodule emergence stage-specific regulator of auxin signalling. *New Phytol* **229**: 1535-1552
- Sinharoy S, Torres-Jerez I, Bandyopadhyay K, Kereszt A, Pislariu CI, Nakashima J, Benedito VA, Kondorosi E, Udvardi MK** (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
- Siol M, Jacquin F, Chabert-Martinello M, Smykal P, Le Paslier MC, Aubert G, Burstin J** (2017) Patterns of Genetic Structure and Linkage Disequilibrium in a Large Collection of Pea Germplasm. *G3 (Bethesda)* **7**: 2461-2471
- Smyth GK** (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**: Article3
- Soyano T, Shimoda Y, Kawaguchi M, Hayashi M** (2019) A shared gene drives lateral root development and root nodule symbiosis pathways in *Lotus*. *Science* **366**: 1021-1023

- Stonoha-Arther C, 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
- Storey JD, Tibshirani R** (2003) Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* **100**: 9440-9445
- Tiwari SB, Belachew A, Ma SF, Young M, Ade J, Shen Y, Marion CM, Holtan HE, Bailey A, Stone JK, Edwards L, Wallace AD, Canales RD, Adam L, Ratcliffe OJ, Repetti PP** (2012) The EDLL motif: a potent plant transcriptional activation domain from AP2/ERF transcription factors. *Plant J* **70**: 855-865
- Trujillo DI, Silverstein KA, Young ND** (2014) Genomic characterization of the LEED..PEEDs, a gene family unique to the medicago lineage. *G3 (Bethesda)* **4**: 2003-2012
- Ulmasov T, Hagen G, Guilfoyle TJ** (1997) ARF1, a transcription factor that binds to auxin response elements. *Science* **276**: 1865-1868
- Van de Velde W, Zehirov G, Szatmari A, Debreczeny M, Ishihara H, Kevei Z, Farkas A, Mikulass K, Nagy A, Tiricz H, Satiat-Jeunemaitre B, Alunni B, Bourge M, Kucho K, Abe M, Kereszt A, Maroti G, Uchiumi T, Kondorosi E, Mergaert P** (2010) Plant peptides govern terminal differentiation of bacteria in symbiosis. *Science* **327**: 1122-1126
- Verdier J, Lalanne D, Pelletier S, Torres-Jerez I, Righetti K, Bandyopadhyay K, Leprince O, Chatelain E, Vu BL, Gouzy J, Gamas P, Udvardi MK, Buitink J** (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, Moreau S, de Billy F, Plet J, Combier JP, Rogers C, Oldroyd G, Frugier F, Niebel A, 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 JM, Fedorova E, Cebolla A, Kevei Z, Horvath G, Kelemen Z, Tarayre S, Roudier F, Mergaert P, Kondorosi A, Kondorosi E** (2003) Endoreduplication Mediated by the Anaphase-Promoting Complex Activator CCS52A Is Required for Symbiotic Cell Differentiation in *Medicago truncatula* Nodules. *Plant Cell* **15**: 2093-2105
- Wang C, Yu H, Luo L, Duan L, Cai L, He X, Wen J, Mysore KS, Li G, Xiao A, Duanmu D, Cao Y, Hong Z, Zhang Z** (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, Long S** (2010) A nodule-specific protein secretory pathway required for nitrogen-fixing symbiosis. *Science* **327**: 1126-1129
- Xiao TT, Schilderink S, Moling S, Deinum EE, Kondorosi E, Franssen H, Kulikova O, Niebel A, Bisseling T** (2014) Fate map of *Medicago truncatula* root nodules. *Development* **141**: 3517-3528
- Xie K, Minkenberg B, Yang Y** (2015) Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc Natl Acad Sci U S A* **112**: 3570-3575
- Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP** (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* **30**: e15
- Yu H, Xiao A, Dong R, Fan Y, Zhang X, Liu C, Wang C, Zhu H, Duanmu D, Cao Y, 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
- Zhang X, Han L, Wang Q, Zhang C, Yu Y, Tian J, Kong Z** (2019) The host actin cytoskeleton channels rhizobia release and facilitates symbiosome accommodation during nodulation in *Medicago truncatula*. *New Phytol* **221**: 1049-1059
- Zürcher E, Tavor-Deslex D, Lituiev D, Enkerli K, Tarr PT, Müller B** (2013) A robust and sensitive synthetic sensor to monitor the transcriptional output of the cytokinin signaling network in planta. *Plant Physiol* **161**: 1066-1075

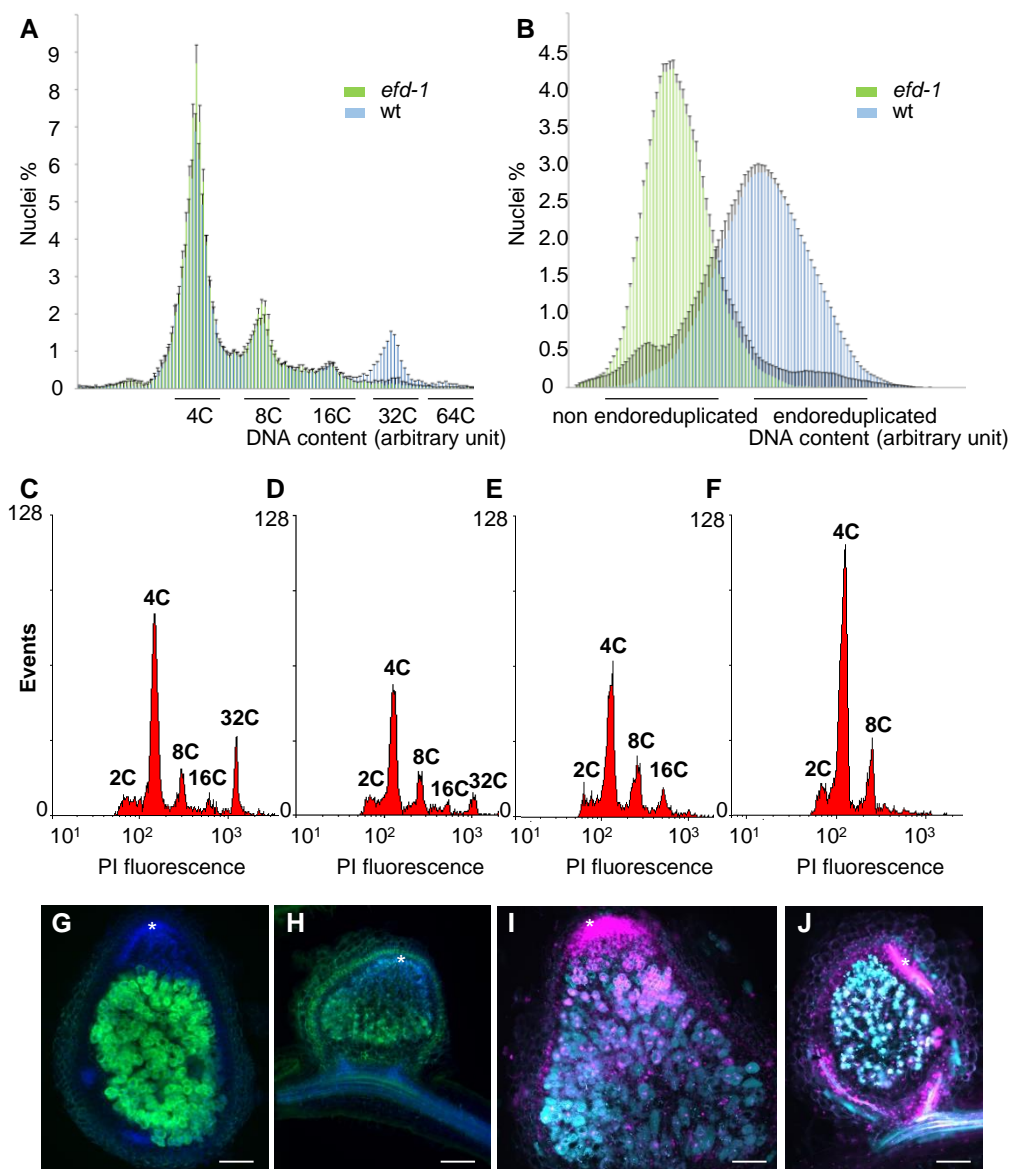


Figure 1

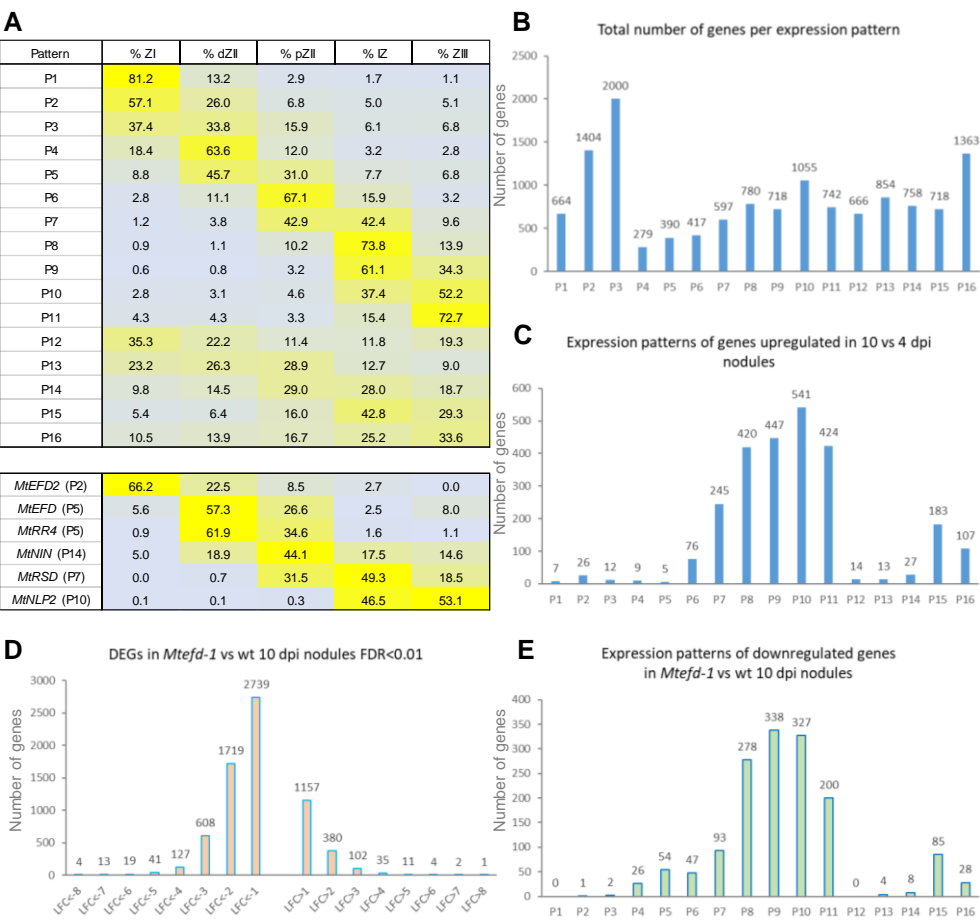


Figure 2



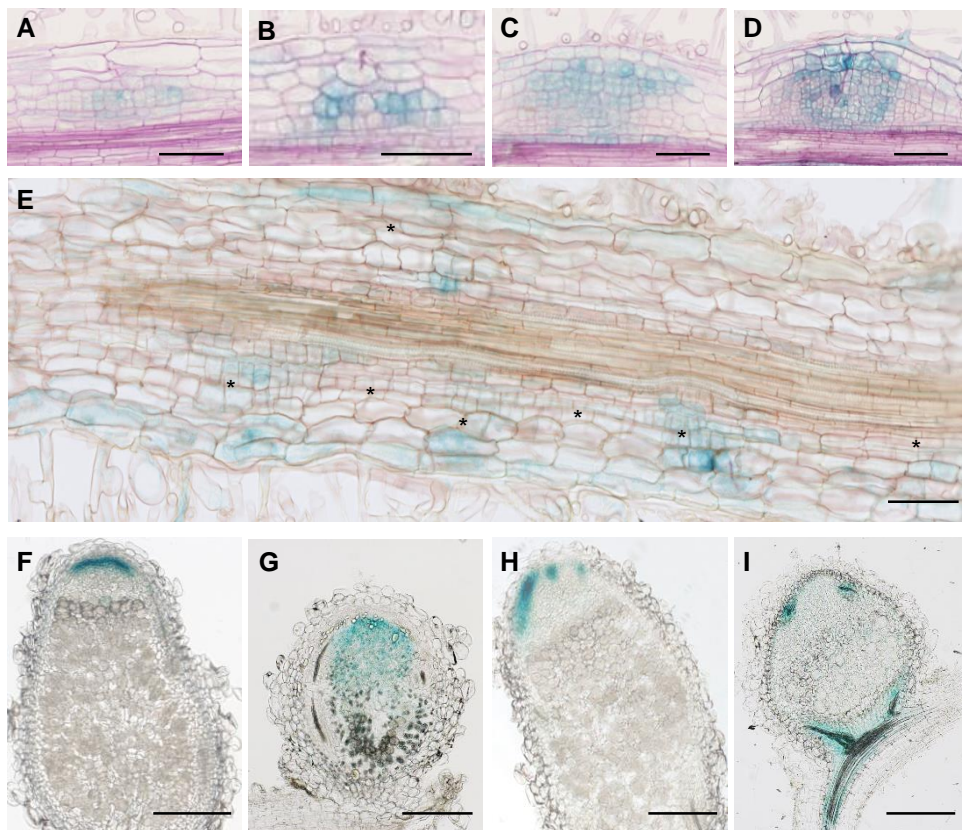


Figure 3

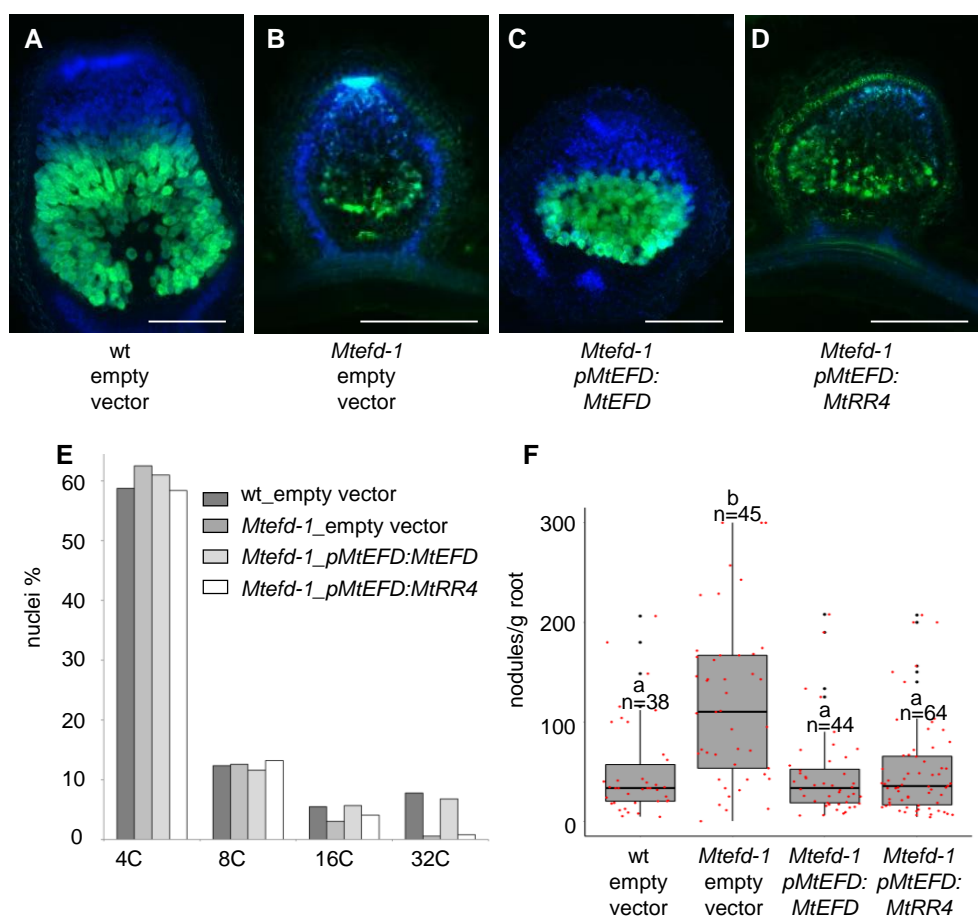


Figure 4

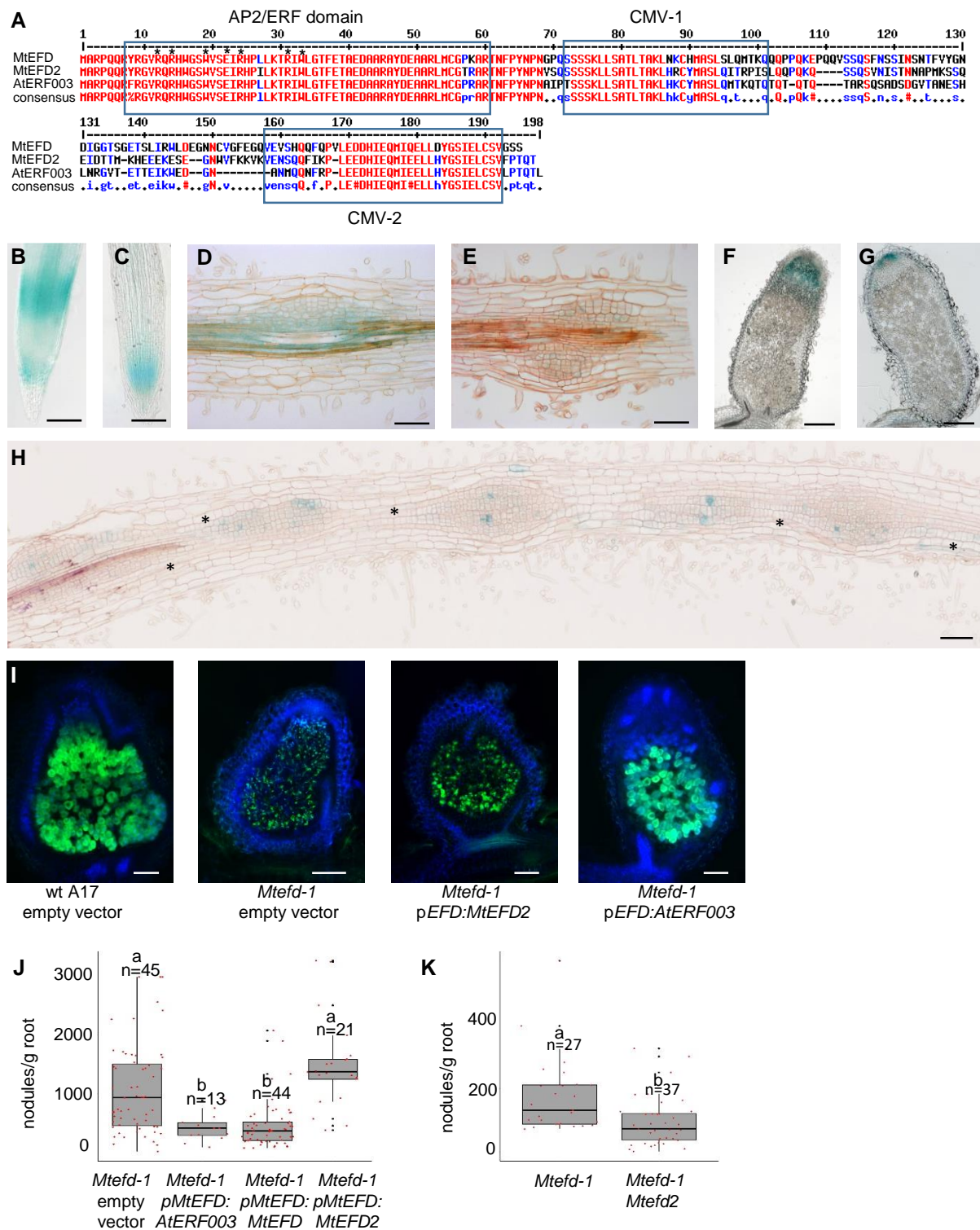


Figure 5



**Supplemental Table S1.** Percentage of 16C, 32C and 64C nuclei in nodules of wt A17, *Mtefd-1* and other fix minus mutants

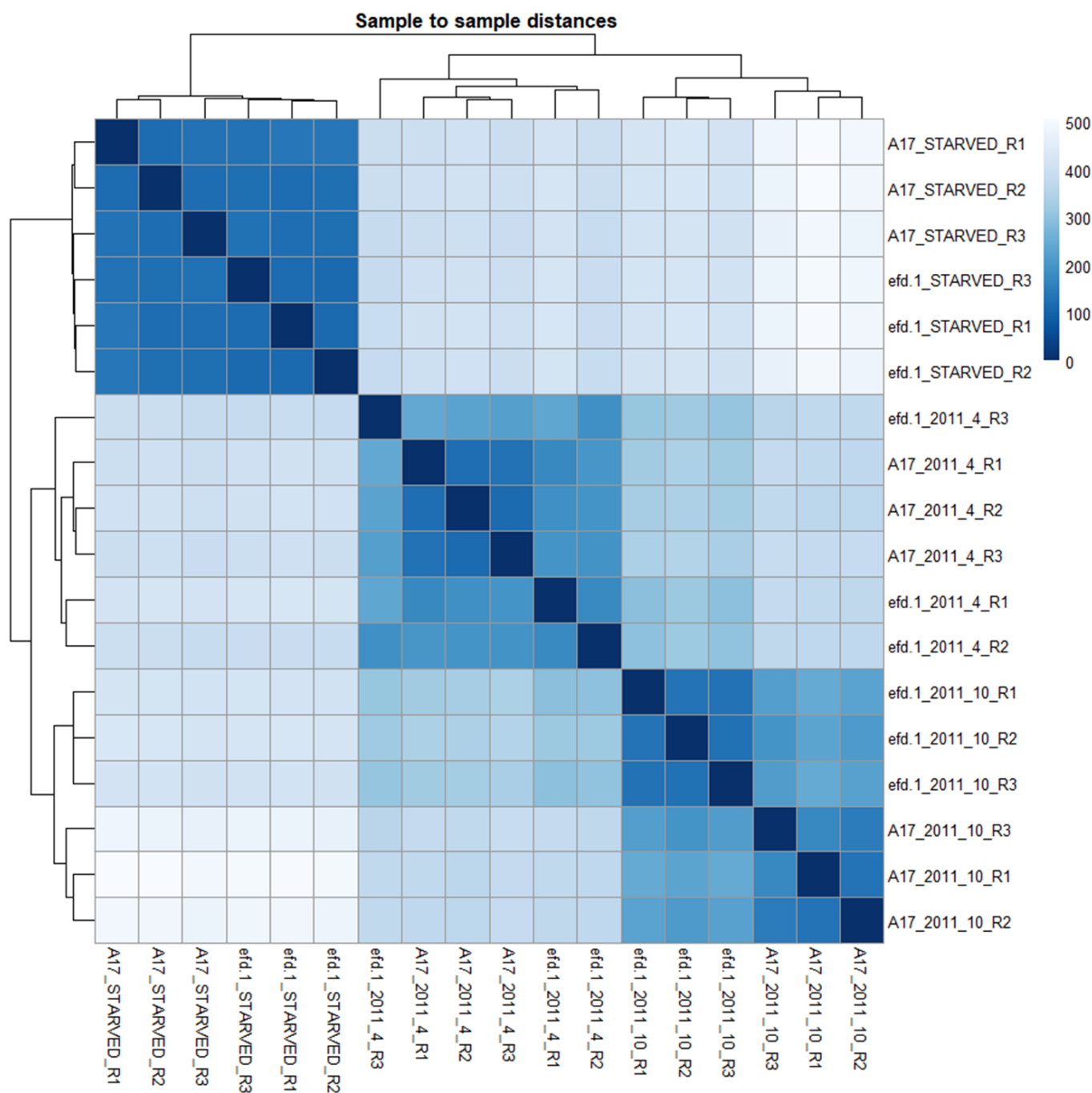
		Aeroponic growth condition		Attapulgit growth condition									
		14 dpi		13 dpi								30 dpi	
genotype		wt (n=19)	<i>Mtefd-1</i> (n=19)	wt (n=22)	TR36p (n=14)	TRV36 (n=14)	TRV43 (n=9)	<i>Mtefd-1</i> (n=19)	TR183 (n=16)	TR3 (n=20)	TR36w (n=11)	wt (n=22)	<i>Mtefd-1</i> (n=27)
DNA content	16C (%)	4.8 <sup>a</sup>	4.7 <sup>a</sup>	4.2 <sup>a</sup>	3.8 <sup>ab</sup>	2.7 <sup>bc</sup>	2.4 <sup>bcd</sup>	3.5 <sup>ab</sup>	1.7 <sup>d</sup>	2 <sup>cd</sup>	1.4 <sup>d</sup>	2.3 <sup>a</sup>	2 <sup>a</sup>
	32C (%)	9.9 <sup>a</sup>	1.2 <sup>b</sup>	9.4 <sup>a</sup>	7.9 <sup>a</sup>	2.9 <sup>b</sup>	3.1 <sup>b</sup>	0.8 <sup>c</sup>	1 <sup>c</sup>	0.5 <sup>c</sup>	0,5 <sup>c</sup>	10.5 <sup>a</sup>	0.6 <sup>b</sup>
	64C (%)	1.2 <sup>a</sup>	0.2 <sup>b</sup>	0.4 <sup>ab</sup>	0.5 <sup>a</sup>	0.2 <sup>bc</sup>	0.1 <sup>cd</sup>	0.1 <sup>d</sup>	0.1 <sup>d</sup>	0.1 <sup>d</sup>	0 <sup>d</sup>		
group				I	I	II	II	III	IV	IV	IV		

Values with different letters are significantly different (*P*-val<0.05; Anova Tukey post hoc test); distinct groups were identified in function of the percentage of nuclei in 16C, 32C and 64C. TR36 produced pink (TR36p) and white (TR36w) nodules with different levels of endoreduplication. Groups I, II, III and IV are as defined in Maunoury et al., 2010.

**Supplemental Table S5. Primers and guides**

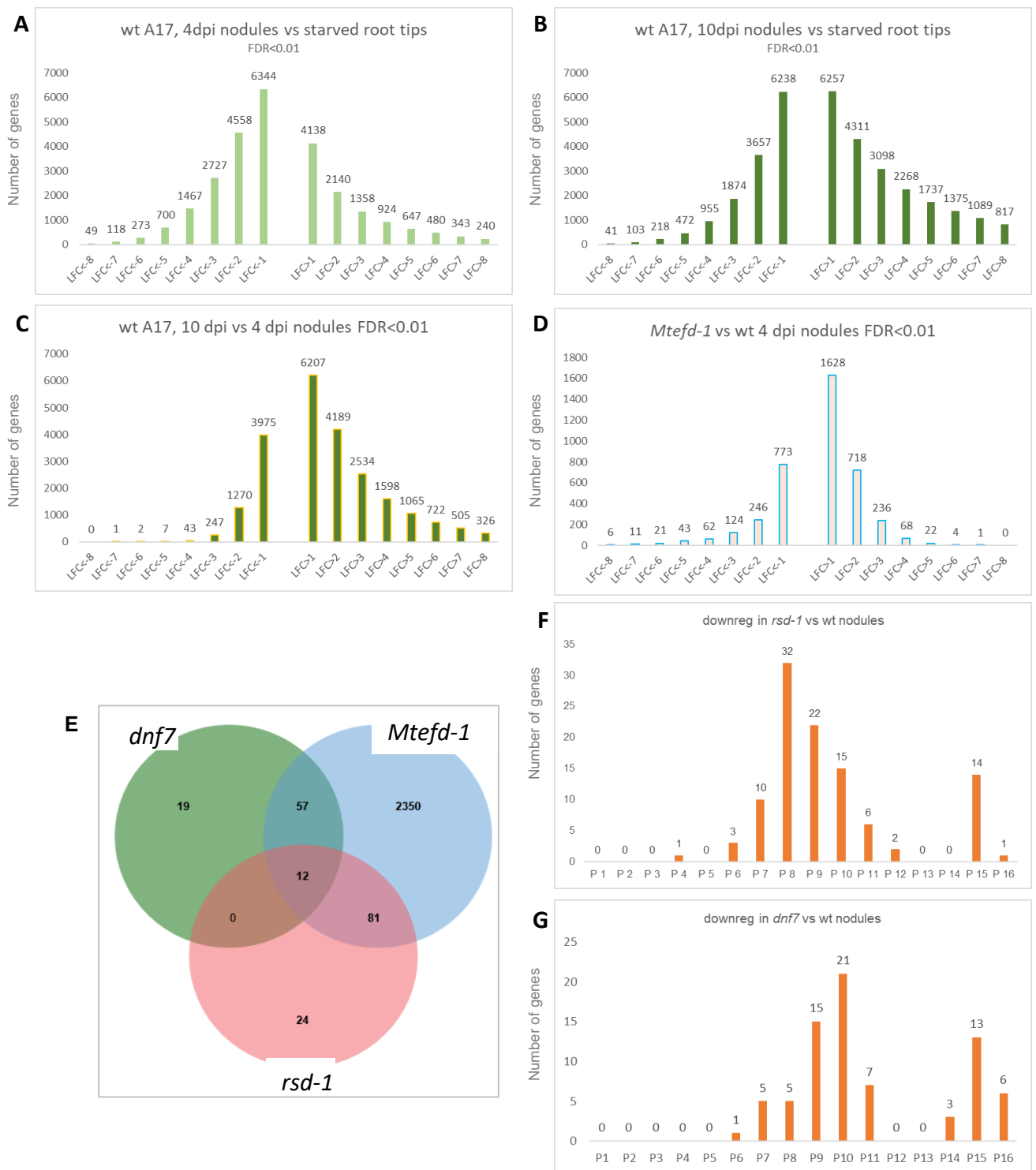
	FORWARD	REVERSE
<b>CLONING</b>		
<i>MtEFD</i> promoter	ggtctcaggagGGTACCCACCCGAACCC	ggtctcacattCAGCTTTGCACATATTTATAT
<i>MtEFD2</i> promoter	ggtctcaggagATAAGACTTAAACATATATATAACT	ggtctcacattACCTTGTITTTCTTTATGTAT
<i>MtWOX5</i> promoter	ggtctccaaatCTTAAATTCCTATCATCCAATTTTGG	ggtctccaaatATTTCATTCTAGATGTTTTACTCTGAAGTC
<i>MtEFD</i>	ggtctcaaagtCTTCACCTTCACCTTCACCTTAAGTCAC	ggtctcaaagcTAAGATGAACCAACAGACAAAGCTCAATTGAACC
<i>MtEFD2</i>	ggtctcaaagtATGGCTAGACCACAACAAGTTACCGAGGC	ggtctcaaagcGCTCTTTGACAAAACCCCTTACATTAC
<i>MtEFD</i> mut	ggtctccgga <b>A</b> CAAAAGCACGTACCAATTTCCCATACAACCCAAATGG	ggtctcc <b>T</b> TCCACACATTAATCTTGCTGCTTCATCATATGCTCTTGC
<i>MtEFD2</i> mut	ggtctccgga <b>C</b> CAAGGGCAAGAACAATTTCCCTTAC	ggtctcc <b>G</b> TCCACACATTAATCTTGCTGCTTCATCG
<i>MtEFD:MtEFD2</i> chim	ggtctcctgggAAATTGGTACGTGCTTTTGGTCCACACATT	ggtgtgggggaTACAACCCCAATGTGTCACAATCATCATCTT
<i>MtEFD2:MtEFD</i> chim	ggtctcccccTACAACCCCAATGGACCACAATCTTCTTCAT	ggtctccagggaAAATTGTCTTGCCCTTGTTCACACA
<i>AtERF003</i> (At5G25190)	ggtctcaaagtAGGTCTCGATTGCTCTTCACTCTCATCATCTTTC	ggtctcaaagcAGGTCTCGCGTAATACATAACATAGCATATGATC
<i>MtRR4</i>	ggtctcaaagtATGGAGACAAACAGTGTGT	ggtctcaaagcTTCGGTGCCGGTCATTTTAAG
<b>NF9808 analysis</b>		
<i>MtEFD2</i> primers	GACCACAACAACGTTACC	GAGGACACTAGCTGAATG
<b>CRISPR GUIDES</b>		
mg <i>MtEFD2a</i> -F	TAGGTCTCGACAAGGCATTGTTTTAGAGCTAGAA	CGGGTCTCTTTGTGCGACGCCTGCACCAGCCGGG
mg <i>MtEFD2b</i> -F	TAGGTCTCACACGTTACCGGTTTTAGAGCTAGAA	CGGGTCTCCGTGTTGTGGTCTTGACCAGCCGGG
mg <i>MtEFD2c</i> -F	TAGGTCTCCTAGAATATGGCTGTTTTAGAGCTAGAA	CGGGTCTCTTCTAGTTTTCTTGACCAGCCGGG
mg <i>MtEFD2d</i> -F	TAGGTCTCATGTGTGGAACAAGTTTAGAGCTAGAA	CGGGTCTCACACATTAACTTTGACCAGCCGGG
<b><i>MtEFD2</i> genotyping</b>		
<i>MtEFD2</i>	TCTACAAC TAGGAATGCAAAATGTC	CCCCAACTTGCAATTGACCCAA
<i>MtEFD2</i> nested	AAATCCTTTGTGGGACCCTC	GAGGACACTAGCTGAATG
<b>RT-qPCR</b>		
<i>MtEFD</i>	TTCATCATCATGGCAAGACCAC	AGAGACCCAAGAACCCCAATG
<i>MtRR4</i>	TTGTCTTCACCGACGTTCTTG	TCATTTCGGTGCCGGTCATT
<i>MtEFD2</i>	TGTGCATGTTTCTTGCACT	GACAAAACCTCCCTTACATTACAGAC
<i>MtUbiquitin</i> <i>MttrunA17</i> _Chr8g0342711	TTGTGTGTTGAATCCTAAGCA	CAAGACCCATGCAACAAGTTC
<i>Translation elongation factor</i> <i>MttrunA17</i> _Chr6g0458091	GACAAGCGTGTGATCGAGAGATT	TTTCACGCTCAGCCTTAAGCT
<b><i>EFD</i> analysis in <i>Pea</i> accessions</b>		
<i>PsEF-1</i>	ATGCACAAAGCTGCTTCACCTTCAC	GCTTAGATGAAGAAGATTGTGGTCC
<i>PsEF-2</i>	TTGAAACAGCAGAAGATGCAGCAAG	GTCTCAGACGAATTATCAGAAAACC

Lower case letters indicate the *BsaI* restriction site and spacer used for Golden Gate cloning; letters in bold correspond to point mutations



**Supplemental Figure S1.** Clustering analysis of RNAseq data produced using non-inoculated N-starved root tips and isolated *S.meliloti* 2011-induced nodules at 4 and 10 days post-inoculation, in wt (A17) and *Mtefd-1* backgrounds.

Three biological replicates (R1, R2, R3) per sample. Color scale : euclidian distances between samples.



**Supplemental Figure S2.** Genes up- and down-regulated in wt nodules, and comparison of genes downregulated in the *Mtefd-1*, *rsd-1* and *dnf-7* mutants.

A, B, C, range of up- and down-regulation in wt nodules (cumulative values)

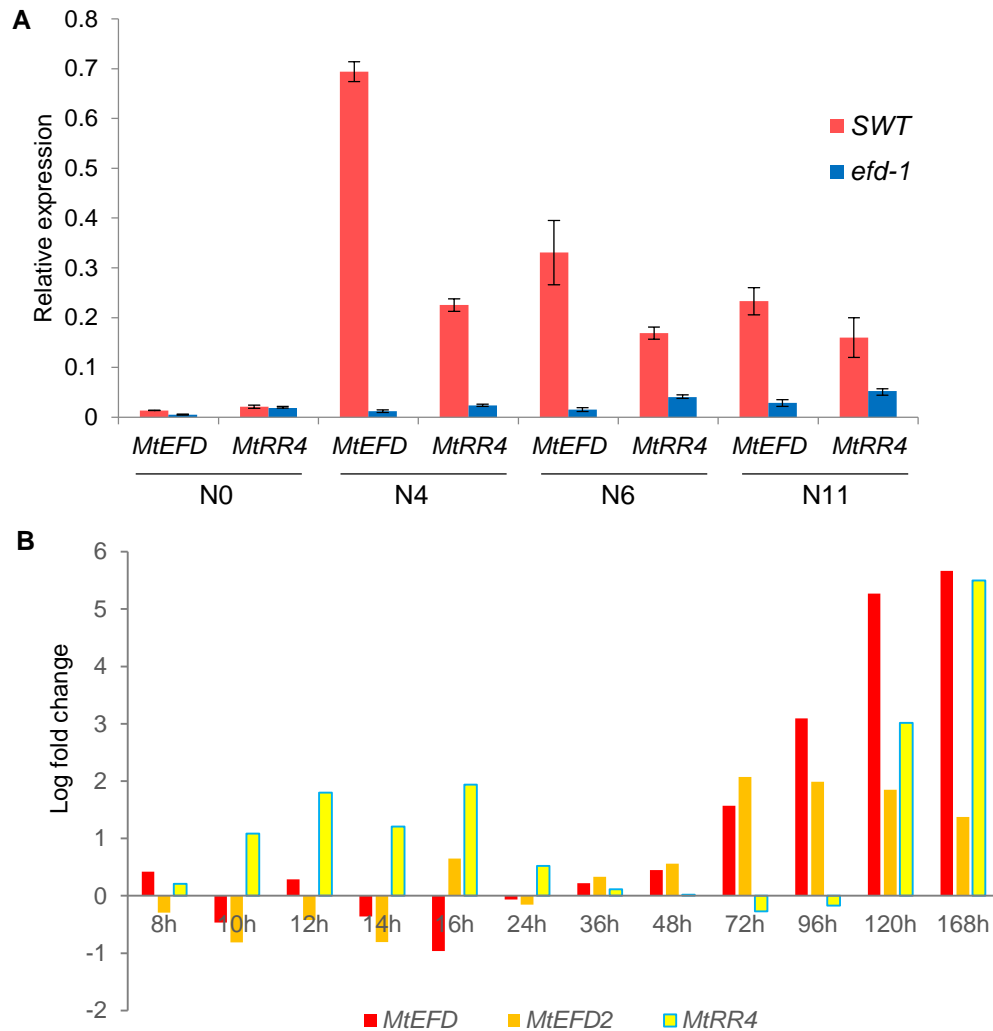
D, range of up- and down-regulation in *Mtefd-1* nodules 4 dpi (cumulative values)

E, Venn diagram showing the overlap between genes downregulated in the *Mtefd-1* (10 dpi), *rsd-1* (8 dpi) and *dnf7* (15 dpi) mutant nodules ( $\geq 2$ -fold downregulation; based on available Mt5.1.8 corresponding loci)

E, Venn diagram showing the overlap between genes downregulated in the *Mtefd-1* (10 dpi), *rsd-1* (8 dpi) and *dnf7* (15 dpi) mutant nodules ( $\geq 2$ -fold downregulation; based on available Mt5.1.8 corresponding loci)

F, G: Distribution of down-regulated genes in *rsd-1* and *dnf7* mutant nodules, amongst the 16 expression patterns defined in Pecrix et al. 2018 (see Fig. 2A)



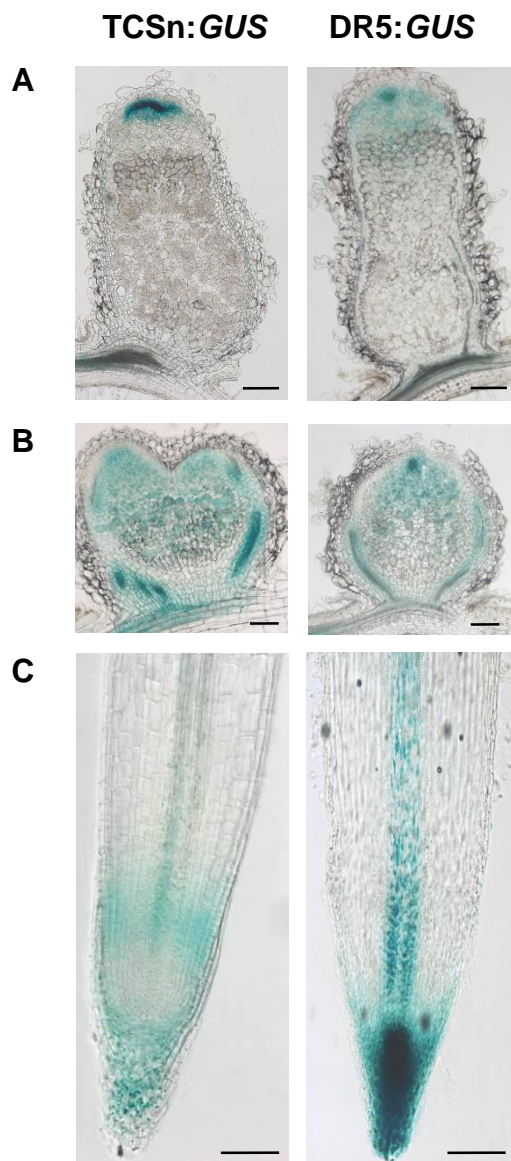


**Supplemental Figure S3.** Time course analysis of *MtEFD*, *MtRR4* and *MtEFD2* expression.

A, Q-RT-PCR analysis of expression in wild-type siblings (SWT) or *Mtefd-1* mutant.

N0 = non-inoculated roots; N4, N6, N11= isolated nodules at 4, 6 and 11 days post-inoculation (dpi), Respectively. Mean of four biological replicates, with 20 plants per replicate. Bars =  $\pm$  SE.

B, RNAseq analysis of spot-inoculated roots vs mock treatment; data extracted from (Schiessl et al., 2019)



**Supplemental Figure S4.** Expression pattern of the auxin-responsive reporter DR5:*GUS*, as compared to the cytokinin-responsive reporter TCSn:*GUS* in wt and *Mtefd-1* nodules.

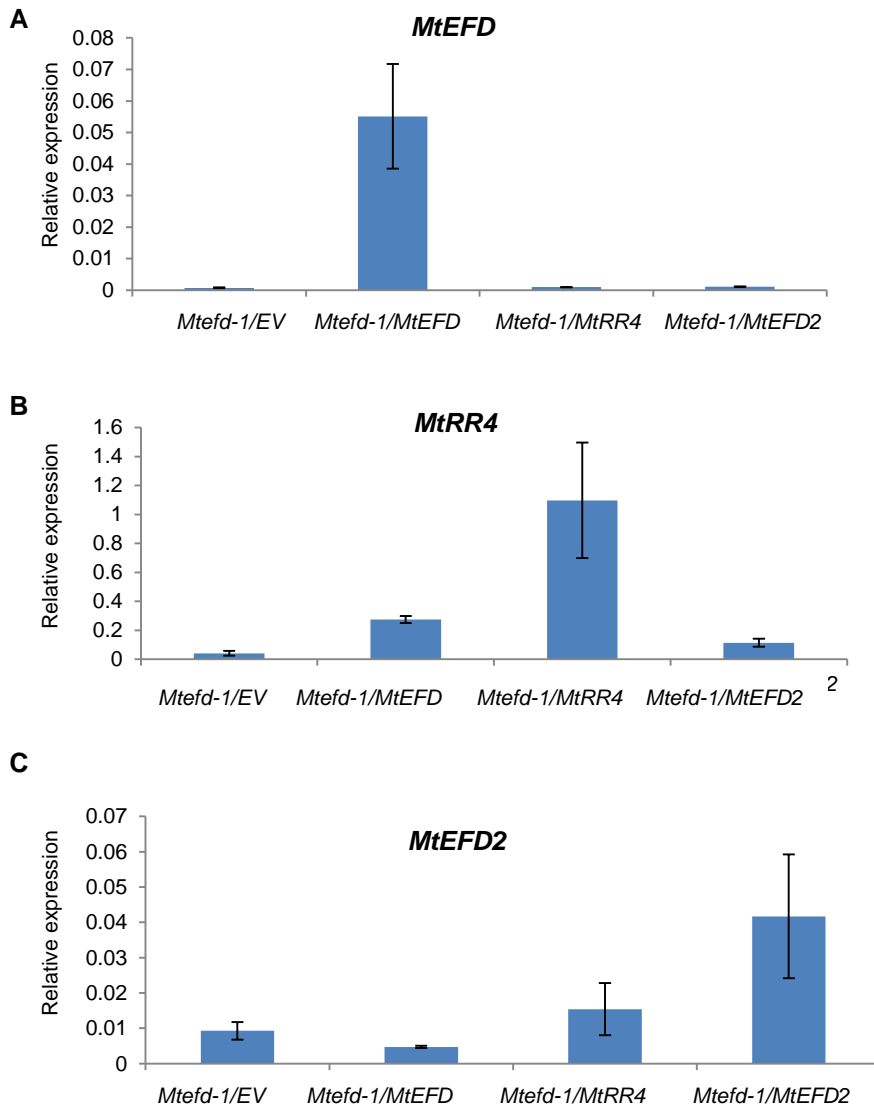
A, wild-type nodules

B, *Mtefd-1* mutant nodules

C, wild-type root tips, showing the expected pattern of TCSn:*GUS* and DR5:*GUS* expression (maximal in the transition and division zones, respectively)

Nodules harvested at 17 days post inoculation with *S.meliloti* 2011.

Scale bars = 250  $\mu$ m (A) or 100  $\mu$ m (B, C).



**Supplemental Figure S5.** Quantitative RT-PCR analysis of gene expression in *Mtefd-1* complementation assays. Relative expression of *MtEFD* (A), *MtRR4* (B) and *MtEFD2* (C) in *Mtefd-1* nodules transformed with an empty vector (EV) or a vector containing the *MtEFD* promoter fused to the *MtEFD*, *MtRR4* or *MtEFD2* coding region. RNA from 40 dpi nodules harvested from 8 (EV, *MtEFD*), 11 (*MtRR4*) and 13 (*MtEFD2*) independently transformed roots (2 biological replicates). Transcript levels were normalized using the mean of two housekeeping genes. Error bars =  $\pm$  SE.

# AP2/ERF domain

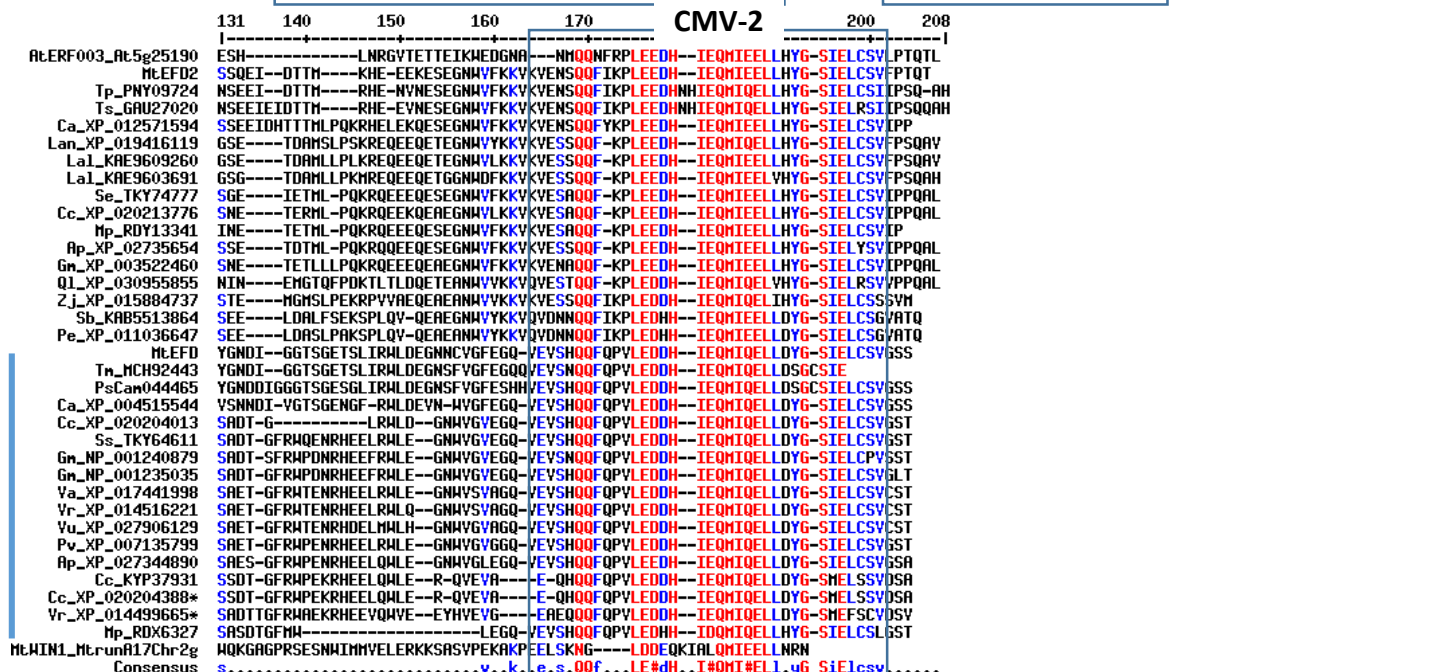
# CMV-1

## MtEFD2 subgroup

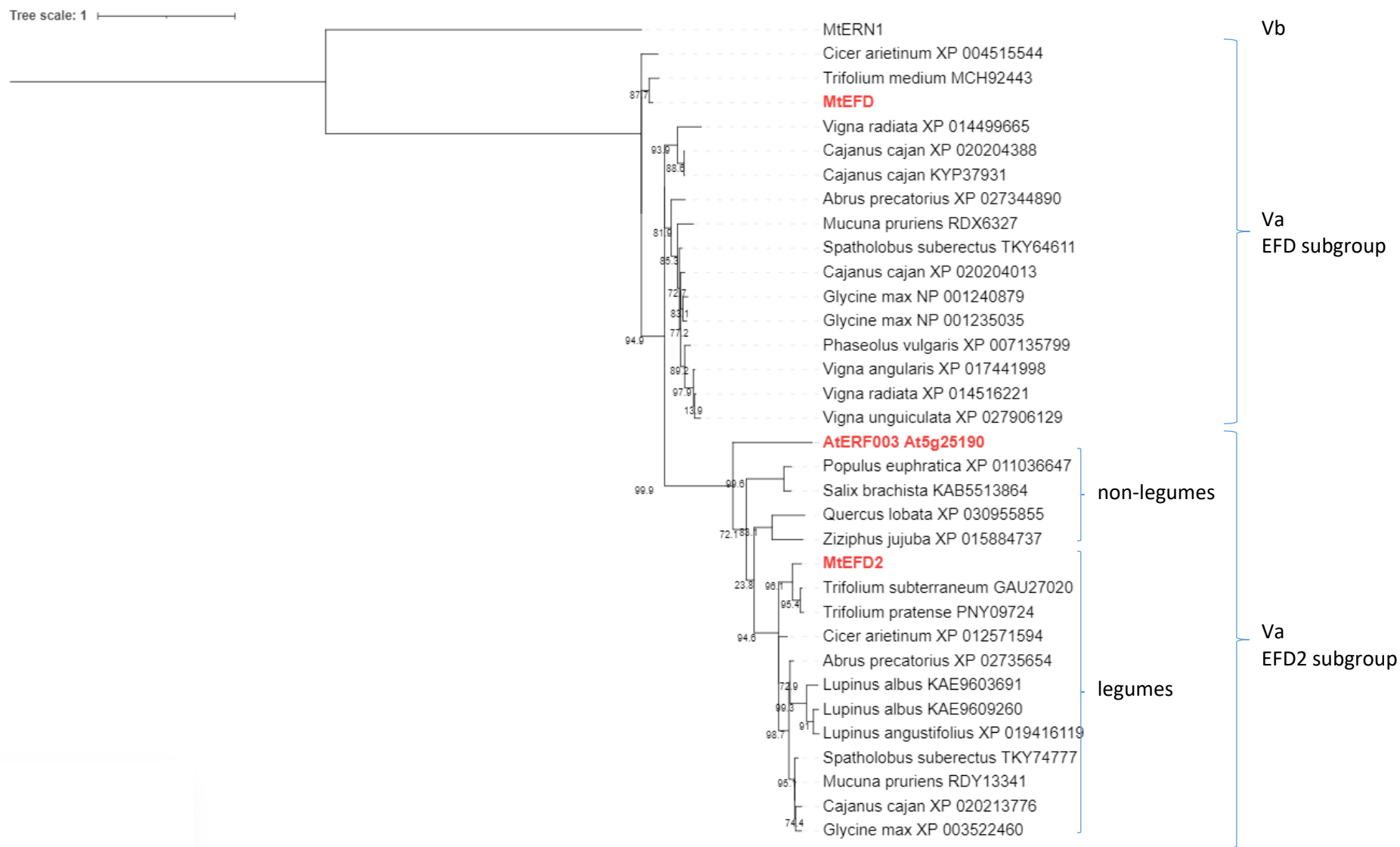
## MtEFD subgroup

## MtEFD2 subgroup

## MtEFD subgroup



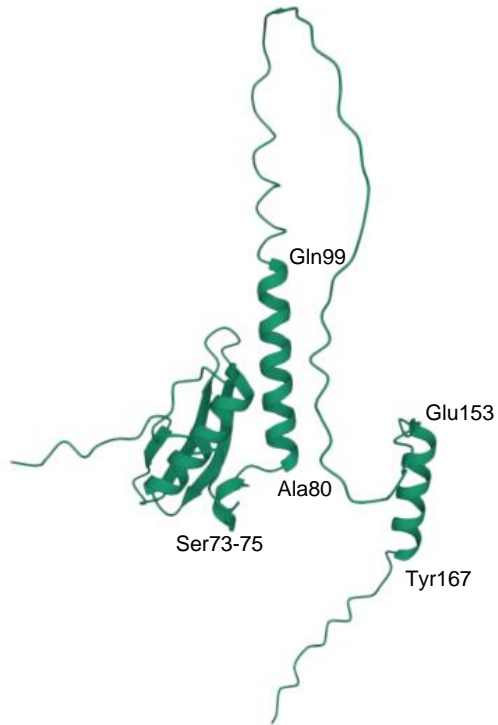
**Supplemental Figure S6.** Multiple alignment analysis of MtEFD- and MtEFD2-related proteins (performed with Multalin). Red and blue fonts: high and low consensus value (>90% and >50%, respectively).



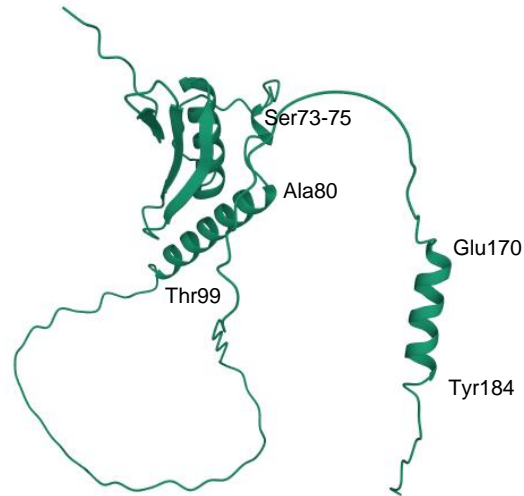
**Supplemental Figure S7.** Phylogenetic tree of MtEFD-related proteins.

The tree was generated using the protein fraction following the AP2/ERF DNA-binding domain, with MtERN1 (ERF group Vb) as an outgroup, AtERF003 and the 15 best blastp hits for MtEFD and MtEFD2 (both belonging to ERFgroup Va) in the nr protein database (ncbi blastp as of 20200204; default parameters).

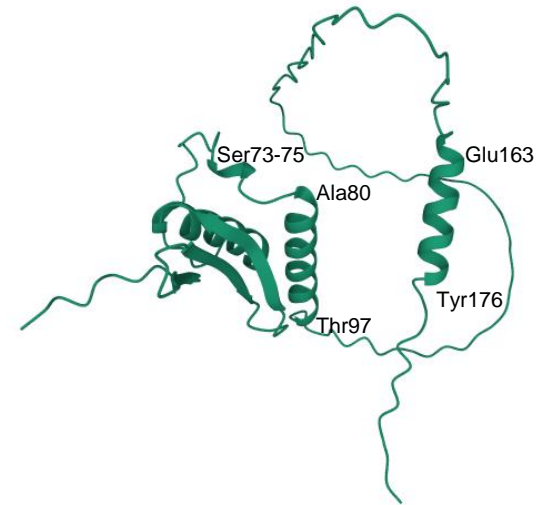
AtERF003



MtEFD

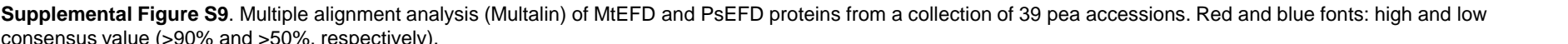


MtEFD2

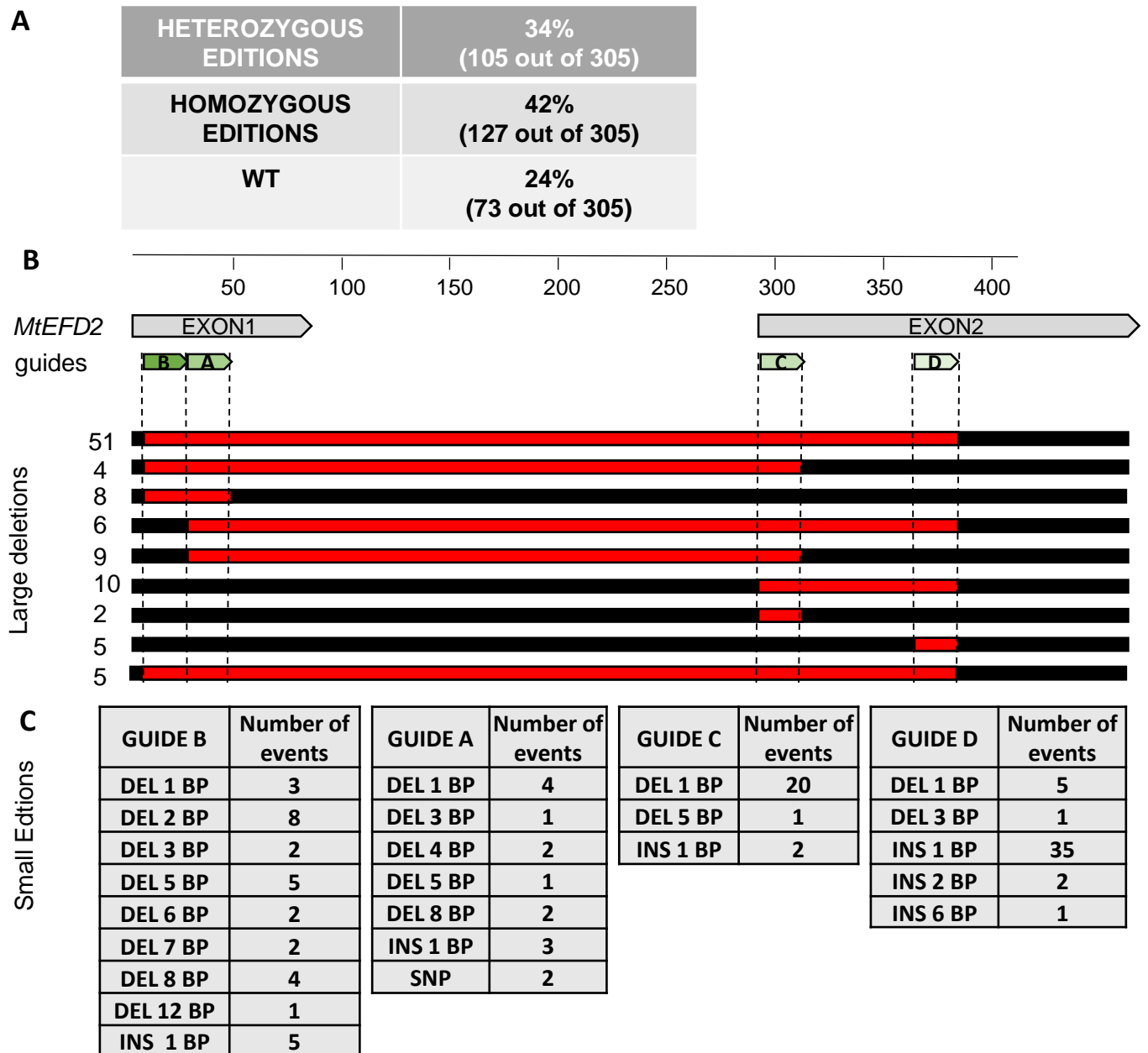


**Supplemental Figure S8.** Protein structure of AtERF003 (At5g25190), MtEFD and MtEFD2 predicted by AlphaFold v2.0 (Jumper et al. 2021) and visualized using Mol\* Viewer (Sehna et al., 2021).

The AP2/ERF domain comprises three  $\beta$ -sheets and two  $\alpha$ -helices, as indicated in Nakano et al. 2006; the conserved CMV-1 comprises a small  $\beta$ -sheet and one  $\alpha$ -helix; the CMV-2 motifs comprises one  $\alpha$ -helix. Only the positions of the predicted structural motifs in CMV-1 and CMV-2 are shown.







**Supplemental Figure S10.** Multi-guide CRISPR-Cas9-induced mutations in *MtEFD2*, obtained following *Agrobacterium rhizogenes* transformation. Data obtained by genotyping (by nested genomic PCR and amplicon sequencing) of 309 randomly picked transformed roots.

A. Frequency of heterozygous and homozygous editions

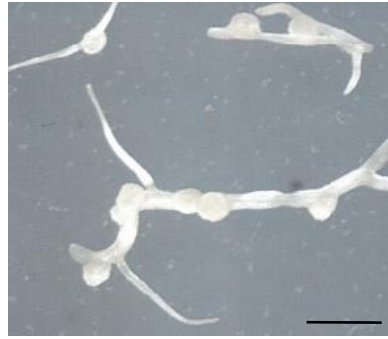
B. Number and position (in red) of the large homozygous deletions obtained, with regard to the position of the four guides used for *MtEFD2*

C. Number of small homozygous editions obtained for each guide. DEL= deletions; INS = insertions; SNP = single nucleotide polymorphism





*Mtefd-1*



*Mtefd-1 Mtefd2*

**Supplemental Figure S11.** Macroscopic development of nodules in the *Mtefd-1* mutant and the *Mtefd-1 Mtefd2* double mutant. Nodules harvested at 17 days post inoculation with *S.melliloti* 2011. Scale bar = 2mm