Title: A novel *cis*-element enabled bacterial uptake by plant cells

Authors:

Chloé Cathebras[1]†, Xiaoyun Gong[1]†, Rosa Elena Andrade[1], Ksenia Vondenhoff[1], Jean Keller[2], Pierre-Marc Delaux[2], Makoto Hayashi[3], Maximilian Griesmann[1], Martin Parniske[1]*

Affiliations:

[1]Faculty of Biology, Genetics, LMU Munich, Großhaderner Strasse 2-4, 82152 Martinsried, Germany

[2]Laboratoire de Recherche en Sciences Végétales, Université de Toulouse, CNRS, UPS, 24 chemin de Borde Rouge, Auzeville, BP42617, 31326 Castanet Tolosan, France

[3]Center for Sustainable Resource Science, RIKEN, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama City, Kanagawa, 230-0045, Japan

† these authors contribute equally to this work

*Correspondence to: Martin Parniske; parniske@lmu.de

Abstract:

The root nodule symbiosis (RNS) of plants with nitrogen-fixing bacteria is phylogenetically restricted to a single clade of flowering plants, which calls for yet unidentified trait acquisitions and genetic changes in the last common ancestor. Here we discovered – within the promoter of the transcription factor gene *Nodule Inception* (*NIN*) – a *cis*-regulatory element (*PACE*), exclusively present in members of this clade. *PACE* was essential for restoring infection threads (ITs) in *nin* mutants of the legume *Lotus japonicus*. *PACE* sequence variants from RNS-competent species appeared functionally equivalent. Evolutionary loss or mutation of *PACE* is associated with loss of this symbiosis. During early stages of nodule development, *PACE* dictates gene expression in a spatially restricted domain containing cortical cells carrying ITs. Consistent with its expression domain, *PACE*-driven *NIN* expression restored the formation of cortical ITs, also when engineered into the *NIN* promoter of tomato. Our data pinpoint *PACE* as a key evolutionary invention that connected *NIN* to a pre-existing symbiosis signal transduction cascade that governs the intracellular accommodation of arbuscular mycorrhiza fungi and is conserved throughout land plants. This connection enabled bacterial uptake into plant cells via intracellular support structures like ITs, a unique and unifying feature of this symbiosis.

Introduction

Nitrogen is essential for plant growth and development (LeBauer & Treseder, 2008). A wide phylogenetic variety of land plants ranging from mosses, gymnosperms to angiosperms have evolved symbioses with nitrogen-fixing bacteria that convert atmospheric nitrogen into ammonium (Masson-Boivin & Sachs, 2018). For example, the fern *Azolla* maintains colonies of nitrogen-fixing cyanobacteria in specialised apoplastic cavities, outside the plant cell wall enclosure (Peters & Meeks, 1989). A major biological breakthrough was the evolution of the nitrogen-fixing root nodule symbiosis (RNS) characterised by the intracellular accommodation of bacteria in lateral organs ("nodules") formed on roots (Kistner & Parniske, 2002; Parniske, 2000, 2018). The occurrence of the RNS is restricted to a monophyletic clade, encompassing four angiosperm orders: the Fabales, Fagales, Rosales and Cucurbitales (FaFaCuRo) (Soltis et al., 1995). Because of this phylogenetic restriction and scattered occurrence of RNS within the FaFaCuRo, Soltis and colleagues (Soltis et al., 1995) postulated that the last common ancestor of the FaFaCuRo clade acquired a genetic change, a "predisposition", which enabled members of this clade to subsequently evolve RNS multiple times independently (Soltis et al., 1995). Intracellular accommodation of bacteria and root nodule development are two genetically separable and to this extend independent features of RNS (Finan et al., 1985; Murray et al., 2007). It is therefore genetically possible that they did evolve sequentially and not at the same time. The phylogenetic diversity of bacterial symbionts plus the variation of nodule anatomy and development across the RNS-competent FaFaCuRo species (Pawlowski & Demchenko, 2012; Sprent et al., 2017) together with the gap of 30 million years between the last common ancestor and the oldest fossil root nodules in this clade (Doyle, 2011) further fuelled the hypothesis that nodule organogenesis evolved several times independently and was not a feature of the last common ancestor (Doyle, 2016; Sprent, 2007). The recent discovery of multiple losses of RNS within the FaFaCuRo clade (Griesmann et al., 2018; van Velzen et al., 2018) has initiated a discussion about whether this genetic change in the common ancestor was perhaps sufficient for the formation of RNS (van Velzen et al., 2019). Nonetheless, the precise nature of this key event in the evolution of nodulation has remained a mystery for more than two decades (Doyle, 2016).

We asked which evolutionary acquisitions by the last common ancestor, in the form of novel traits and the underlying genetic causes, enabled the evolution of the RNS. From a phylogenetic perspective, such acquisitions should be: 1) exclusively present in the FaFaCuRo clade and absent outside of this clade and 2) conserved throughout the FaFaCuRo clade or at least maintained in RNS-competent (hereafter called "nodulating") species. The uptake of bacteria into living plant cells is, with one exception (*Gunnera*), phylogenetically restricted to the FaFaCuRo clade (Parniske, 2000). Uptake of bacteria requires the localised

70 lysis of the plant cell wall, which threatens cell integrity because of the turgor pressure imposed
71 by the protoplast (Parniske, 2018). A systematic comparison of features associated with the
72 RNS across the entire FaFaCuRo clade pinpoints a single unique and shared trait – the uptake
73 of bacteria into living plant cells with intracellular physical support structures – that fulfils both
74 abovementioned criteria to be acquired by the common ancestor (Parniske, 2018). These
75 structures come in a diversity of shapes (infection threads and infection pegs) and in at least
76 two different cell types (epidermal and cortical), but are all characterised by the apposition of
77 matrix material which is thought to maintain cell integrity during the localised lysis of the plant
78 cell wall. While this matrix material is a common feature of all analysed successful bacteria
79 uptake events in FaFaCuRo species, only one type, cortical infection threads (ITs), can be
80 found in almost all nodulating species (Parniske, 2018). Cortical IT formation is an evolutionary
81 breakthrough because it allowed clonal selection of bacteria (Gage, 2002), specific control of
82 nutrient exchange and increased nitrogen fixation efficiency (Carvalho et al., 2014). By
83 contrast, in *Gunnera*, cell integrity is maintained by physical closure of a multicellular cavity by
84 extracellular matrix material (Johansson & Bergman, 1992). This difference, together with the
85 phylogenetic distance of *Gunnera* from the FaFaCuRo clade, suggests an independent origin
86 of bacterial uptake in this genus (Parniske, 2018). To search for gene gains specific for the
87 FaFaCuRo clade, a genome-wide comparative phylogenomic analysis was performed,
88 however, not a single gene following the aforementioned evolutionary pattern was identified
89 (Griesmann et al., 2018).

90 Here, we tested the hypothesis that the "predisposition" event involved gain of novel *cis*-
91 regulatory elements. It has been shown that changes in gene regulation are important drivers
92 of functional and morphological evolution. Emergence or loss of even a single *cis*-regulatory
93 element can lead to dramatic phenotypic consequences, e.g. novel organ formation (Kvon et
94 al., 2016; Wittkopp & Kalay, 2011). Phylogeny has dated the common ancestor of the
95 FaFaCuRo clade to approximately 90 Mya (Bell et al., 2010; Li et al., 2015; Wang et al., 2009).
96 A long standing hypothesis states that the evolution of RNS involved co-opting genes from the
97 arbuscular mycorrhiza (AM) symbiosis (Kistner & Parniske, 2002; Parniske, 2000), which can
98 be traced back to the earliest land plant fossils 450 Mya (Redecker et al., 2000; Remy et al.,
99 1994). This hypothesis is underpinned by similarities in intracellular accommodation structures
100 (Parniske, 2018) and the common requirement of both symbioses for a set of so-called
101 "common symbiosis genes" (Kistner & Parniske, 2002) that are conserved across land plant
102 species able to form AM, and encode symbiotic signal transduction and intracellular
103 restructuring machineries (Banba et al., 2008; Chen et al., 2007; Gutjahr et al., 2008;
104 Markmann et al., 2008; Yano et al., 2008).

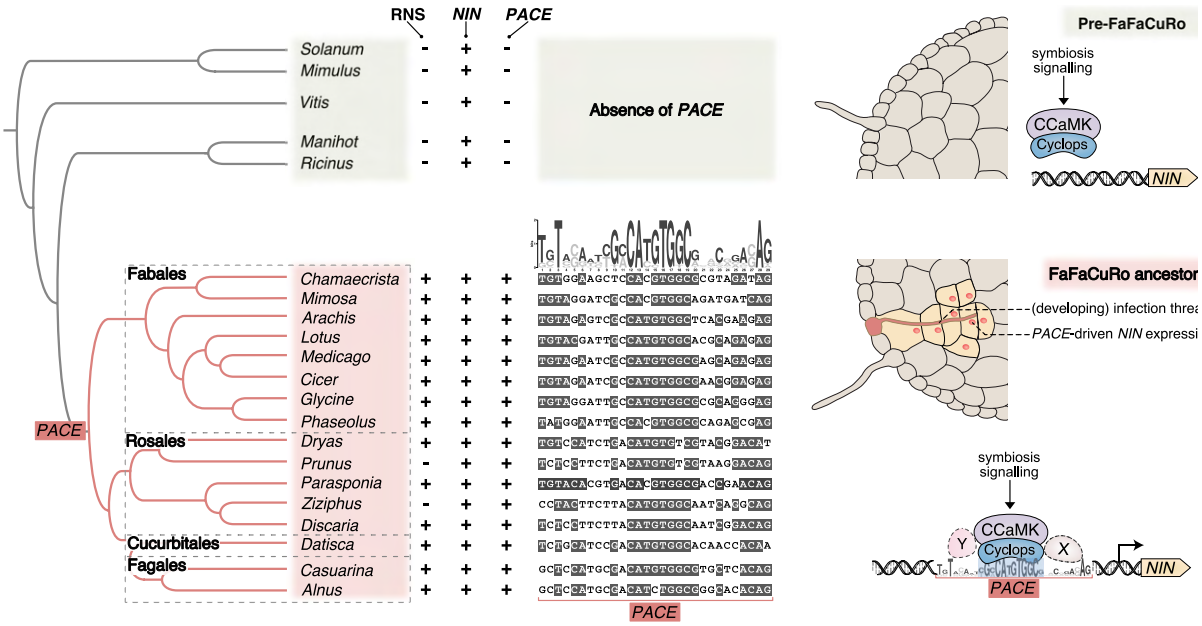


Figure 1. Acquisition of *PACE* was a key step in the evolution of RNS.
Left: Schematic illustration of phylogenetic relationships between species inside (light red shade) and outside (light grey shade) the FaFaCuRo clade and presence (+) and absence (-) pattern of RNS, *NIN* and *PACE* (See **Figure 1 – figure supplement 1 and 2** for additional data support). **Middle:** *PACE* sequence alignment of the displayed species in which grey shadings indicate more than 50% sequence identity. On top of the alignment the *PACE* consensus sequence depicted as Position Weight Matrix calculated from the displayed RNS-competent species. **Right:** Graphical illustration of how *PACE* connected *NIN* to symbiotic transcriptional regulation by CCaMK/Cyclops, enabling IT development in the root cortex. This acquisition coincided with the predisposition event. X and Y: hypothetical proteins binding to *PACE* outside of the Cyclops binding site.

Results and discussion

Discovery of *PACE*

The transcription factor-encoding *Nodule Inception* (*NIN*) gene (Schauser et al., 1999; Soyano et al., 2013) is positioned at the top of a RNS-specific transcriptional regulatory cascade and is indispensable for RNS (Schauser et al., 1999; Singh et al., 2014; Soyano & Hayashi, 2014). The promoter of *NIN* is a potential physical target for such a co-option event, because it defines the molecular interface between common symbiotic signal transduction and the specific transcriptional networks underlying RNS development (Soyano & Hayashi, 2014). We therefore compared the *NIN* promoter sequences of 37 angiosperm species including 27 FaFaCuRo members and identified only one motif fulfilling the aforementioned criteria, which we called *Predisposition-Associated Cis-regulatory Element* (*PACE*) (**Figure 1**; **Figure 1 – figure supplement 1A - D** and **Figure 1 – Source Data 1**). The phylogenetic distribution of *PACE* was further investigated in an expanded search comprising 163 plant species in the promoter of *NIN* and the entire *NIN-like protein* (*NLP*) gene family, including *NLP1* from which *NIN* diverged at the base of the eudicots (Soyano & Hayashi, 2014) (**Figure 1 – figure supplement 1E**; **Figure 1 – figure supplement 2** and **Figure 1 – Source Data 2**). *PACE*

121 was found in all nodulating FaFaCuRo members and four non-nodulating species that have
122 lost RNS but maintained *NIN* (**Figure 1 – figure supplement 1E** and **Figure 1 – Source Data**
123 **3**). Importantly, *PACE* was absent from all the *NLP* promoters analysed (**Figure 1 – figure**
124 **supplement 2**). Thus, the phylogenetic distribution pattern of *PACE* is FaFaCuRo-clade
125 specific and is consistent with a model in which *PACE* was acquired by the *NIN* promoter of
126 the last common FaFaCuRo ancestor. Intriguingly, the 29 nucleotides-long *PACE*
127 encompassed and extended beyond the previously identified binding site of the transcription
128 factor Cyclops, encoded by a common symbiosis gene required for the development of both
129 AM and RNS (Singh et al., 2014; Yano et al., 2008).

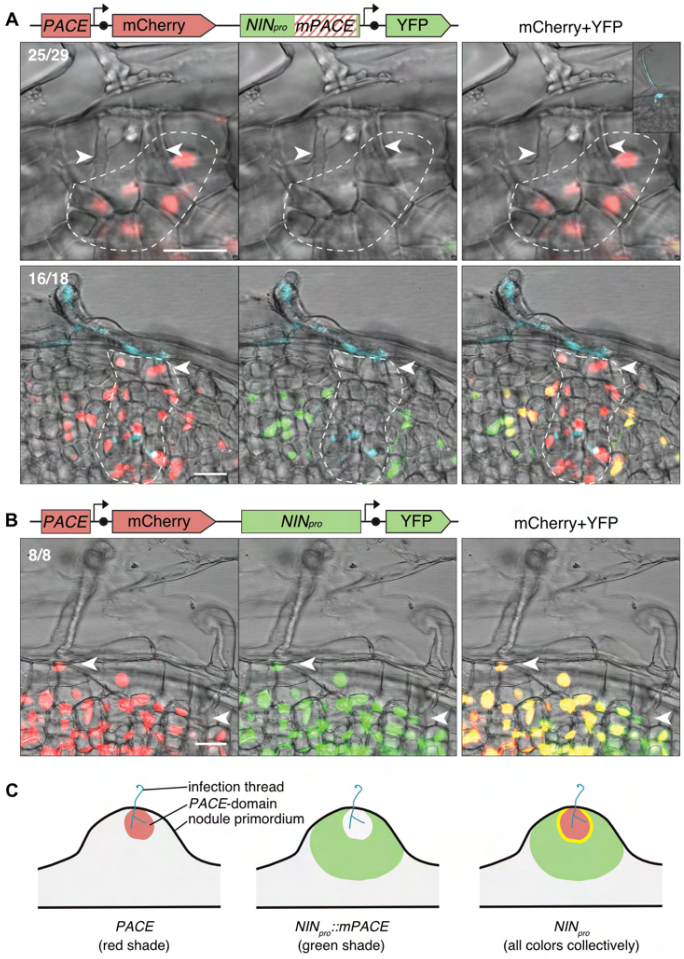


Figure 2. *PACE* drives the expression of *NIN* during IT development in the cortex. Sections of representative *L. japonicus* nodule primordia formed upon inoculation with *M. loti* R7A expressing CFP (blue) imaged by confocal laser-scanning microscopy. Comparison of the expression domains determined by (**A**) *PACE* (*PACE:NINmin$_{pro}$:NLS-mCherry*; red) and a *NIN* promoter carrying a mutated *PACE* (*NIN$_{pro}$::mPACE:NLS-YFP*; green); or (**B**) *PACE* (red) and the intact *NIN* promoter (*NIN$_{pro}$:NLS-YFP*; green). Dashed lines demarcate a group of cortical cells in the *PACE* core territory. Arrowheads indicate ITs. Numbers: nodule primordia showing the presented expression pattern / total number of nodule primordia sectioned and inspected. Bars, 20 µm. (**C**) graphical interpretation of expression patterns presented in (**A and B**). Yellow: overlapping region.

130 Given this clade-specific distribution of *PACE*, we searched for conserved motifs in the
131 promoter sequences of two genes encoding transcriptional regulators, *ERF Required for*
132 *Nodulation 1* (*ERN1*) (Cerri et al., 2017) (**Figure 1 – figure supplement 3**) and *Reduced*
133 *Arbuscular Mycorrhiza 1* (*RAM1*) (Pimprikar et al., 2016) (**Figure 1 – figure supplement 4**)
134 that are also known Cyclops targets. We identified motifs within the promoters of both, *ERN1*
135 and *RAM1*, encompassing the previously identified Cyclops binding sites (Cerri et al., 2017;
136 Pimprikar et al., 2016). In sharp contrast to *PACE*, their presence extended beyond the
137 FaFaCuRo clade (**Figure 1 – figure supplement 3** and **Figure 1 – figure supplement 4**).

138 We tested the functional relevance of these distinct phylogenetic distribution patterns in
139 transcriptional activation assays in *Nicotiana benthamiana* leaf cells. Transactivation by
140 Cyclops was restricted to *NIN* promoters from FaFaCuRo species, but extended to non-
141 FaFaCuRo species for *RAM1* promoters (**Figure 1 – figure supplement 5**). Importantly,
142 *PACE* was necessary and sufficient for the activation of the *NIN* promoter by Cyclops (**Figure**
143 **1 – figure supplement 6**). Together with the exclusive occurrence of *PACE* in the *NIN*
144 promoter of the FaFaCuRo clade, these results are in line with the hypothesis that the
145 mechanistic link between Cyclops and the *NIN* promoter was established in the last common
146 ancestor of this clade (**Figure 1**).

147 ***PACE* drives the expression of *NIN* during IT development in the cortex**

148 *NIN* is indispensable for IT development (Schauser et al., 1999; Soyano et al., 2013) and its
149 precise spatiotemporal expression is essential for this process (Liu et al., 2019; Soyano et al.,
150 2013; Vernié et al., 2015; Yoro et al., 2014). Because *cis*-regulatory elements are master
151 determinants of gene expression patterns (Buecker & Wysocka, 2012), we investigated the
152 impact of *PACE* on the expression of *NIN* in physical relation to the bacterial uptake and
153 accommodation stages during nodule development. We used the model legume *Lotus*
154 *japonicus* in combination with its compatible nitrogen-fixing bacterium *Mesorhizobium loti* as
155 experimental system. The process by which *M. loti* is taken up by *L. japonicus* can be
156 subdivided into successive stages: (1) entrapment of bacteria in a pocket formed by a curled
157 root hair (Perrine-Walker et al., 2014), (2) development of an IT within that root hair (Perrine-
158 Walker et al., 2014), (3) IT progression into and through the outer cortical cell layers (Van
159 Spronsen et al., 2001), (4) IT branching and extension within the nodule primordium (Yoon et
160 al., 2014) (5) release of bacteria from ITs into plant membrane-enclosed organelle-like
161 structures called symbiosomes (Yoon et al., 2014) leading to (6) mature nodules characterised
162 by infected cells densely packed with symbiosomes and the pink colour of leghemoglobin (Ott
163 et al., 2005).

164 To determine the *PACE*-mediated spatiotemporal expression domain, we introduced a *GUS*
165 reporter gene driven by *PACE* fused to a region comprising the *NIN* minimal promoter and the

166 5'UTR (Singh et al., 2014) (*PACE:NINmin$_{pro}$:GUS*) into *L. japonicus* wild-type roots. Roots
167 were subsequently inoculated with *M. loti* MAFF 303099 expressing *Ds*Red (*M. loti Ds*Red)
168 facilitating detection of the bacteria through their fluorescence signal in root hairs and nodules.
169 The *NIN* minimal promoter did not mediate reporter gene expression at any stage of bacterial
170 infection (**Figure 2 – figure supplement 1E**). Intriguingly, the earliest detectable GUS activity
171 mediated by *PACE:NINmin$_{pro}$:GUS* was clearly restricted to a zone in the nodule primordia
172 (panel I - II in **Figure 2 – figure supplement 1D**) that roughly correlated with the site of
173 bacterial infection (indicated by a local accumulation of *Ds*Red signal) and later expanded to
174 the entire central tissue of the nodule (panel III in **Figure 2 – figure supplement 1D**). *PACE*-
175 driven reporter expression was neither detected in root hairs harbouring ITs (**Figure 2 – figure**
176 **supplement 1G**) nor in nodules in which cortical cells were filled with symbiosomes (panel IV
177 in **Figure 2 – figure supplement 1D**). Importantly, *PACE*-mediated expression was distinct
178 from that mediated by the *LjNIN* 3 kb promoter (*NIN$_{pro}$*) or the *NIN$_{pro}$* with *PACE* mutated or
179 deleted (*NIN$_{pro}$::mPACE* and *NIN$_{pro}$::ΔPACE*, respectively) that conferred reporter expression
180 across the central tissue of the nodule (panels II - IV in **Figure 2 – figure supplement 1A -**
181 **C**). We concluded based on these observations that the *PACE*-mediated expression domain
182 is temporally and spatially restricted and possibly accompanies the development of bacterial
183 accommodation structures in the nodule.

184 To further resolve this relationship between *PACE* driven gene expression and bacterial
185 accommodation at the cellular level, we compared – simultaneously in the same tissue – the
186 progression of bacterial infection with the expression pattern mediated by *PACE* fused to the
187 *NIN* minimal promoter (*PACE:NINmin*) and by a *NIN* promoter with mutated *PACE*
188 (*NIN$_{pro}$::mPACE*). A red and a yellow fluorescent protein (mCherry and YFP, respectively)
189 targeted to the nucleus by fusion to a nuclear localization signal (NLS) were used as reporters.
190 The resulting promoter:reporter fusions (*PACE:NINmin$_{pro}$:NLS-mCherry* and
191 *NIN$_{pro}$::mPACE:NLS-YFP*) were placed in tandem on the same T-DNA allowing a nucleus-by-
192 nucleus comparison of their relative expression. This T-DNA construct was introduced into *L.*
193 *japonicus* wild-type roots that were subsequently inoculated with *M. loti* R7A expressing the
194 cyan fluorescent protein (CFP) to facilitate detection (**Figure 2**). In sections of developing
195 nodules, in which infection had progressed to stage 3 or 4, *PACE*-mediated *mCherry* was
196 expressed specifically in a – hereafter called "infection zone" – comprising cortical cells that
197 carried ITs and in some, but not all, directly adjacent cells (25 out of 29 nodules inspected;
198 **Figure 2A**). Intriguingly, the expression domains marked by mCherry and YFP fluorescence
199 were distinct from each other: while the *PACE*-driven mCherry signal was consistently marking
200 the infection zone, the *NIN$_{pro}$::mPACE*-driven YFP signal was observed in cortical cells
201 surrounding this zone (16 out of 18 nodules inspected; **Figure 2A, 2C**). The thin (approx. 1-2

202 cells thick) border between the two domains was characterised by nuclei emitting both YFP
203 and mCherry signals (**Figure 2A**). In so-marked cells, ITs were typically not detected. The
204 expression pattern mediated by the *NIN* promoter (containing *PACE*) was congruent with the
205 sum of both promoter fragments (8 out of 8 nodules inspected; **Figure 2B, 2C**).

206 Based on these clearly distinct and complementary reporter expression domains governed by
207 *PACE* versus the remaining promoter, we concluded that 1) *PACE* directs *NIN* expression to
208 a specific infection zone and that 2) the *NIN* promoter comprises *cis*-regulatory elements that
209 drive expression outside the *PACE* territory i.e. in root hairs (together with *PACE*), non-
210 infected cortical cells and cells filled with symbiosomes. These additional *cis*-regulatory
211 elements might be addressed by other transcription factors that have been reported to bind to
212 this promoter (Hirsch et al., 2009; Singh et al., 2014; Xiao et al., 2020; Zhu et al., 2008).

213 **Mutational dissection of *PACE* reveals a quantitative impact of regions flanking the**
214 ***CYC-box* on IT development**

215 To test the relevance and specific role of *PACE* in nodule and IT development, we performed
216 complementation experiments using plants homozygous for the *nin-2* or *nin-15* mutant alleles
217 (Schauser et al., 1999). The *nin-2* mutant allele harbours a frameshift mutation of the *NIN*
218 gene, leading to a *NIN* loss-of-function phenotype, i.e. absence of both IT formation and
219 nodule organogenesis (Schauser et al., 1999) while the *nin-15* mutant allele carries a *Lotus*
220 *Retrotransposon 1* insertion within the *NIN* promoter 143 bp 3' of *PACE* (**Figure 3 – figure**
221 **supplement 1**). We examined the restoration of bacterial infection 21 days after inoculation
222 with *M. loti Ds*Red by quantifying the number of root hairs harbouring ITs and the number of
223 infected nodules (**Figure 3** and **Figure 3 – Source Data 1**).

224 Nodule development in the legume *Medicago truncatula* is dependent on *NIN* expression
225 mediated by a regulatory region containing several putative cytokinin responsive elements
226 (*CE*) (Liu et al., 2019). In *L. japonicus*, a similar *CE* region is positioned 45 kb upstream of the
227 *NIN* transcriptional start site (Liu et al., 2019). To enable transgenic complementation
228 experiments, we synthetically fused a 1 kb or 5 kb region encompassing this distant *CE* to the
229 5' end of a 3 kb *NIN* promoter. The *NIN* gene driven by these promoters (CE_{1kb}:NIN_{pro}:*NIN* and
230 CE_{5kb}:NIN_{pro}:*NIN*) restored the formation of root hair ITs on 78% and 95% and infected nodules
231 on 40% and 88% of *nin-2* transgenic root systems, respectively (**Figure 3A**; **Figure 3 – figure**
232 **supplement 2, 3, 4, 5 and 6**). Importantly, this complementation success relied on the
233 presence of *PACE*; *nin-2* roots transformed with the same fusion design but carrying a
234 mutation of *PACE* (CE_{1kb}:NIN_{pro}::*mPACE:NIN* and CE_{5kb}:NIN_{pro}::*mPACE:NIN*) did not restore
235 root hair ITs but nodule formation was not impaired when using the cytokinin element-
236 containing region of 5 kb (CE_{5kb}:NIN_{pro}::*mPACE:NIN*). We concluded that *PACE* is
237 indispensable for bacterial infection but not for nodule development.

238 The 29 bp long *PACE* sequence encompasses and extends beyond the previously identified
239 Cyclops binding site (**Figure 1**). To dissect the specific contributions of the Cyclops binding
240 site (*CYC-box* (Singh et al., 2014), "box") and *PACE* sequences flanking the *CYC-box*
241 ("flanking") to *PACE* function, we mutated the box and the flanking region independently
242 (*CE:NIN$_{pro}$::mbox:NIN* and *CE:NIN$_{pro}$::mflanking:NIN*, respectively). Mutation of the *CYC-box*
243 abolished root hair ITs. Interestingly, mutation of the flanking sequences led to a 50%
244 reduction of the number of transgenic root systems carrying infected nodules, while the
245 formation of root hair ITs was not impaired (**Figure 3 – figure supplement 2, 3, 4, 5 and 6**).
246 This mutational dissection revealed two separable functions of *PACE*: while the *PACE*-
247 Cyclops connection is essential for IT development, the flanking sequences are positively
248 regulating the infection of newly developed nodules and possibly act as binding sites for
249 additional, yet undefined, transcription factors (**Figure 1**). The existence of transcription
250 factors that act synergistically with Cyclops is in line with the original *cyclops* mutant phenotype
251 description which called for the existence of CCaMK phosphorylation targets that can partially
252 compensate for the loss of Cyclops in nodule development but not in infection (Yano et al.,
253 2008). Our data suggest that *PACE* comprises synergistic binding sites for both Cyclops and
254 cooperating transcription factors.

255 *PACE*-mediated *NIN* expression defined an infection zone in the nodule cortex (**Figure 2**). To
256 genetically separate the initiation of nodule development from IT formation and thereby enable
257 a focussed analysis of the role of *PACE* in cortical IT formation, we utilised the *nin-15* mutant,
258 which is impaired in IT formation but retains the capacity to form nodules. Most of these
259 nodules were uninfected (92% and 86% plants carrying no root hair ITs and no infected
260 nodules, respectively) and cortical cells filled with symbiosomes were never observed (**Figure**
261 **3 – figure supplement 1**). This mutant therefore provided an ideal background to study the
262 role of *PACE* in cortical IT formation, circumventing the negative epistatic effect of the inability
263 of *nin* loss-of-function mutants to initiate cell divisions (Clavijo et al., 2015; Liu et al., 2019;
264 Schauser et al., 1999; Vernié et al., 2015; Yoro et al., 2014) (**Figure 3B - D**; **Figure 3 – figure**
265 **supplement 7**; **Figure 4**; **Figure 4 – figure supplement 1** and **Figure 3 – Source Data 1**).

266 Transformation with the *L. japonicus NIN* gene driven by the *NIN* minimal promoter
267 (*NINmin$_{pro}$:NIN*) did not alter the symbiotic phenotype of *nin-15* roots (**Figure 3B**). In contrast,
268 the *NIN* gene driven by the *NIN* promoter (*NIN$_{pro}$:NIN*) led to restoration of the complete
269 infection process in *nin-15* roots from root hair ITs to symbiosome formation (100% and 92%
270 of transgenic root systems carried root hairs ITs and infected nodules, respectively; **Figure**
271 **3B**). Similar to observations in complementation experiments of *nin-2*, mutation or deletion of
272 *PACE* (*NIN$_{pro}$::mPACE:NIN* and *NIN$_{pro}$:: ΔPACE:NIN,* respectively) drastically reduced the

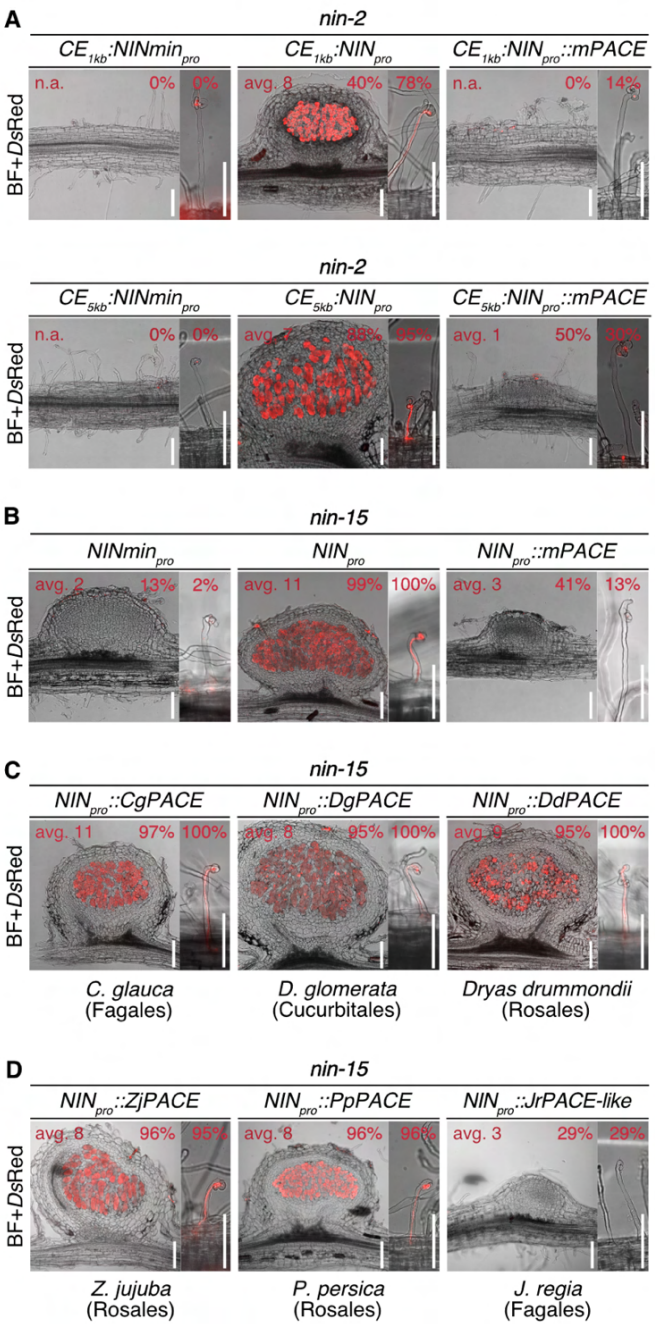


Figure 3. *PACE* is necessary for bacterial infection and functionally conserved across the FaFaCuRo clade. Microscopy images of representative nodule sections or root hairs harbouring an IT or an infection pocket from (**A**) *nin-2* or (**B - D**) *nin-15* roots transformed with the *LjNIN* gene driven by: (**A - B**) indicated promoters, and in (**C - D**) the *L. japonicus NIN* promoter in which *LjPACE* was replaced by *PACE* from (**C**) nodulating and (**D**) non-nodulating FaFaCuRo species or with a *PACE*-like sequence identified in the *JrNLP1b* promoter. %: percentage of transgenic root systems carrying infected nodules or root hair ITs. Avg.: average number of infected nodules on plants carrying infected nodules. n.a.: not applicable. Bars, 100 µm.

273 restoration of bacterial infection in root hairs and nodules in *nin-15* (**Figure 3B**; **Figure 3 –**
274 **figure supplement 2, 3, 4, 5, 6 and 7**).

275 ***PACEs* from different nodulating FaFaCuRo species are functionally equivalent**
276 *PACE* was detected by MEME searches as a conserved motif within *NIN* promoters of the
277 FaFaCuRo clade. However, the individual *PACE* sequences from different species differed
278 from each other (**Figure 1** and **Figure 1 – figure supplement 1**). We therefore tested whether
279 and to what extend this sequence variation of *PACE* would affect its function. Replacement of
280 *PACE* within the *L. japonicus* (Fabales) 3 kb *NIN* promoter with *PACE* sequence variants
281 (NIN_{pro}*::Species abbreviation PACE:NIN*) originating from *Casuarina glauca* (*Cg*, Fagales),
282 *Datisca glomerata* (*Dg*, Cucurbitales) or *Dryas drummondii* (*Dd*, Rosales) restored the
283 complete infection process in *nin-15* to similar level as NIN_{pro}*:NIN*, demonstrating the
284 functional conservation of *PACE* from nodulating species across the entire FaFaCuRo clade
285 (**Figure 3C** and **Figure 3 – figure supplement 7**). Similarly, the *PACE* versions from two non-
286 nodulating Rosales that maintained the *NIN* gene, *Ziziphus jujuba* and *Prunus persica*,
287 restored the complete infection process in *nin-15* (**Figure 3D** and **Figure 3 – figure**
288 **supplement 7**). The results of these complementation experiments were consistent with the
289 conserved expression pattern mediated by *PACEs* in *L. japonicus* (**Figure 3 – figure**
290 **supplement 8**) and the CCaMK/Cyclops-mediated transactivation via these *PACE* variants
291 (Extended Data Fig. 3a) or chimeric promoter:reporter fusions (**Figure 1 – figure supplement**
292 **6B**) tested in *N. benthamiana* leaves.

293 **Loss of *PACE* is associated with a loss of nodulation**
294 Griesmann et al. (Griesmann et al., 2018) and van Velzen et al. (van Velzen et al., 2018)
295 discovered that RNS was lost multiple times independently during evolution, via independent
296 truncations or losses of the *NIN* gene. However, at least 10 out of 28 FaFaCuRo species that
297 lost RNS have maintained a full-length *NIN* open reading frame (**Figure 1 – Source Data 3**).
298 Based on our complementation data, *PACE* is indispensable for the *NIN* promoter function in
299 symbiosis (**Figure 3** and **Figure 3 – figure supplement 2, 3, 4, 5, 6 and 7**). Therefore, the
300 absence of *PACE* from 5 out of these 10 species (**Figure 1 – Source Data 1, 2 and 3**), is
301 potentially sufficient to explain these losses of RNS. Consequently, at least 82% of all losses
302 can now be attributed to either the *NIN* ORF (18/28, 64%) or loss of *PACE* (5/28, 18%) (**Figure**
303 **1 – Source Data 3**). The presence of *PACE* in all nodulating species (31/31, 100%; **Figure 1**
304 **– Source Data 1 and 2**) together with a correlation between the absence of *PACE* with the
305 absence of RNS adds strong support for the evolutionary relevance of *PACE* both in the gain
306 and potential loss of RNS.

307 *PACE* was not detected in the promoters of *NIN-like protein* (*NLP*) genes (**Figure 1 – figure**
308 **supplement 2** and **Figure 1 – Source Data 1 and 2**) with the possible exception of the curious
309 case of *Juglans regia* (Fagales). While it was also absent from the promoter of the so-
310 annotated *NIN* gene, a *PACE*-like motif was identified in the promoter of the closest gene

311 family member, *NIN-like protein 1 JrNLP1b* (*JrPACE-like*; **Figure 1 – Source Data 1**). This
312 *PACE*-like element was not able to restore IT formation in *nin-15* (**Figure 3D** and **Figure 3 –**
313 **figure supplement 7**). Regardless of whether this exceptional presence/absence pattern of
314 *PACE* may be caused by a miss-annotation of *NIN* and *NLP1* in *J. regia*, either a loss-of-
315 function mutation within *PACE* or a loss of the entire *PACE* element in the *JrNIN* promoter
316 could explain the absence of the RNS observed in this species.

317 ***PACE* is sufficient to restore cortical IT formation in *nin-15***

318 We tested whether *PACE* on its own, only supported by the minimal *NIN* promoter
319 (*PACE:NINmin*$_{pro}$) is sufficient to restore IT development in cortical cells. For this purpose, we
320 transformed *nin-15* roots with *PACE:NINmin*$_{pro}$ fused to the transcribed region of the *NIN* gene.
321 *PACE*-mediated *NIN* expression led to an increased success in restoration of infection (49%
322 of transgenic root systems carried infected nodules) compared to *NINmin*$_{pro}$*:NIN*-transformed
323 roots (17%; **Figure 4A**; **Figure 4 – figure supplement 1** and **Figure 3 – Source Data 1**).
324 Root hair ITs were rarely observed on *PACE:NINmin*$_{pro}$*:NIN*-transformed *nin-15* roots (**Figure**
325 **4A**) and infected nodules harbouring cortical cells filled with symbiosomes were never
326 observed (**Figure 4 – figure supplement 1**), consistent with the restricted expression domain
327 defined by *PACE* (**Figure 2 – figure supplement 1**).

328 Strikingly, the vast majority of infected nodules transformed with *PACE:NINmin*$_{pro}$*:NIN* (25 out
329 of 28 nodules inspected) carried ITs in the outer cortex, originating from a focused
330 hyperaccumulation of bacteria, locally constricted by root cell wall boundaries (**Figure 4A**).
331 This phenomenon was not observed in most of the rarely occurring infected nodules formed
332 on *NINmin*$_{pro}$*:NIN*-transformed *nin-15* roots (11 out of the 16 nodules inspected did not carry
333 ITs in the outer cortex). Bacterial colonies within cell wall boundaries resembling this
334 phenomenon have been described in a variety of legumes including *Sesbania* and *Mimosa*
335 (D'Haeze et al., 1998; De Faria et al., 1988). Our data imply that *PACE* promotes this type of
336 cortical IT initiation. Altogether, these findings revealed that *PACE* promotes IT development
337 in cortex cells but not within root hairs.

338 ***PACE* insertion into the tomato *NIN* promoter confers RNS capability**

339 To artificially recapitulate the functional consequence of *PACE* acquisition into a non-
340 FaFaCuRo *NIN* promoter, we chose tomato (*Solanum lycopersicum*) which belongs to the
341 Solanaceae, a family phylogenetically distant from the FaFaCuRo clade. Consistent with the
342 absence of *PACE*, a *GUS* reporter gene driven by the tomato *NIN* promoter (*SlNIN*$_{pro}$) was not
343 transactivated by Cyclops in *N. benthamiana* leaf cells (**Figure 1** and **Figure 1 – figure**
344 **supplement 5B and 6C**), while the insertion of the *L. japonicus PACE* (*SlNIN*$_{pro}$*::PACE*), but
345 not of a mutated *PACE* (*SlNIN*$_{pro}$*::mPACE*) conferred transactivation by Cyclops (**Figure 1 –**
346 **figure supplement 6C**).

347 We tested the ability of the *LjNIN* expressed under the control of these synthetic promoters to
348 restore the bacterial infection process in *nin-15*. Similar to *NINmin*$_{pro}$*:NIN*-transformed *nin-15*
349 roots, *SlNIN*$_{pro}$*:NIN* did not restore bacterial infection (0% and 7% of transgenic root systems
350 carried root hair ITs and infected nodules, respectively; **Figure 4A - C**). In contrast, *nin-15*
351 roots transformed with *SlNIN*$_{pro}$*::PACE:NIN* restored the formation of root hair ITs and infected
352 nodules on 36% and 26% of transgenic root systems, respectively (**Figure 4B** and **Figure 4**

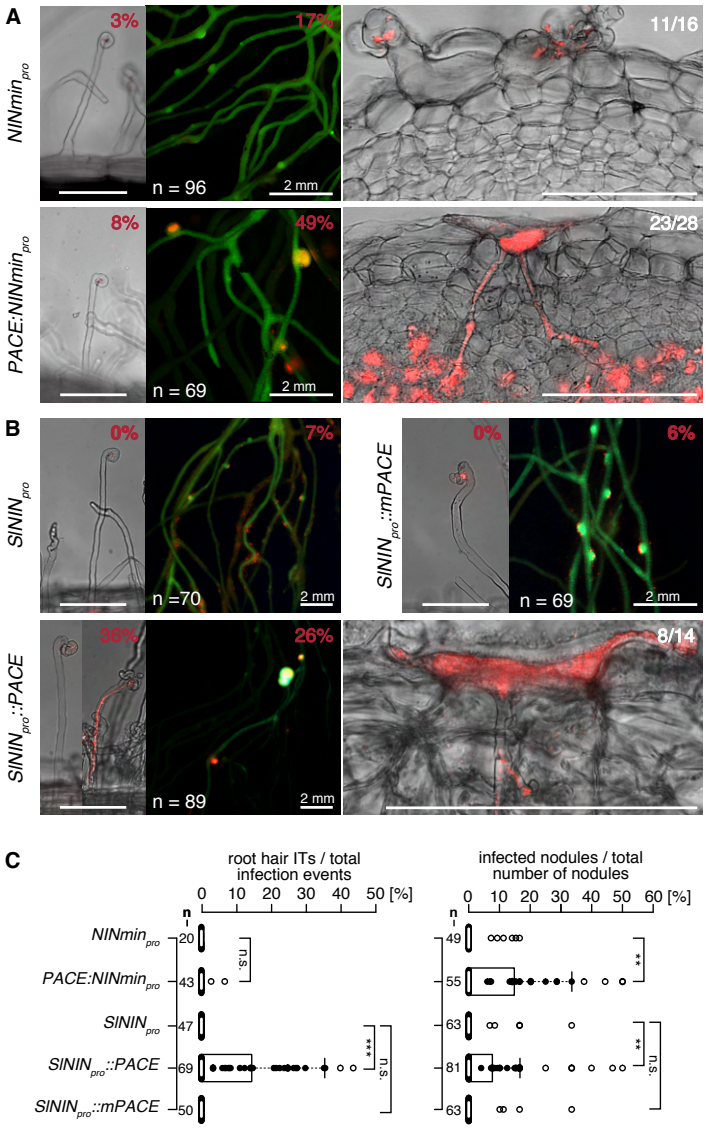


Figure 4. *PACE* enables IT formation in the cortex. Representative pictures of *nin-15* root hairs, root and nodule sections (see Extended Data Fig. 10 for overview pictures), transformed with the *L. japonicus NIN* gene driven by (**A**) *NINmin*$_{pro}$ or *PACE:NINmin*$_{pro}$; (**B**) *S. lycopersicum NIN* promoter (*SlNIN*$_{pro}$) and *SlNIN*$_{pro}$ with *LjPACE* or *mPACE* inserted. %: percentage of transgenic root systems carrying root hair ITs or infected nodules; Ratios: number of nodules showing the presented pattern / total number of nodules sectioned and inspected. (**C**) Boxplots displaying the percentage of root hair ITs and infected nodules per transgenic root system. n, number of transgenic root systems or root pieces analysed. The applied statistical method was Fisher's exact test: $^{**}p<0.01$; $^{***}p<0.001$; n.s., not significant. Unlabelled scale bars, 100 µm.

353 – **figure supplement 1**). This increase in infection success was not observed on
354 *SlNIN$_{pro}$::mPACE:NIN*-transformed roots. ITs in the outer cortex that originated from a focal
355 accumulation of bacteria were also observed in the *SlNIN$_{pro}$::PACE:NIN*-transformed *nin-15*
356 nodules (8 out of 14 nodules inspected; **Figure 4B**) resembling those in the
357 *PACE:NINmin$_{pro}$:NIN*-transformed *nin-15* nodules (**Figure 4A**). The gained ability of the
358 *SlNIN::PACE* promoter to restore root hair ITs suggested that additional *cis*-regulatory
359 elements within the *SlNIN* promoter function together with *PACE* for root hair IT formation. All
360 together, these findings obtained with the tomato *NIN* promoter carrying an artificially inserted
361 *PACE* agree with the hypothesis that the acquisition of *PACE* by a non-FaFaCuRo *NIN*
362 promoter enabled its regulation via Cyclops and laid the foundation for IT formation in cortical
363 cells.

364 **Conclusion**

365 The mechanistic connection between *PACE* and cortical IT formation together with their
366 congruent phylogenetic distribution strongly support the idea that the acquisition of *PACE* by
367 the latest common ancestor of the FaFaCuRo clade enabled cortical ITs and thus laid the
368 foundation for the evolution of present day RNS. Our findings support an evolutionary model
369 in which an ancestral symbiotic transcription factor complex (comprising CCaMK and
370 Cyclops), that facilitated intracellular symbiosis with AM fungi already in the earliest land plants
371 (Delaux et al., 2014, 2015), gained control over the transcriptional regulation of the *NIN* gene
372 by the acquisition of *PACE* (**Figure 1**). This genetic innovation in the last common ancestor of
373 the FaFaCuRo clade extended the function of the ancestral CCaMK complex to initiate cortical
374 IT development. The *NIN*-like protein family underwent important evolutionary steps preceding
375 the origin of RNS including a gene duplication leading to *NIN* and *NLP1* as closest paralogs
376 (Liu & Bisseling, 2020). It is very likely that the NIN protein itself underwent changes that
377 enabled its role in nodulation (Soyano & Hayashi, 2014). Loss of *NIN* events associated with
378 the loss of nodulation are scattered across all four FaFaCuRo orders (Griesmann et al., 2018;
379 van Velzen et al., 2018), suggesting that *NIN* acquired its relevance for nodulation probably
380 before or latest in the last common ancestor. Because our phylogenomic analysis dates the
381 acquisition of *PACE* to the latest ancestor of the FaFaCuRo clade, we conclude that the critical
382 changes within NIN must have occurred simultaneously or earlier. From a statistical point of
383 view, it is likely that the *PACE* acquisition and the RNS-enabling changes within NIN occurred
384 independently from each other. It will be interesting to determine what these critical changes
385 within NIN are and where they occurred phylogenetically.

386 A “young” primary cell wall characteristic for recently divided cells is considered an important
387 prerequisite for cortical IT initiation (Geurts et al., 2016; Parniske, 2018) but cell division is not
388 restricted to the formation of novel organs (Murray et al., 2007). It is therefore conceptually

389 possible that the common ancestor of the FaFaCuRo clade was forming ITs in recently divided
390 cortical cells but in the absence of root nodules. Multiple lines of evidence indicate that the
391 diverse types of lateral organs harbouring nitrogen-fixing bacteria ("nodules") evolved multiple
392 times independently. Indeed, *CE*-mediated *NIN* expression is important for nodule
393 organogenesis in legumes, but upon searching for this regulatory element in a region of 0.1
394 Mb upstream and downstream of the *NIN* gene, Liu and colleagues (Liu et al., 2019) found its
395 presence to be restricted to legume species, indicating an evolutionary emergence
396 independently of and significantly later than the last common ancestor (Liu et al., 2019; Liu &
397 Bisseling, 2020). ITs in root hairs are only found in Fabales and Fagales and therefore also
398 considered a more recent acquisition (Madsen et al., 2010; Parniske, 2018). *CE* only in
399 combination with *PACE* facilitates root hair ITs (**Figure 3 – figure supplement 2, 3, 4, 5 and**
400 **6**) and additional elements in the 3 kb promoter are necessary for nodule and cortical IT
401 development (**Figure 3 – figure supplement 2, 3, 4, 5 and 6**). These observations highlight
402 the enormous complexity of concerted activity of *cis*-elements and transcription factors
403 underlying the spatiotemporal expression control by present day *NIN* promoters in RNS-
404 competent species. Our data pinpoint the acquisition of *PACE* as a key event during the
405 evolution of nodulation. Together with our discovery that multiple independent losses of *PACE*
406 are associated with multiple losses of RNS within the FaFaCuRo clade, our data underpin the
407 essential position of *PACE* in the evolutionary gain and loss of RNS.

408 **Materials and Methods:**

409 **Bioinformatic analyses**

410 Based on the phylogenetic classification of the RWP-RK gene family (Clavijo et al., 2015;
411 Griesmann et al., 2018), 144 NIN/NLP genes were selected from 37 plant species and 13
412 orders ranging from monocotyledons to dicotyledons including the FaFaCuRo clade (**Figure**
413 **1 – Source Data 1**). For each selected gene, 3 kb of sequence upstream of the translational
414 start site including promoter and 5'UTR region was defined and extracted from the
415 corresponding species' genomic sequence, if contig length allowed it. If contig length was
416 limiting, the longest possible sequence stretch was extracted. For the identification of a *cis*-
417 regulatory element specific for *NIN* promoters of the FaFaCuRo clade, the tool MEME
418 (http://meme-suite.org/tools/meme) was used in discriminative mode (option: "search given
419 strand only", default parameters) with *NIN* promoter regions of only nodulating plants. The
420 control group consisted of promoter regions of all *NIN* genes outside of the FaFaCuRo clade
421 and all NLP genes listed in **Figure 1 – Source Data 1**. The highest scoring motif (E-value:
422 1.6e-058) was 27 bp long and contained the much shorter previously described *CYC-box*
423 (Singh et al., 2014) (**Figure 1 – figure supplement 1A**).

424 To refine the conserved region in this motif, MEME analysis was performed again in normal
425 mode (option: “search given strand only”, default parameters) with *NIN* promoters from only
426 nodulating species. This analysis revealed that the most conserved nucleotides are found
427 within 29 nucleotides (nucleotides 10 to 38 in **Figure 1 – figure supplement 1B**). The
428 previous MEME analysis was repeated, but an exactly 29 nucleotides long motif was searched
429 for (resulting in a motif in **Figure 1 – figure supplement 1C**) and the best scoring *NIN* paralog
430 per searched species, i.e. lowest p-value per species were identified (**Figure 1 – Source Data**
431 **1**). In a final step, one best scoring *NIN* promoter region per nodulating species were analysed
432 with MEME (option: “search given strand only”, default parameters) by searching for an exactly
433 29 nucleotides long motif. The resulting motif was named *PACE* (**Figure 1 – figure**
434 **supplement 1D**). This final MEME run was done for two reasons: first in order to avoid a
435 sequence bias towards a single species with multiple *NIN* paralog promoter regions (e.g.
436 soybean); second to avoid a potential sequence bias generated by the promoter region of a
437 *NIN* paralog that might be no longer functional and therefore has mutated sites in its promoter
438 region due to positive selection.

439 As a control, the FIMO tool was used (http://meme-suite.org/tools/fimo, option: “scan given
440 strand only”, default parameters, fdr < 0.1) to search all 144 *NIN* and *NLP* promoter regions
441 (**Figure 1 – Source Data 1**). *PACE* was found within *NIN* promoter regions of all nodulating
442 species analysed and two non-nodulating FaFaCuRo species (*Prunus persica* and *Ziziphus*
443 *jujuba*) (**Figure 1 – Source Data 1**).

444 The presence or absence of *PACE* was further investigated in promoters of *NIN* and *NLPs* in
445 an expanded database of 163 species covering all Viridiplantae lineages. Orthologs of the
446 whole *NLP* family were retrieved using tBLASTn v2.11.0+ (Camacho et al., 2009) with
447 reference sequences from *Medicago truncatula* as query and a cut-off e-value of 1e-10.
448 Sequences were then aligned using MAFFT v7.380 (Katoh & Standley, 2013) with default
449 parameters. Alignments were subjected to phylogenetic analysis to identify orthologs using
450 Maximum Likelihood approach and the IQ-TREE v1.6.7 software (Nguyen et al., 2015). Prior
451 to phylogenetic reconstruction, the best-fitting evolution model was determined for each
452 alignment using ModelFinder (Kalyaanamoorthy et al., 2017) as implemented in IQ-TREE.
453 Branch support was tested using 10,000 replicates of UltraFast Bootstraps using UFBoot2
454 (Hoang et al., 2018). For each identified ortholog, a 5 kb region upstream of the translational
455 start site was extracted. The three different consensuses identified in the previous MEME
456 analyses (**Figure 1 – figure supplement 1A, B, D**) were then searched in all *NLP* upstream
457 regions using FIMO 5.0.2 and a q-value threshold of 0.1. If several motifs were identified in a
458 given upstream region, only the one with the lowest q-value was conserved for further analysis
459 (**Figure 1 – figure supplement 1E and 2** and **Figure 1 – Source Data 2**). *PACE* from

460 *Parasponia andersonii* (a nodulating species from Rosales) was identified in this analysis and
461 included to generate the consensus in **Figure 1**.

462 Promoters of *ERN1* and *RAM1* genes were analysed independently of the previous analysis,
463 using 87 plant genomes covering the main Angiosperms orders (**Supplementary file 1**).
464 Orthologs of each gene were retrieved using tBLASTn v2.7.1+ (Camacho et al., 2009) with
465 reference sequences from *Medicago truncatula* as query and a cut-off e-value of 1e-10.
466 Sequences were then aligned using MAFFT v7.380 (Katoh & Standley, 2013) with default
467 parameters. Alignments were subjected to phylogenetic analysis to identify orthologs using
468 Maximum Likelihood approach and the IQ-TREE v1.6.7 software (Nguyen et al., 2015). Prior
469 to phylogenetic reconstruction, the best-fitting evolution model was determined for each
470 alignment using ModelFinder (Kalyaanamoorthy et al., 2017) as implemented in IQ-TREE.
471 Branch support was tested using 10,000 replicates of UltraFast Bootstraps using UFBoot2
472 (Hoang et al., 2018). For each identified ortholog, regions upstream of the translational start
473 site of different lengths (1 to 5 kb) were extracted. For each length of region upstream of the
474 translational start site, sequences were analysed using MEME software v5.0.1 (Bailey &
475 Elkan, 1994) with the following parameters: a motif size comprised between 5 - 45 bp and 5 -
476 25 bp for *ERN1* and *RAM1* respectively and a maximum number of discovered motifs of 20
477 (**Figure 1 – figure supplement 3 and 4**). In addition, MEME search was set on "zoops" mode,
478 assuming that each sequence can contain zero or one occurrence of the motif.

479 **Biological material**

480 *L. japonicus* seed bags, bacterial strains and days post inoculation for each experiment are
481 listed in **Supplementary file 2**.

482 **Plant growth conditions and symbiotic inoculations**

483 *Lotus japonicus* seeds were scarified and surface-sterilized as described (Gossmann et al.,
484 2012) before germination on ½ Gamborg's B5 medium solidified with 0.8 % Bacto™ agar in
485 square plates (12 x 12 x 1.7 cm) (Gamborg et al., 1968). Plates were kept in dark for three
486 days before transferring to light condition in a Panasonic growth cabinet (MLR-352H-PE) at
487 24 °C under a 16 h/8 h light/dark regime (50 µmol·m-2·s-1). Six-days-old seedlings were (1)
488 subject to hairy root transformation as described (Charpentier et al., 2008) (**Figure 2, 3 and**
489 **4**; **Figure 2 – figure supplement 1**; **Figure 3 – figure supplement 1, 3, 4, 5, 6, 7 and 8** and
490 **Figure 4 – figure supplement 1**); or (2) transferred to Weck jars (SKU 745 or 743; J.Weck
491 GmbH u. Co. KG) containing 300 ml of sand:vermiculite mixture (2:1) and 20 ml of a modified
492 ¼ strength Hoagland's medium with Fe-EDDHA used as iron source (Hoagland & Arnon,
493 1950) (**Figure 3 – figure supplement 1E**). For *in vivo* promoter expression analysis (**Figure**
494 **2**), transgenic roots expressing a kanamycin-resistance gene were kept on square plates
495 supplemented with kanamycin (25 µg/ml) 10 days after *Agrobacterium rhizogenes* inoculation.

496 Plants with transformed roots were kept on 0.8% Bacto™ agar including a nitrogen-reduced
497 version of FAB medium (500 µM $MgSO_4 \cdot 7H_2O$; 250 µM KH_2PO_4; 250 µM KCl; 250 µM
498 $CaCl_2 \cdot 2H_2O$; 100 µM KNO_3; 25 µM Fe-EDDHA; 50 µM H_3BO_3; 25 µM $MnSO_4 \cdot H_2O$; 10 µM
499 $ZnSO_4 \cdot 7H_2O$; 0.5 µM $Na_2MoO_4 \cdot 2H_2O$; 0.2 µM $CuSO_4 \cdot 5H_2O$; 0.2 µM $CoCl_2 \cdot 6H_2O$; pH 5.7) in
500 square plates for 1 week before transferring to a growth chamber at 24 °C under a 16 h/8 h
501 light/dark regime (275 µmol·m-2·s-1) in Weck jars (SKU 745 or 743) containing 300 ml of
502 sand:vermiculite mixture (2:1) and 30 ml of nitrogen-reduced FAB medium containing
503 *Mesorhizobium loti* MAFF 303099 *Ds*Red (*M. loti Ds*Red; **Figure 3 - 4**; **Figure 3 – figure**
504 **supplement 1, 3, 4, 5, 6 and 7** and **Figure 4 – figure supplement 1**) or *M. loti* R7A CFP
505 (**Figure 2**) set to a final optical density at 600 nm (OD_{600}) of 0.05. For **Figure 2 – figure**
506 **supplement 1** and **Figure 3 – figure supplement 8**, plants were grown in Weck jars (SKU
507 745 or 743) containing 300 ml of sand:vermiculite mixture (2:1) and 60 ml of nitrogen-reduced
508 FAB medium containing *M. loti Ds*Red or MAFF 303099 *lacZ* (*M. loti lacZ*) (OD_{600} = 0.01).

509 **Cloning and DNA constructs**

510 For the construction of promoter:*NIN* fusions for complementation experiments (**Figure 3**;
511 **Figure 3 – figure supplement 2, 3, 4, 5, 6 and 7** and **Figure 4 – figure supplement 1**), the
512 *NIN* genomic sequence without the 5' and 3'UTRs served as a cloning module. A 3 kb region
513 of the *L. japonicus NIN* promoter plus the 244 bp *NIN* 5'UTR was cloned from *L. japonicus*
514 Gifu and used for complementation experiments (**Figure 3**; **Figure 3 – figure supplement 2,**
515 **3, 4, 5, 6 and 7**), dual luciferase assays (**Figure 1 – figure supplement 5**), fluorimetric GUS
516 assay (**Figure 1 – figure supplement 6**) and promoter activity analysis (**Figure 2**; **Figure 2 –**
517 **figure supplement 1**). For all the other versions of the *L. japonicus NIN* promoter tested
518 (**Figure 2, 3 and 4**; **Figure 1 – figure supplement 6**; **Figure 2 – figure supplement 1**; **Figure**
519 **3 – figure supplement 1, 2, 3, 4, 5, 6, 7 and 8** and **Figure 4 – figure supplement 1**), the
520 *LjNIN* minimal promoter (98 bp (Singh et al., 2014)) plus the *LjNIN* 5'UTR was fused to 3' end
521 of the promoter. A 472 bp region containing multiple cytokinin response elements and highly
522 conserved in eight legume species was identified 5' of the *NIN* transcriptional start site by Liu
523 et al. (Liu et al., 2019). We used this conserved region of 472 bp from *L. japonicus* and added
524 flanking regions (192 bp upstream and 366 bp downstream; 2399 bp upstream and 2231 bp
525 downstream, respectively) to obtain cytokinin element-containing regions of 1 kb and 5 kb
526 (CE_{1kb} and CE_{5kb}, respectively). The *Solanum lycopersicum* gene ID Solyc01g112190.2.1 was
527 identified as the closest homologue of *LjNIN* gene based on phylogenetic analysis (Griesmann
528 et al., 2018), and is referred to as *SlNIN*. A 3 kb region of the *SlNIN* promoter plus the 238 bp
529 *SlNIN* 5'UTR was cloned from *S. lycopersicum* cv. "Moneymaker" and *PACE* or *mPACE*
530 (**Figure 1 – figure supplement 6A**) was inserted 184 bp upstream of the *SlNIN* 5'UTR and
531 used for complementation experiments (**Figure 4** and **Figure 4 – figure supplement 1**), dual

532 luciferase assays (**Figure 1 – figure supplement 5**) and fluorimetric GUS assay (**Figure 1 –**
533 **figure supplement 6**). A detailed description of constructs and a list of oligonucleotides can
534 be found in **Supplementary file 3** and **Supplementary file 4**, respectively. Constructs were
535 generated with the Golden Gate cloning system (Binder et al., 2014).

536 **Imaging**

537 Microscope and scanner settings as well as parameters for image acquisition are listed in
538 **Supplementary file 5**.

539 **Phenotypic analysis and quantification of infection events**

540 Infected and non-infected nodules were discriminated by the presence and absence of a
541 *Ds*Red signal (representing *M. loti Ds*Red) detected or not detected inside of the nodules,
542 respectively. Presence or absence of bacteria was later confirmed by examination of sections
543 of representative nodules. Infection threads (ITs) and *M. loti* entrapments in root hairs were
544 detected by their *Ds*Red fluorescence (for microscope settings see **Supplementary file 5**).

545 For phenotypic analysis of *nin-15* (**Figure 3 – figure supplement 1C, F**), quantification was
546 performed 21 days after inoculation (dpi) with *M. loti Ds*Red as follows: (1) the total number of
547 nodules (including infected and non-infected) was determined under white light illumination
548 (WLI); (2) the number of infected nodules and root hair ITs were counted as described above.
549 Shoot dry weight was measured after drying the shoot at 60 °C for 1 h (**Figure 3 – figure**
550 **supplement 1E**).

551 For complementation experiments of *nin-2* and *nin-15* (**Figure 3 and 4**; **Figure 3 – figure**
552 **supplement 2, 3, 4, 5, 6 and 7** and **Figure 4 – figure supplement 1**), quantifications and
553 sectioning were performed 21 or 35 dpi with *M. loti Ds*Red with the microscope settings listed
554 in **Supplementary file 5** in the following order: (1) transgenic roots were identified by GFP
555 fluorescence-emanating nuclei with a GFP filter; (2) infected nodules were counted as
556 described above; (3) the total number of nodules (including infected and non-infected ones)
557 was then determined under WLI; (4) the number of non-infected nodules was calculated by
558 subtracting the number of infected nodules from the total number of nodules. To quantify
559 infection events in root hairs, the number of bacterial entrapment and ITs in root hairs were
560 counted on a 0.5 cm root piece for each transgenic root system, excised from a region where
561 bacterial accumulation was detected by *Ds*Red fluorescence. Sectioning was performed on
562 non-infected and infected nodules and the presence/absence of ITs and symbiosomes in
563 cortical cells was examined. Nodule primordia and nodules were embedded in 6% low-melting
564 agarose and sliced into 40 - 50 µm thick sections using a vibrating-blade microtome (Leica
565 VT1000 S).

566 **Transient expression in *Nicotiana benthamiana* leaves**

567 *Agrobacterium tumefaciens* strain AGL1 carrying promoter:reporter fusions on T-DNA were
568 infiltrated as described in (Cerri et al., 2012) with the acetosyringone concentration in the
569 infiltration buffer modified to 150 µM. *A. tumefaciens* strains AGL1 and GV3101 containing
570 plasmids $35S_{pro}$*:3xHA-Cyclops* (Singh et al., 2014) and $35S_{pro}$*:CCaMK*$^{1-314}$*-mOrange* (Takeda
571 et al., 2015), respectively, were co-infiltrated with the reporter constructs as indicated. An
572 AGL1 strain carrying a K9 plasmid constitutively expressing red fluorescent protein was used
573 as needed to equalize the density of the *A. tumefaciens* suspension infiltrated per leaf,
574 together with an *A. tumefaciens* strain carrying a plasmid for the expression of the viral P19
575 silencing suppressor to reduce post-transcriptional gene silencing (Voinnet et al., 2003)
576 (**Figure 1 – figure supplement 5 and 6**). *N. benthamiana* leaf discs with a diameter of 0.5
577 cm were harvested 60 hours post infiltration and used for quantitative fluorometric GUS assay
578 and dual luciferase assay.

579 **Dual luciferase assay**

580 The dual luciferase assay (**Figure 1 – figure supplement 5**) was based on the Dual-
581 Luciferase® reporter assay system (Promega). *N. benthamiana* leaf discs were ground to a
582 fine powder in liquid nitrogen in 2 ml Eppendorf Safe-Lock tube (1 leaf disc in each tube) and
583 subsequently incubated 5 min at room temperature (RT) with 200 µl of the Passive lysis buffer
584 (Promega catalog number E1910). The resulting crude leaf extract was centrifuged at 20,000
585 g for 2 min at RT. An aliquot of the supernatant was subjected to the dual luciferase assay
586 according to manufacturer's instruction for Promega Dual Luciferase Kit (catalog number
587 E1910) and chemiluminescence was quantified with a fluorescence plate reader (TECAN
588 Infinite® 200 PRO; TECAN Group Ltd.) in white 96 well plates (Greiner Bio-One International
589 GmbH). For each reporter construct, the promoter of interest was fused to the *Firefly* luciferase
590 gene and constitutively expressed *Renilla* luciferase from the same vector was used for
591 normalization (**Supplementary file 3**). The ratio of the two signals (*Firefly* luciferase signal to
592 the *Renilla* luciferase) was calculated and normalized to the vector control. A total number of
593 at least 4 biological and 2 technical replicates per indicated vector were analysed in at least 2
594 independent experiments.

595 **Quantitative fluorometric GUS assay and analysis**

596 Quantitative fluorometric GUS assays (**Figure 1 – figure supplement 6**) were performed as
597 described (Jefferson, 1987) adapted to 96 well format. A total number of 4 to 8 leaf discs per
598 indicated vector combination were analysed in at least 2 independent experiments performed
599 in different weeks.

600 **Promoter activity analysis**

601 For promoter activity analyses with the *GUS* reporter gene (**Figure 2 – figure supplement**
602 **1A - F** and **Figure 3 – figure supplement 8**), transgenic nodule primordia and nodules were
603 excised 10 - 14 dpi with *M. loti Ds*Red and stained for GUS activity using 5-bromo-4-chloro-3-
604 indolyl-beta-D-glucuronic acid (X-Gluc; x-gluc.com) as catalytic substrate (Cerri et al., 2012)
605 for 3 h at 37 °C. To visualise the root hair ITs together with the promoter activity with the *GUS*
606 reporter gene (**Figure 2 – figure supplement 1G**), plants with transgenic root systems were
607 inoculated with *M. loti lacZ*. Transgenic roots were first stained for GUS activity with X-Gluc
608 for 3 h at 37 °C and then for *lacZ* expression with Magenta-Gal for 18 h at 28 °C (as described
609 in (Cerri et al., 2012), which were visualized in blue and purple colour after staining,
610 respectively). For promoter activity analyses with fluorescent reporters (**Figure 2**), transgenic
611 root systems were harvested 10 - 14 dpi. Nodule primordia with bacterial infection at stage 3
612 or 4 (for stage description see main text) were selected by locating the CFP signal (*M. loti* R7A
613 CFP) via rapid (around 10 seconds) Z-stack analysis with the confocal light scanning
614 microscope (**Supplementary file 5**). Nodule primordia were sectioned and imaged as
615 described in **Supplementary file 5**.

616 **Data visualization and statistical analysis**

617 Statistical analyses and data visualization were performed with RStudio 1.1. 383 (RStudio
618 Inc.). Boxplots were used to display data in **Figure 4**; **Figure 1 – figure supplement 5 and**
619 **6**; **Figure 3 – figure supplement 1, 2, 3, 4 and 7** and **Figure 4 – figure supplement 1** (thick
620 black or white lines: median; box: interquartile range; whiskers: lowest and highest data point
621 within 1.5 interquartile range (IQR); black filled circles, data points inside 1.5 IQR; white filled
622 circles, data points outside 1.5 IQR of the upper/lower quartile). The R package "beeswarm"
623 with the method "center" was used to plot the individual data points for the boxplots
624 (http://CRAN.R-project.org/package=beeswarm). The R package "agricolae" was used to
625 perform ANOVA statistical analysis with post hoc Tukey and statistical results were displayed
626 in small letters where different letters indicated statistical significance
627 (https://cran.rproject.org/web/packages/agricolae/index.html). Tests applied are stated in the
628 figure legend.

629 **Data availability:**

630 Correspondence and requests for materials should be addressed to Martin Parniske.

631 **Acknowledgments:**

632 The authors thank David Chiasson for providing the MAFF 303099 *lacZ* and R7A CFP *M. loti*
633 strains, Niels Sandal and Jens Stougaard for providing the *nin-15* mutant. This project has
634 received funding from the European Research Council (ERC) under the European Union's

Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 340904 (EvolvingNodules) which supported the work of MG and XG. MP acknowledges funding from the German Research Foundation (DFG) in the context of the SFB924 'Molecular mechanisms regulating yield and yield stability in plants' which supported the work of REA and CC and the ANR-DFG project 'COME-IN' which supported the work of CC. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No H2020-MSCA-IF-2015-703186 to KV and postdoctoral fellowship from Alexander von Humboldt Foundation to KV. PMD and JK belong to the LRSV laboratory, which is part of the TULIP Laboratoire d'Excellence (LABEX) (ANR-10-LABX-41). MH received funding from the JSPS KAKENHI Grant Number 17H06472.

Author contributions:

MG performed bioinformatic analysis of *NIN* promoters and discovered *PACE* presented in **Figure 1** and **Figure 1 – figure supplement 1A - D**. CC performed *in vivo* expression analysis presented in **Figure 2** and prepared all confocal and light microscopy images of root hairs and nodule sections (**Figure 3 and 4**; **Figure 3 – figure supplement 5 and 6**; **Figure 4 – figure supplement 1**). Complementation experiments of *nin-15* were performed by REA (**Figure 3B - D**, **Figure 4A, 4C**; **Figure 3 – figure supplement 7**; **Figure 4 – figure supplement 1A, 1B**), CC (**Figure 3 and 4**; **Figure 3 – figure supplement 7C**; **Figure 4 – figure supplement 1**) and XG (**Figure 3D and 4B - C**; **Figure 3 – figure supplement 7**; **Figure 4 – figure supplement 1C, 1D**). Complementation experiments of *nin-2* were performed by CC and XG (**Figure 3A**; **Figure 3 – figure supplement 2, 3, 4, 5 and 6**). REA, CC and XG performed *nin-15* mutant phenotyping (**Figure 3 – figure supplement 1**). XG performed transient expression assays presented in **Figure 1 – figure supplement 6** and promoter expression analysis in **Figure 2 – figure supplement 1** and **Figure 3 – figure supplement 8**. KV performed transient expression assays presented in **Figure 1 – figure supplement 5** and drafted **Figure 1**. JK and PMD performed motif search in *ERN1*, *NIN* and *RAM1* promoters presented in **Figure 1 – figure supplement 1E, 2, 3 and 4**. CC, XG, REA, KV, PMD, MH, MG and MP formulated research hypothesis. CC, XG, REA, KV, MG and MP designed experiments. MP conceived and supervised the project. XG and MP coordinated research activities. MP, CC and XG wrote the manuscript and XG finalized all figures with inputs and comments from co-authors.

Competing interests:

Authors declare no competing interests.

References:

Bailey, T. L., & Elkan, C. (1994). Fitting a mixture model by expectation maximization to

670 discover motifs in biopolymers. *Proc. 2nd Int. Conf. on Intelligent Systems for Molecular*
671 *Biology*, *2*, 28–36. https://doi.org/citeulike-article-id:878292

672 Banba, M., Gutjahr, C., Miyao, A., Hirochika, H., Paszkowski, U., Kouchi, H., & Imaizumi-
673 Anraku, H. (2008). Divergence of evolutionary ways among common sym genes:
674 CASTOR and CCaMK show functional conservation between two symbiosis systems
675 and constitute the root of a common signaling pathway. *Plant and Cell Physiology*,
676 *49*(11), 1659–1671. https://doi.org/10.1093/pcp/pcn153

677 Bell, C. D., Soltis, D. E., & Soltis, P. S. (2010). The age and diversification of the
678 angiosperms re-revisited. *American Journal of Botany*, *97*(8), 1296–1303.
679 https://doi.org/10.3732/ajb.0900346

680 Binder, A., Lambert, J., Morbitzer, R., Popp, C., Ott, T., Lahaye, T., & Parniske, M. (2014). A
681 modular plasmid assembly kit for multigene expression, gene silencing and silencing
682 rescue in plants. *PLoS ONE*, *9*(2). https://doi.org/10.1371/journal.pone.0088218

683 Buecker, C., & Wysocka, J. (2012). Enhancers as information integration hubs in
684 development: Lessons from genomics. *Trends in Genetics*, *28*(6), 276–284.
685 https://doi.org/10.1016/j.tig.2012.02.008

686 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden,
687 T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*, 1–9.
688 https://doi.org/10.1186/1471-2105-10-421

689 Carvalho, T. L. G., Balsemão-Pires, E., Saraiva, R. M., Ferreira, P. C. G., & Hemerly, A. S.
690 (2014). Nitrogen signalling in plant interactions with associative and endophytic
691 diazotrophic bacteria. *Journal of Experimental Botany*, *65*(19), 5631–5642.
692 https://doi.org/10.1093/jxb/eru319

693 Cerri, M. R., Frances, L., Laloum, T., Auriac, M.-C., Niebel, A., Oldroyd, G. E. D., Barker, D.
694 G., Fournier, J., & de Carvalho-Niebel, F. (2012). *Medicago truncatula* ERN
695 transcription factors: regulatory interplay with NSP1/NSP2 GRAS factors and
696 expression dynamics throughout rhizobial Infection. *Plant Physiology*, *160*(4), 2155–
697 2172. https://doi.org/10.1104/pp.112.203190

698 Cerri, M. R., Wang, Q., Stolz, P., Folgmann, J., Frances, L., Katzer, K., Li, X., Heckmann, A.
699 B., Wang, T. L., Downie, J. A., Klingl, A., de Carvalho-Niebel, F., Xie, F., & Parniske, M.
700 (2017). The *ERN1* transcription factor gene is a target of the CCaMK/CYCLOPS
701 complex and controls rhizobial infection in *Lotus japonicus*. *New Phytologist*, *215*(1),
702 323–337. https://doi.org/10.1111/nph.14547

703 Charpentier, M., Bredemeier, R., Wanner, G., Takeda, N., Schleiff, E., & Parniske, M.

(2008). *Lotus japonicus* Castor and Pollux are ion channels essential for perinuclear calcium spiking in legume root endosymbiosis. *Plant Cell*, *20*(12), 3467–3479. https://doi.org/10.1105/tpc.108.063255

Chen, C., Gao, M., Liu, J., & Zhu, H. (2007). Fungal symbiosis in rice requires an ortholog of a legume common symbiosis gene encoding a Ca^{2+}/calmodulin-dependent protein kinase. *Plant Physiology*, *145*(4), 1619–1628. https://doi.org/10.1104/pp.107.109876

Clavijo, F., Diedhiou, I., Vaissayre, V., Brottier, L., Acolatse, J., Moukouanga, D., Crabos, A., Auguy, F., Franche, C., Gherbi, H., Champion, A., Hocher, V., Barker, D., Bogusz, D., Tisa, L. S., & Svistoonoff, S. (2015). The *Casuarina NIN* gene is transcriptionally activated throughout *Frankia* root infection as well as in response to bacterial diffusible signals. *New Phytologist*, *208*(3), 887–903. https://doi.org/10.1111/nph.13506

D'Haeze, W., Gao, M., De Rycke, R., Van Montagu, M., Engler, G., & Holsters, M. (1998). Roles for azorhizobial nod factors and surface polysaccharides in intercellular invasion and nodule penetration, respectively. *Molecular Plant-Microbe, 11*(10), 999–1008. https://doi.org/10.1094/MPMI.1998.11.10.999

De Faria, S. M., Hay, G. T., & Sprent, J. I. (1988). Entry of rhizobia into roots of *Mimosa scabrella* bentham occurs between epidermal cells. *Microbiology*, *134*(8), 2291–2296. https://doi.org/10.1099/00221287-134-8-2291

Delaux, P. M., Radhakrishnan, G. V., Jayaraman, D., Cheema, J., Malbreil, M., Volkening, J. D., Sekimoto, H., Nishiyama, T., Melkonian, M., Pokorny, L., Rothfels, C. J., Sederoff, H. W., Stevenson, D. W., Surek, B., Zhang, Y., Sussman, M. R., Dunand, C., Morris, R. J., Roux, C., … Ane, J. M. (2015). Algal ancestor of land plants was preadapted for symbiosis. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(43), 13390–13395. https://doi.org/10.1073/pnas.1515426112

Delaux, P. M., Varala, K., Edger, P. P., Coruzzi, G. M., Pires, J. C., & Ané, J. M. (2014). Comparative phylogenomics uncovers the impact of symbiotic associations on host genome evolution. *PLoS Genetics*, *10*(7). https://doi.org/10.1371/journal.pgen.1004487

Doyle, J. J. (2011). Phylogenetic perspectives on the origins of nodulation. *Molecular Plant-Microbe Interactions*, *24*(11), 1289–1295. https://doi.org/10.1094/MPMI-05-11-0114

Doyle, J. J. (2016). Chasing unicorns: nodulation origins and the paradox of novelty. *American Journal of Botany*, *103*(11), 1865–1868. https://doi.org/10.3732/ajb.1600260

Finan, T. M., Hirsch, A. M., Leigh, J. A., Johansen, E., Kuldau, G. A., Deegan, S., Walker, G. C., & Signer, E. R. (1985). Symbiotic mutants of *Rhizobium meliloti* that uncouple plant from bacterial differentiation. *Cell*, *40*(4), 869–877. https://doi.org/10.1016/0092-

8674(85)90346-0

Gage, D. J. (2002). Analysis of infection thread development using GFP- and *Ds*Red-expressing *Sinorhizobium meliloti*. *Journal of Bacteriology*, *184*(24), 7042–7046. https://doi.org/10.1128/JB.184.24.7042-7046.2002

Gamborg, O. L., Miller, R. A., & Ojima, K. (1968). Nutrient requirements of suspension cultures of soybean root cells. *Experimental Cell Research*, *50*(1), 151–158. https://doi.org/10.1016/0014-4827(68)90403-5

Geurts, R., Xiao, T. T., & Reinhold-Hurek, B. (2016). What does it take to evolve a nitrogen-fixing endosymbiosis? *Trends in Plant Science*, *21*(3), 199–208. https://doi.org/10.1016/j.tplants.2016.01.012

Gossmann, J. A., Markmann, K., Brachmann, A., Rose, L. E., & Parniske, M. (2012). Polymorphic infection and organogenesis patterns induced by a *Rhizobium leguminosarum* isolate from *Lotus* root nodules are determined by the host genotype. *New Phytologist*, *196*(2), 561–573. https://doi.org/10.1111/j.1469-8137.2012.04281.x

Griesmann, M., Chang, Y., Liu, X., Song, Y., Haberer, G., Crook, M., Billault-Penneteau, B., Lauressergues, D., Keller, J., Imanishi, L., Roswanjaya, Y., Kohlen, W., Pujic, P., Battenberg, K., Alloisio, N., Liang, Y., Hilhorst, H., Salgado, M., Hocher, V., … Cheng, S. (2018). Phylogenomics reveals multiple losses of the nitrogen-fixing root nodule symbiosis. *Science*, *13*(361). https://doi.org/10.1126/science.aat1743

Gutjahr, C., Banba, M., Croset, V., An, K., Miyao, A., An, G., Hirochika, H., Imaizumi-Anraku, H., & Paszkowski, U. (2008). Arbuscular mycorrhiza-specific signaling in rice transcends the common symbiosis signaling pathway. *The Plant Cell*, *20*(11), 2989–3005. https://doi.org/10.1105/tpc.108.062414

Hirsch, S., Kim, J., Muñoz, A., Heckmann, A. B., Downie, J. A., & Oldroyd, G. E. D. (2009). GRAS proteins form a DNA binding complex to induce gene expression during nodulation signaling in *Medicago truncatula*. *Plant Cell*, *21*(2), 545–557. https://doi.org/10.1105/tpc.108.064501

Hoagland, D. R., & Arnon, D. I. (1950). The water-culture method for growing plants without soil. Circular California Agricultural Experiment Station 347, ed. 2, Berkeley: University of California.

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Le, S. V. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*, *35*(2), 518–522. https://doi.org/10.5281/zenodo.854445

771 Jefferson, R. A. (1987). Assaying chimeric genes in plants: The *GUS* gene fusion system.
772 *Plant Molecular Biology Reporter*, *5*(4), 387–405. https://doi.org/10.1007/BF02667740

773 Johansson, C., & Bergman, B. (1992). Early events during the establishment of the *Gunnera*
774 / *Nostoc* symbiosis. *Planta*, *188*(3), 403–413.

775 Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A., & Jermiin, L. S. (2017).
776 ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature*
777 *Methods*, *14*(6), 587–589. https://doi.org/10.1038/nmeth.4285

778 Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version
779 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4),
780 772–780. https://doi.org/10.1093/molbev/mst010

781 Kistner, C., & Parniske, M. (2002). Evolution of signal transduction in intracellular symbiosis.
782 *Trends in Plant Science*, *7*(11), 511–518. https://doi.org/10.1016/S1360-
783 1385(02)02356-7

784 Kvon, E. Z., Kamneva, O. K., Melo, U. S., Barozzi, I., Osterwalder, M., Mannion, B. J.,
785 Tissières, V., Pickle, C. S., Plajzer-Frick, I., Lee, E. A., Kato, M., Garvin, T. H.,
786 Akiyama, J. A., Afzal, V., Lopez-Rios, J., Rubin, E. M., Dickel, D. E., Pennacchio, L. A.,
787 & Visel, A. (2016). Progressive loss of function in a limb enhancer during snake
788 evolution. *Cell*, *167*(3), 633–642. https://doi.org/10.1016/j.cell.2016.09.028

789 LeBauer, D., & Treseder, K. (2008). Nitrogen limitation of net primary productivity in
790 terrestrial ecosystems is globally distributed. *Ecology*, *89*(7), 371–379.
791 https://doi.org/10.1007/BF00153104

792 Li, H. L., Wang, W., Mortimer, P. E., Li, R. Q., Li, D. Z., Hyde, K. D., Xu, J. C., Soltis, D. E., &
793 Chen, Z. D. (2015). Large-scale phylogenetic analyses reveal multiple gains of
794 actinorhizal nitrogen-fixing symbioses in angiosperms associated with climate change.
795 *Scientific Reports*, *5*(January), 1–8. https://doi.org/10.1038/srep14023

796 Liu, J., & Bisseling, T. (2020). Evolution of *NIN* and *NIN-like* genes in relation to nodule
797 symbiosis. *Genes*, *11*(7), 1–15. https://doi.org/10.3390/genes11070777

798 Liu, J., Rutten, L., Limpens, E., Van Der Molen, T., Van Velzen, R., Chen, R., Chen, Y.,
799 Geurts, R., Kohlen, W., Kulikova, O., & Bisseling, T. (2019). A remote *cis*-regulatory
800 region is required for nin expression in the pericycle to initiate nodule primordium
801 formation in *Medicago truncatula*. *Plant Cell*, *31*(1), 68–83.
802 https://doi.org/10.1105/tpc.18.00478

803 Madsen, L. H., Tirichine, L., Jurkiewicz, A., Sullivan, J. T., Heckmann, A. B., Bek, A. S.,

Ronson, C. W., James, E. K., & Stougaard, J. (2010). The molecular network governing nodule organogenesis and infection in the model legume *Lotus japonicus*. *Nature Communications*, *1*(1). https://doi.org/10.1038/ncomms1009

Markmann, K., Giczey, G., & Parniske, M. (2008). Functional adaptation of a plant receptor-kinase paved the way for the evolution of intracellular root symbioses with bacteria. *PLoS Biology*, *6*(3), 0497–0506. https://doi.org/10.1371/journal.pbio.0060068

Masson-Boivin, C., & Sachs, J. L. (2018). Symbiotic nitrogen fixation by rhizobia — the roots of a success story. *Current Opinion in Plant Biology*, *44*, 7–15. https://doi.org/10.1016/j.pbi.2017.12.001

Murray, J. D., Karas, B. J., Sato, S., Tabata, S., Amyot, L., & Szczyglowski, K. (2007). A cytokinin perception mutant colonized by *Rhizobium* in the absence of nodule organogenesis. *Science*, *305*(5808), 101–104.

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Ott, T., Van Dongen, J. T., Günther, C., Krusell, L., Desbrosses, G., Vigeolas, H., Bock, V., Czechowski, T., Geigenberger, P., & Udvardi, M. K. (2005). Symbiotic leghemoglobins are crucial for nitrogen fixation in legume root nodules but not for general plant growth and development. *Current Biology*, *15*(6), 531–535. https://doi.org/10.1016/j.cub.2005.01.042

Parniske, M. (2000). Intracellular accommodation of microbes by plants: a common developmental program for symbiosis and disease? *Curr Opin Plant Biol*, *3*, 320–328.

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

Pawlowski, K., & Demchenko, K. N. (2012). The diversity of actinorhizal symbiosis. *Protoplasma*, *249*(4), 967–979. https://doi.org/10.1007/s00709-012-0388-4

Perrine-Walker, F. M., Lartaud, M., Kouchi, H., & Ridge, R. W. (2014). Microtubule array formation during root hair infection thread initiation and elongation in the *Mesorhizobium-Lotus* symbiosis. *Protoplasma*, *251*(5), 1099–1111. https://doi.org/10.1007/s00709-014-0618-z

Peters, G., & Meeks, J. (1989). The *Azolla-Anabaena* symbiosis : basic biology. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, *40*, 193–210.

837 Pimprikar, P., Carbonnel, S., Paries, M., Katzer, K., Klingl, V., Bohmer, M. J., Karl, L., Floss,
838 D. S., Harrison, M. J., Parniske, M., & Gutjahr, C. (2016). A CCaMK-CYCLOPS-DELLA
839 complex activates transcription of *RAM1* to regulate arbuscule branching. *Current*
840 *Biology*, *26*(8), 987–998. https://doi.org/10.1016/j.cub.2016.01.069

841 Redecker, D., Kodner, R., & Graham, L. E. (2000). Glomalean fungi from the Ordovician.
842 *Science*, *289*(5486), 1920–1921. https://doi.org/8819 [pii]

843 Remy, W., Taylort, T. N., Hass, H., & Kerp, H. (1994). Four hundred-million-year-old
844 vesicular arbuscular mycorrhizae. *Proceedings of National Academy of Sciences, USA*,
845 *91*(December), 11841–11843. https://doi.org/10.1073/pnas.91.25.11841

846 Schauser, L., Roussis, A., Stiller, J., & Stougaard, J. (1999). A plant regulator controlling
847 development of symbiotic root nodules. *Nature*, *402*(6758), 191–195.
848 https://doi.org/10.1038/46058

849 Singh, S., Katzer, K., Lambert, J., Cerri, M., & Parniske, M. (2014). CYCLOPS, A DNA-
850 binding transcriptional activator, orchestrates symbiotic root nodule development. *Cell*
851 *Host and Microbe*, *15*(2), 139–152. https://doi.org/10.1016/j.chom.2014.01.011

852 Soltis, D. E., Soltis, P. S., Morgant, D. R., Swensent, S. M., Mullin, B. C., Dowdi, J. M., &
853 Peter, G. M. (1995). Chloroplast gene sequence data suggest a single origin of the
854 predisposition for symbiotic nitrogen fixation in angiosperms. *Proceedings of National*
855 *Academy of Sciences, USA*, *92*(March), 2647–2651.
856 https://doi.org/10.1073/pnas.92.7.2647

857 Soyano, T., & Hayashi, M. (2014). Transcriptional networks leading to symbiotic nodule
858 organogenesis. *Current Opinion in Plant Biology*, *20*, 146–154.
859 https://doi.org/10.1016/j.pbi.2014.07.010

860 Soyano, T., Kouchi, H., Hirota, A., & Hayashi, M. (2013). NODULE INCEPTION directly
861 targets *NF-Y* bubunit genes to regulate essential processes of root nodule development
862 in *Lotus japonicus*. *PLoS Genetics*, *9*(3). https://doi.org/10.1371/journal.pgen.1003352

863 Sprent, J. I. (2007). Evolving ideas of legume evolution and diversity: A taxonomic
864 perspective on the occurrence of nodulation: Tansley review. *New Phytologist*, *174*(1),
865 11–25. https://doi.org/10.1111/j.1469-8137.2007.02015.x

866 Sprent, J. I., Ardley, J., & James, E. K. (2017). Biogeography of nodulated legumes and their
867 nitrogen-fixing symbionts. *New Phytologist*, *215*(1), 40–56.
868 https://doi.org/10.1111/nph.14474

869 Takeda, N., Handa, Y., Tsuzuki, S., Kojima, M., Sakakibara, H., & Kawaguchi, M. (2015).

870 Gibberellins interfere with symbiosis signaling and gene expression and alter
871 colonization by arbuscular mycorrhizal fungi in *Lotus japonicus*. *Plant Physiology*,
872 *167*(2), 545–557. https://doi.org/10.1104/pp.114.247700

873 Van Spronsen, P. C., Grønlund, M., Bras, C. P., Spaink, H. P., & Kijne, J. W. (2001). Cell
874 biological changes of outer cortical root cells in early determinate nodulation. *Molecular*
875 *Plant-Microbe Interactions*, *14*(7), 839–847.
876 https://doi.org/10.1094/MPMI.2001.14.7.839

877 van Velzen, R., Doyle, J. J., & Geurts, R. (2019). A resurrected scenario: single gain and
878 massive loss of nitrogen-fixing nodulation. *Trends in Plant Science*, *24*(1), 49–57.
879 https://doi.org/10.1016/j.tplants.2018.10.005

880 van Velzen, R., Holmer, R., Bu, F., Rutten, L., van Zeijl, A., Liu, W., Santuari, L., Cao, Q.,
881 Sharma, T., Shen, D., Roswanjaya, Y., Wardhani, T. A. K., Kalhor, M. S., Jansen, J.,
882 van den Hoogen, J., Güngör, B., Hartog, M., Hontelez, J., Verver, J., … Geurts, R.
883 (2018). Comparative genomics of the nonlegume *Parasponia* reveals insights into
884 evolution of nitrogen-fixing rhizobium symbioses. *Proceedings of the National Academy*
885 *of Sciences*, *115*(20), E4700–E4709. https://doi.org/10.1073/pnas.1721395115

886 Vernié, T., Kim, J., Frances, L., Ding, Y., Sun, J., Guan, D., Niebel, A., Gifford, M. L., de
887 Carvalho-Niebel, F., & Oldroyd, G. E. D. (2015). The NIN transcription factor
888 coordinates diverse nodulation programs in different tissues of the *Medicago truncatula*
889 root. *The Plant Cell*, *27*(12), 3410–3424. https://doi.org/10.1105/tpc.15.00461

890 Voinnet, O., Rivas, S., Mestre, P., & Baulcombe, D. (2003). An enhanced transient
891 expression system in plants based on suppression of gene silencing by the p19 protein
892 of tomato bushy stunt virus. *The Plant Journal*, *33*, 949–956.

893 Wang, H., Moore, M. J., Soltis, P. S., Bell, C. D., Brockington, S. F., Alexandre, R., Davis, C.
894 C., Latvis, M., Manchester, S. R., & Soltis, D. E. (2009). Rosid radiation and the rapid
895 rise of angiosperm-dominated forests. *Proceedings of the National Academy of*
896 *Sciences*, *106*(10), 3853–3858. https://doi.org/10.1073/pnas.0813376106

897 Wittkopp, P., & Kalay, G. (2011). *Cis*-regulatory elements: molecular mechanisms and
898 evolutionary processes underlying divergence. *Nat Rev Genet*, *13*(1), 59–69.
899 https://doi.org/10.1038/nrg3095

900 Xiao, A., Yu, H., Fan, Y., Kang, H., Ren, Y., Huang, X., Gao, X., Wang, C., Zhang, Z., Zhu,
901 H., & Cao, Y. (2020). Transcriptional regulation of *NIN* expression by IPN2 is required
902 for root nodule symbiosis in *Lotus japonicus*. *New Phytologist*, *227*(2), 513–528.
903 https://doi.org/10.1111/nph.16553

Yano, K., Yoshida, S., Muller, J., Singh, S., Banba, M., Vickers, K., Markmann, K., White, C., Schuller, B., Sato, S., Asamizu, E., Tabata, S., Murooka, Y., Perry, J., Wang, T. L., Kawaguchi, M., Imaizumi-Anraku, H., Hayashi, M., & Parniske, M. (2008). CYCLOPS, a mediator of symbiotic intracellular accommodation. *Proceedings of the National Academy of Sciences*, *105*(51), 20540–20545. https://doi.org/10.1073/pnas.0806858105

Yoon, H. J., Hossain, M. S., Held, M., Hou, H., Kehl, M., Tromas, A., Sato, S., Tabata, S., Andersen, S. U., Stougaard, J., Ross, L., & Szczyglowski, K. (2014). *Lotus japonicus SUNERGOS1* encodes a predicted subunit A of a DNA topoisomerase VI that is required for nodule differentiation and accommodation of rhizobial infection. *Plant Journal*, *78*(5), 811–821. https://doi.org/10.1111/tpj.12520

Yoro, E., Suzaki, T., Toyokura, K., Miyazawa, H., Fukaki, H., & Kawaguchi, M. (2014). A positive regulator of nodule organogenesis, NODULE INCEPTION, acts as a negative regulator of rhizobial infection in *Lotus japonicus*. *Plant Physiology*, *165*(2), 747–758. https://doi.org/10.1104/pp.113.233379

Zhu, H., Chen, T., Zhu, M., Fang, Q., Kang, H., Hong, Z., & Zhang, Z. (2008). A novel ARID DNA-binding protein interacts with SymRK and is expressed during early nodule development in *Lotus japonicus*. *Plant Physiology*, *148*(1), 337–347. https://doi.org/10.1104/pp.108.119164

Additional references:

Handberg, K. & Stougaard, J. (1992). *Lotus japonicus*, an autogamous, diploid legume species for classical and molecular genetics. *Plant J*ournal, *2*(4), 487–496. https://doi.org/10.1111/j.1365-313X.1992.00487.x

Leong, J. M., Nunes-Duby, S., Lesser, C. F., Youderian, P., Susskind, M. M., Landy, A. (1985). The Φ80 and P22 attachment sites. Primary structure and interaction with *Escherichia coli* integration host factor. *Journal of Biological Chemistry 260*(7), 4468–4477.

Maekawa, T., Maekawa-Yoshikawa, M., Takeda, N., Imaizumi-Anraku, H., Murooka, Y., Hayashi, M. (2009). Gibberellin controls the nodulation signaling pathway in *Lotus japonicus*. *Plant Journal, 58*(2), 183–194. https://doi.org/10.1111/j.1365-313X.2008.03774.x

Małolepszy, A., Mun, T., Niels, S., Gupta, V., Dubin, M., Urbański, D., Shah, N., Bachmann, A., Fukai, E., Hirakawa, H., Tabata, S., Nadzieja, M., Markmann, K., Su, J., Umehara,

Y., Soyano, T., Miyahara, A., Sato, S., Hayashi, M., Stougaard, J., Andersen, S. U. (2016). The *LORE1* insertion mutant resource. *Plant Journal, 88*(2), 306–317. https://doi.org/10.1111/tpj.13243

Stevens, P. F. (2001 onwards). Angiosperm phylogeny website. Version 14, July 2017 http://www.mobot.org/MOBOT/research/APweb/.

943 **Supplementary Figures:**

944 **Figure 1 – figure supplement 1 | Discovery of *PACE* by MEME analyses.** (**A - D**)
945 Consensus sequence of the Position Weight Matrix identified by MEME analyses using the
946 regions upstream of the translational start site (ATG) thus representing the promoters and
947 5'UTRs of *NIN* and *NLP* genes from 37 angiosperm species (see Methods): (**A**) in a
948 discriminative search for a motif that is present in 3 kb upstream regions of the *NIN* genes
949 from nodulating FaFaCuRo species, but absent in 3 kb upstream regions of the *NIN* genes
950 from species outside of the FaFaCuRo clade and absent in the 3 kb upstream regions
951 of *NLP* genes; (**B**) in an independent, non-discriminative search in the 3 kb upstream regions
952 of the *NIN* genes from nodulating species revealing that the most conserved nucleotides span
953 a larger region than (**A**); (**C**) the resulted most conserved 29 nucleotides derived from
954 upstream regions of *NIN* genes of nodulating FaFaCuRo species; (**D**) the resulted most
955 conserved 29 nucleotides (*PACE*) derived from the upstream region of one representative *NIN*
956 gene per species of nodulating FaFaCuRo species. (**E**) Motif analyses by FIMO using the
957 upstream regions of the *NIN* and *NLP* genes from an expanded list of 163 species (see
958 Methods). Left: pruned tree from the whole *NLP* tree (demarcated in the black rectangle in
959 **Figure 1 – figure supplement 2**) corresponding to the *NIN* orthologs. Right: three versions
960 of consensus sequence (left to right, (**A, B and D**), respectively) were used to retrieve motifs
961 from the 5 kb upstream region of *NIN* genes *via* FIMO search, the output of which is displayed
962 underneath the consensus sequences. Sequence names of nodulating species are coloured
963 in blue. Blank lines represent the absence of significant motifs.

964 **Figure 1 – figure supplement 2 | *PACE* is present exclusively in the FaFaCuRo clade.**
965 Motif analyses by FIMO using the using the regions upstream of the translational start site
966 (ATG) thus representing the promoters and 5'UTRs of the *NIN* and *NLP* genes from an
967 expanded list of 163 species (see Methods). Left: Maximum likelihood tree of *NLP* family
968 (model: GTR+F+R10) of 163 species (**Figure 1 – Source Data 2**). Tree was rooted on the
969 Charophytes algae *Klebsormidium nitens*. Black box demarcates the area that is displayed in
970 **Figure 1 – figure supplement 1E**. Blank lines represent the absence of significant motifs.
971 Sequence names of nodulating species are coloured in blue. The different NLP clades are
972 indicated by branch colours and named according to the *Arabidopsis thaliana* nomenclature.
973 Right: three versions of consensus sequence (left to right, **Figure 1 – figure supplement 1A,**
974 **1B and 1D**, respectively) were used to retrieve motifs from the 5 kb upstream region of *NIN*
975 genes via FIMO search, the output of which is displayed underneath the consensuses
976 sequences.

977 **Figure 1 – figure supplement 3 | The presence of a motif encompassing the previously**
978 **identified Cyclops binding site within the regions upstream of the translational start**

979 **site (ATG) thus representing the promoters and 5'UTRs of *ERN1* genes extends beyond**
980 **the FaFaCuRo clade.** (**A**) Maximum likelihood of *ERN1/ERN2* (model: JTT+F+R5) from 87
981 species (**Supplementary file 1**) based on protein sequences. Tree was rooted using
982 sequences from the *ERN1/ERN2* paralog *ERN3* (outgroup clade collapse and indicated by
983 grey triangle marked "Outgroup"). Species belonging to the monophyletic FaFaCuRo clade
984 (Stevens, 2017) are indicated by thicker black branches. *ERN1* and *ERN2* subclades are
985 indicated. Sequence names of nodulating species are coloured in blue. These conserved
986 motifs identified by MEME analyses in the 3 kb upstream regions were aligned and displayed
987 next to species names. While the presence of a conserved Cyclops binding site containing
988 motif in species outside of the FaFaCuRo clade is consistent with the role of *RAM1* in
989 arbuscular mycorrhizal symbiosis, a function of *ERN1* in arbuscular mycorrhizal symbiosis has
990 not been described and suggests a yet unidentified function of *ERN1* outside of the FaFaCuRo
991 clade. (**B**) Consensus sequence of the Position Weight Matrix identified by MEME analyses
992 of all the species. (**C**) Consensus sequence of the Position Weight Matrix identified by MEME
993 analyses when analysing only the nodulating species.

994 **Figure 1 – figure supplement 4 | The presence of a motif encompassing the previously**
995 **identified Cyclops binding site within the regions upstream of the translational start**
996 **site (ATG) thus representing the promoters and 5'UTRs of *RAM1* genes extends beyond**
997 **the FaFaCuRo clade.** (**A**) Maximum likelihood of *RAM1* (model: JTT+F+R4) from 87 species
998 (**Supplementary file 1**) engaging in arbuscular mycorrhizal symbiosis. Tree was rooted on
999 the early-diverging angiosperm *Amborella trichopoda*. Species belonging to the monophyletic
1000 FaFaCuRo clade (Stevens, 2017) are indicated by thicker black branches. Sequence names
1001 of nodulating species are coloured in blue. These conserved motifs identified by MEME
1002 analyses in the 3 kb upstream regions are aligned and displayed next to species names. Blank
1003 lines represent the absence of identified motif due to genome contiguity issues. (**B**)
1004 Consensus sequence of the Position Weight Matrix identified by MEME analyses when
1005 analysing all the mycorrhizal species. (**C**) Consensus sequence of the Position Weight Matrix
1006 identified by MEME analyses when analysing only the nodulating species.

1007 **Figure 1 – figure supplement 5 | Transcriptional activation of *NIN promoter:Firefly***
1008 ***luciferase* reporter gene by CCaMK$^{1-314}$/Cyclops is restricted to *NIN* promoters from**
1009 **species of the FaFaCuRo clade.** *Nicotiana benthamiana* leaf cells were transformed with T-
1010 DNAs carrying a *Firefly luciferase* reporter gene driven by either of the indicated promoters in
1011 tandem with the *AtACT2$_{pro}$*:*Renilla luciferase* reporter fusion that provides a quantitative
1012 internal standard. (**A**) List of species within the FaFaCuRo clade (light red shade) and outside
1013 (light grey shade) and abbreviations. (**B**) Reporter gene activation by *L. japonicus* CCaMK^{1-}
1014 ^{314}/Cyclops via *NIN* promoters (*NIN$_{pro}$*) originating from listed species. (**C**) Comparison of the

1015 transactivation potential of Cyclops versions from *L. japonicus* and *S. lycopersicum*. Note that
1016 the expression of the *Firefly luciferase* reporter gene driven by $LjNIN_{pro}$, the *RAM1* promoters
1017 from *L. japonicus* and *S. lycopersicum* ($LjRAM1_{pro}$ and $SlRAM1_{pro}$, respectively) was induced
1018 in the presence of $CCaMK^{1-314}$/Cyclops regardless of the origin of Cyclops. In contrast, the
1019 transactivation failed with the *SlNIN* promoter (panel (**A**)). The applied statistical method was
1020 ANOVA with *post hoc* Tukey: (**B**), $F_{14,214} = 71.07$, $p < 2x10^{-16}$; (**C**), plots from left to right: $F_{5,18}$
1021 $= 20.58$, $p = 7.14x10^{-7}$; $F_{5,18} = 25.38$, $p = 1.45x10^{-7}$ and $F_{5,18} = 40.49$, $p = 3.55x10^{-9}$,
1022 respectively. Different small letters indicate significant differences.

1023 **Figure 1 – figure supplement 6 | *PACE* sequence variants from species across the**
1024 **FaFaCuRo clade were able to functionally replace *L. japonicus PACE* in a $LjNIN_{pro}$:*GUS***
1025 **reporter fusion.** *N. benthamiana* leaf cells were transformed with T-DNAs carrying a *GUS*
1026 reporter gene driven by either of the indicated promoters: (**A**) the *L. japonicus NIN* promoter
1027 (NIN_{pro}), the *LjNIN* promoter with *PACE* mutated or deleted (NIN_{pro}*::mPACE* and
1028 NIN_{pro}*::ΔPACE*, respectively), or *PACE* sequence variants from the nodulating FaFaCuRo
1029 species fused to the *LjNIN* minimal promoter ($NINmin_{pro}$); (**B**) chimeric promoters where
1030 *LjPACE* in the *LjNIN* promoter was replaced with either one of the *PACE* variants from species
1031 tested in (**A**) or from non-nodulating FaFaCuRo species including the *Juglans regia PACE*-
1032 like motif (*JrPACE-like*); (**C**) the *S. lycopersicum NIN* promoter ($SlNIN_{pro}$), the *SlNIN* promoter
1033 with *LjPACE* ($SlNIN_{pro}$*::PACE*) or *mPACE* ($SlNIN_{pro}$*::mPACE*) inserted. For species
1034 abbreviations see **Figure 1 – figure supplement 5A**. Note in (**A**) that the deletion or mutation
1035 of *PACE* in *LjNIN* promoter resulted in a drastic reduction in reporter gene expression and in
1036 (**C**) insertion of *LjPACE* but not *mPACE* into the *S. lycopersicum* promoter confers
1037 transactivation by $CCaMK^{1-314}$/Cyclops. The applied statistical method was ANOVA with *post*
1038 *hoc* Tukey: (**A**) $F_{20,144} = 51.38$, $p < 2x10^{-16}$; (**B**), $F_{18,166} = 149.1$, $p < 2x10^{-16}$; (**C**) $F_{7,62} = 30.5$, p
1039 $= 7.02x10^{-7}$. Different small letters indicate significant difference. n.d., not determined.

1040 **Figure 2 – figure supplement 1 | Spatio-temporal *GUS* expression driven by *PACE* and**
1041 **the *NIN* promoter in *L. japonicus* roots during the bacterial infection process.** *L.*
1042 *japonicus* wild-type hairy roots were transformed with T-DNAs carrying a $Ubq10_{pro}$*:NLS-GFP*
1043 transformation marker together with a *GUS* reporter gene driven by either of the indicated
1044 promoters: (**A**) the 3 kb *LjNIN* promoter (NIN_{pro}); the *LjNIN* promoter with *PACE* (**B**) mutated
1045 ($LjNIN_{pro}$*::mPACE*) or (**C**) deleted (NIN_{pro}*::ΔPACE*); (**D**) *PACE* fused to the *LjNIN* minimal
1046 promoter (*PACE:*$NINmin_{pro}$) or (**E**) the *LjNIN* minimal promoter ($NINmin_{pro}$). The progression
1047 of bacterial infection was determined by the *Ds*Red signal 10 - 14 days post inoculation (dpi)
1048 with *M. loti Ds*Red. Nodules undergoing different stages of infection (panels I to IV) were
1049 stained with X-Gluc to reveal the *GUS* expression pattern. Note the overlapping bacterial
1050 invasion zone and *PACE:*$NINmin_{pro}$*:GUS* expression in early infection stages (red and blue

1051 arrowheads in (**D**)) as well as the differences between *PACE:NINmin$_{pro}$:GUS* and the much
1052 broader *NIN$_{pro}$:GUS* expression at that stage (red and blue arrows in (**A**)). Red arrow and
1053 arrowheads: *M. loti Ds*Red. Blue arrow and arrowheads: GUS activity in root hairs bearing ITs
1054 and nodule primordia, respectively. The *NINmin$_{pro}$:GUS* fusion gave only rarely detectable
1055 signal, and if so in the vasculature (yellow arrowhead in (**E**)). Only pictures taken under white
1056 light illumination (WLI) are displayed for nodules in panel VI to reveal the pink colour of
1057 leghemoglobin, characteristic for mature and fully infected nodules. Note that
1058 *PACE:NINmin$_{pro}$:GUS* expression was absent at this stage, whereas the *NIN$_{pro}$:GUS* resulted
1059 in strong blue staining in the nodule regardless of the presence of *PACE* (compare panel IV
1060 in (**D**) and (**A - C**)). (**F**) Quantification of transgenic root systems exhibiting *GUS* expression in
1061 different cell types and tissues exemplarily displayed in (**A - E**). (**G**) *PACE* drove *GUS* reporter
1062 gene expression in the central tissue of primordia and nodules, but was not sufficient for
1063 expression in root hairs. Transgenic roots carrying promoter:GUS fusions same as in (**A, D**
1064 **and E**) were inoculated with *M. loti lacZ* and dual-stained with X-Gluc and Magenta-Gal.
1065 Purple: *M. loti lacZ*. Blue: GUS activity. Note the co-existence of blue and purple staining in
1066 root hairs on roots transformed by *NIN$_{pro}$:GUS*, but not that transformed by
1067 *PACE:NINmin$_{pro}$:GUS*. Bars, 250 µm.

1068 **Figure 3 – figure supplement 1 | *L. japonicus nin-15* mutant phenotype.** (**A**) A
1069 representative picture of *L. japonicus* wild-type (WT, left) and *nin-15* (right) plants 21 dpi with
1070 *M. loti Ds*Red. (**B**) Position of the *Lotus Retrotransposon 1* (*LORE1*) insertion within the *NIN*
1071 promoter in the *nin-15* mutant. (**C**) Representative pictures of *nin-15* root hairs and nodule
1072 sections 21 dpi with *M. loti Ds*Red. Forty-nine plants with a total number of 436 nodules were
1073 analysed: only four plants bore one or two IT(s) within root hairs and seven plants bore one
1074 or two infected nodule(s). Deformed or curled root hairs in the presence of *M. loti Ds*Red were
1075 abundant but infection threads were rarely found. Arrowheads: uninfected nodules.
1076 Unlabelled bars, 100 µm. (**D - E**) Phenotype of *nin-15* in the presence of a symbiosis-
1077 independent nitrogen source (15 mM KNO_3) for 28 days. (**D**) Pictures documenting the healthy
1078 status of *L. japonicus* WT and *nin-15* plants (compare (**D**) and (**A**)) and (**E**) quantitative
1079 assessment of parameters displayed in boxplots. Thirty plants per genotype were analysed.
1080 Each dot represents one plant. Lateral root density: number of lateral roots/primary root length
1081 (cm). The applied statistical method was pairwise *t*-test: $*p < 0.05$; $**p < 0.01$; n.s.: not
1082 significant. (**F**) Segregation analysis of *nin-15.* The applied statistical method was ANOVA with
1083 *post hoc* Tukey: $F_{3,120} = 84.1$, $p = 2x10^{-16}$. Different small letters indicate significant difference.
1084 (**G**) Representative pictures of *nin-15* plants with hairy roots transformed with the *NIN* gene
1085 driven by the *L. japonicus NIN* minimal promoter (*NINmin$_{pro}$*) or the 3 kb *NIN* promoter (*NIN$_{pro}$*)
1086 24 dpi with *M. loti Ds*Red. WLI: white light illumination.

1087 **Figure 3 – figure supplement 2 | The *CYC-box* and flanking sequences of *PACE* are**
1088 **required for the complete restoration of the bacterial infection process in the *L.***
1089 ***japonicus nin-2* mutant.** *nin-2* roots were transformed with T-DNAs carrying a $Ubq10_{pro}$*:NLS-*
1090 *GFP* transformation marker in tandem with the *LjNIN* gene driven by either of the following
1091 promoter versions: the cytokinin element-containing region of 1 kb (CE_{1kb}) fused to the 3 kb or
1092 9 kb *LjNIN* promoter (CE_{1kb}*:*NIN_{pro} or CE_{1kb}*:*NIN_{9kbpro}, respectively); CE_{1kb}*:*NIN_{pro} or
1093 CE_{1kb}*:*NIN_{9kbpro} with *PACE* mutated (CE_{1kb}*:*NIN_{pro}*::mPACE* or CE_{1kb}*:*NIN_{9kbpro}*::mPACE*,
1094 respectively); CE_{1kb}*:*NIN_{pro} carrying a mutated Cyclops binding site (*CYC-box*)
1095 (CE_{1kb}*:*NIN_{pro}*::mbox*); CE_{1kb}*:*NIN_{pro} carrying mutated sequences flanking the *CYC-box* in *PACE*
1096 (CE_{1kb}*:*NIN_{pro}*::mflanking*); CE_{1kb} fused to the *LjNIN* minimal promoter (CE_{1kb}*:*$NINmin_{pro}$); CE_{1kb}
1097 fused to *PACE* and to $NINmin_{pro}$ (CE_{1kb}*:PACE:*$NINmin_{pro}$); NIN_{pro}, *PACE:*$NINmin_{pro}$ or
1098 $NINmin_{pro}$. (**A**) Representative overview pictures of transgenic root systems. Roots were
1099 analysed 21 dpi with *M. loti Ds*Red. White asterisks and arrowheads: infected and non-
1100 infected nodules, respectively. Bars, 2 mm. (**B - C**) Boxplots displaying the number of root hair
1101 ITs or infected nodules and the percentage of root hair ITs among total infection events (sum
1102 of bacterial entrapments and ITs). Each dot represents one transgenic *nin-2* root system or
1103 root piece. *L. japonicus* WT roots transformed with NIN_{pro}*:NIN* or CE_{1kb}*:*NIN_{pro}*:NIN* were
1104 included as controls. Note the loss of restoration of nodules and IT formation associated with
1105 the mutation of *PACE* or only the *CYC-box* in *PACE*; and the reduction of same when
1106 sequences flanking the *CYC-box* in *PACE* were mutated. n: number of transgenic root
1107 systems or root pieces analysed. Numbers above the boxplots: the value of individual data
1108 points outside of the plotting area. n.d.: not determined. WLI: white light illumination.

1109 **Figure 3 – figure supplement 3 | The *CYC-box* and flanking sequences of *PACE* are**
1110 **required for the complete restoration of the bacterial infection process in the *L.***
1111 ***japonicus nin-2* mutant.** Roots were from a subset of plants from the same experiment
1112 depicted in **Figure 3 – figure supplement 2** but analysed 35 dpi with *M. loti Ds*Red. (**A - B**)
1113 Boxplots displaying the number of root hair ITs or infected nodules and the percentage of root
1114 hair ITs among total infection events (sum of bacterial entrapments and ITs). Each dot
1115 represents one transgenic *nin-2* root system or root piece. *L. japonicus* WT roots transformed
1116 with NIN_{pro}*:NIN* or CE_{1kb}*:*NIN_{pro}*:NIN* were included as controls. Note that the results follow the
1117 same trend as those obtained 21 dpi with *M. loti Ds*Red (**Figure 3 – figure supplement 2**). n:
1118 number of transgenic root systems or root pieces analysed. Numbers above the boxplots: the
1119 value of individual data points outside of the plotting area.

1120 **Figure 3 – figure supplement 4 | The *CYC-box* and flanking sequences of *PACE* are**
1121 **required for the complete restoration of the bacterial infection process but are**
1122 **dispensable for the nodule organogenesis process in the *L. japonicus nin-2* mutant.**

1123 *nin-2* roots were transformed with T-DNAs carrying a *Ubq10*$_{pro}$*:NLS-GFP* transformation
1124 marker in tandem with the *LjNIN* gene driven by either of the following promoter versions: the
1125 cytokinin element-containing region of 5 kb (*CE*$_{5kb}$) fused to the 3 kb *LjNIN* promoter
1126 (*CE*$_{5kb}$*:NIN*$_{pro}$); *CE*$_{5kb}$*:NIN*$_{pro}$ with *PACE* mutated (*CE*$_{5kb}$*:NIN*$_{pro}$*::mPACE*); *CE*$_{5kb}$*:NIN*$_{pro}$ carrying
1127 a mutated Cyclops binding site (*CYC-box*) (*CE*$_{5kb}$*:NIN*$_{pro}$*::mbox*); *CE*$_{5kb}$*:NIN*$_{pro}$ carrying mutated
1128 sequences flanking the *CYC-box* in *PACE* (*CE*$_{5kb}$*:NIN*$_{pro}$*::mflanking*); *CE*$_{5kb}$ fused to the *LjNIN*
1129 minimal promoter (*CE*$_{5kb}$*:NINmin*$_{pro}$); *CE*$_{5kb}$ fused to *PACE* and to *NINmin*$_{pro}$
1130 (*CE*$_{5kb}$*:PACE:NINmin*$_{pro}$) or *NIN*$_{pro}$. (**A**) Representative overview pictures of transgenic root
1131 systems. Roots were analysed 21 dpi with *M. loti Ds*Red. White asterisks and arrowheads:
1132 infected and non-infected nodules, respectively. Bars, 2 mm. (**B**) Boxplots displaying the
1133 number of infected nodules, the percentage of infected nodules among total organogenesis
1134 events (sum of infected and non-infected nodules) and the number of organogenesis events.
1135 (**C**) Boxplots displaying the number of root hair ITs and the percentage of root hair ITs among
1136 total infection events (sum of bacterial entrapments and ITs). Each dot represents one
1137 transgenic *nin-2* root system or root piece. *L. japonicus* WT roots transformed with *NIN*$_{pro}$*:NIN*
1138 or *CE*$_{5kb}$*:NIN*$_{pro}$*:NIN* were included as controls. Note that the mutation of *PACE* or only the
1139 *CYC-box* in *PACE* led to an almost complete loss of IT formation and infected nodules per
1140 root system while nodule organogenesis was not significantly reduced; and that mutation of
1141 sequences flanking the *CYC-box* in *PACE* led to a reduction of the number of infected nodules
1142 per root systems. n: number of transgenic root systems or root pieces analysed. Numbers
1143 above the boxplots: the value of individual data points outside of the plotting area. n.a.: not
1144 applicable. WLI: white light illumination.

1145 **Figure 3 – figure supplement 5 | The *CYC-box* and flanking sequences of *PACE* are**
1146 **required for the complete restoration of the bacterial infection process but are**
1147 **dispensable for the nodule organogenesis process in the *L. japonicus nin-2* mutant.**
1148 Pictures of nodule sections or roots from *L. japonicus nin-2* roots 21 dpi with *M. loti Ds*Red
1149 from the same experiments depicted in **Figure 3 – figure supplement 2** (**A**) and **Figure 3 –**
1150 **figure supplement 4** (**B**). Nodule sections from *L. japonicus* WT roots transformed with
1151 *NIN*$_{pro}$*:NIN* and *CE*$_{5kb}$*NIN*$_{pro}$*:NIN* were included for comparison. Note that when the cytokinin
1152 element-containing region of 1 kb was fused to *NIN*$_{pro}$ nodule organogenesis was abolished
1153 by mutation of *PACE* or only the *CYC-box* in *PACE* and that these mutations did not abolish
1154 organogenesis when the cytokinin element-containing region of 5 kb was fused to *NIN*$_{pro}$. Bars,
1155 100 µm.

1156 **Figure 3 – figure supplement 6 | The *CYC-box* and flanking sequences of *PACE* are**
1157 **required for the complete restoration of the bacterial infection process in the *L.***
1158 ***japonicus nin-2* mutant.** Pictures of nodule sections or roots from *L. japonicus nin-2* roots 35

1159 dpi with *M. loti Ds*Red from the same experiments depicted in **Figure 3 – figure supplement**
1160 **3**. Upper left corner: a nodule section from a *L. japonicus* WT root transformed with NIN_{pro}*:NIN*
1161 was included for comparison. Bars, 100 µm.

1162 **Figure 3 – figure supplement 7 | *PACEs* from FaFaCuRo species are functionally**
1163 **equivalent in restoring bacterial infection in the *L. japonicus nin-15* mutant.** *L. japonicus*
1164 *nin-15* roots were transformed with T-DNAs carrying a $Ubq10_{pro}$*:NLS-GFP* transformation
1165 marker in tandem with the *LjNIN* gene driven by either of the following promoters: (**A**) the 3
1166 kb *LjNIN* promoter (NIN_{pro}), the *LjNIN* minimal promoter ($NINmin_{pro}$), the 3 kb *LjNIN* promoter
1167 with *PACE* deleted (NIN_{pro}*::ΔPACE*) or mutated (NIN_{pro}*::mPACE*); (**B**) the 3 kb *LjNIN* promoter
1168 with *LjPACE* replaced with either of the *PACE* sequence variants from nodulating or non-
1169 nodulating FaFaCuRo species and analysed 21 dpi with *M. loti Ds*Red. (**A - B**) Representative
1170 overview pictures of *nin-15* transgenic roots systems. Sections of representative nodules are
1171 displayed in **Figure 3**. Note the drastic reduction of restoration of infection in nodules and root
1172 hairs associated with the mutation or deletion of *PACE* as well as the replacement of *PACE*
1173 with *JrPACE-like* in the context of the *LjNIN* promoter. White asterisks and arrowheads:
1174 infected and non-infected nodules, respectively. (**C - E**) Boxplots displaying **c,** the percentage
1175 of root hair ITs among total infection events (sum of bacterial entrapments and ITs) and (**D -**
1176 **E**) the number of infected nodules from two independent experiments. Each dot in (**D - E**)
1177 represents one *nin-15* transgenic root system. (**C**) displays merged data from all experiments
1178 as the percentage represents a normalised value calculated for each root piece (see **Figure**
1179 **3 – Source Data 1**). n: number of transgenic root systems or root pieces analysed. For
1180 species abbreviations see **Figure 1 – figure supplement 5A**. The applied statistical method
1181 was ANOVA with *post hoc* Tukey: (**C**) $F_{9,313} = 106.7$, $p < 2x10^{-16}$; (**D**) $F_{6,346} = 82.89$, $p < 2x10^{-}$
1182 ^{16}; (**E**) $F_{4,135} = 20.18$, $p = 4.76x10^{-13}$. Different small letters indicate significant differences. Bars,
1183 2 mm. WLI: white light illumination.

1184 **Figure 3 – figure supplement 8 | Spatio-temporal *GUS* expression driven by *PACE***
1185 **variants in *L. japonicus* roots during the bacterial infection process.** *L. japonicus* WT
1186 roots were transformed with T-DNAs carrying a $Ubq10_{pro}$*:NLS-GFP* transformation marker
1187 together with a *GUS* reporter gene driven by either of the *PACE* variants from nodulating
1188 FaFaCuRo species fused to the *LjNIN* minimal promoter ($NINmin_{pro}$). For species
1189 abbreviations and experimental details see **Figure 1 – figure supplement 5A and Figure 2**
1190 **– figure supplement 1**, respectively. Note the overlapping bacterial invasion zone and
1191 *PACE:*$NINmin_{pro}$*:GUS* expression in early infection stages (red and blue arrowheads in (**A -**
1192 **C**)). Red arrowheads: *M. loti Ds*Red. Blue arrowheads: GUS activity in nodule primordia. Only
1193 pictures taken under white light illumination (WLI) are displayed for nodules in panel VI to
1194 reveal the pink colour of leghemoglobin, characteristic for mature and fully infected nodules.

1195 Note that like *LjPACE*, the *PACE* variants-driven *GUS* expressions were absent at this stage
1196 (panel IV in (**A - C**) and panel IV in **Figure 2 – figure supplement 1D**). (**D**) Quantification of
1197 transgenic root systems exhibiting *GUS* expression in different cell types and tissues
1198 exemplarily displayed in (**A - C**). n.d.: not determined. Bars, 250 µm.

1199 **Figure 4 – figure supplement 1 | *PACE* alone or in the context of the *S. lycopersicum***
1200 ***NIN* promoter (a species outside of the FaFaCuRo clade) enables IT formation in the**
1201 **cortex.** (**A - D**) Representative pictures of sections of nodules formed on *L. japonicus nin-15*
1202 roots transformed with T-DNAs carrying a $Ubq10_{pro}$*:NLS-GFP* transformation marker together
1203 with the *LjNIN* gene driven by either of the following promoters: (**A - B**) the *L. japonicus NIN*
1204 minimal promoter ($NINmin_{pro}$) or *PACE* fused to $NINmin_{pro}$ ($PACE{:}NINmin_{pro}$); (**C - D**) the 3 kb
1205 *S. lycopersicum NIN* promoter ($SlNIN_{pro}$), the 3 kb *SlNIN* promoter with mutated *PACE*
1206 ($SlNIN_{pro}$*::mPACE*) or with *L. japonicus PACE* inserted ($SlNIN_{pro}$*::PACE*), 21 dpi with *M. loti*
1207 *Ds*Red (from the same experiments depicted in **Figure 4**). Black rectangles in (**A**) demarcate
1208 the enlarged area displayed in **Figure 4A and 4B** to focus on the initial infection structures.
1209 Note the absence of cells filled with symbiosomes in nodules transformed with the *LjNIN* gene
1210 driven by $PACE{:}NINmin_{pro}$ or $NINmin_{pro}$. By contrast, infected cells were often filled with
1211 symbiosomes in the $SlNIN_{pro}$*::PACE:NIN*-transformed nodules, like those resulted by
1212 NIN_{pro}*:NIN* (see (**C**) and compare the two sections in (**D**)). (**E - F**) Boxplots displaying the
1213 percentage of root hair ITs among total infection events (sum of bacterial entrapments and
1214 ITs) or the percentage of infected nodules among total number of nodules (**E**) 21 dpi and (**F**)
1215 35 dpi with *M. loti Ds*Red, respectively. Each dot represents one *nin-15* transgenic root system
1216 or root piece. (**E**) displays results from an independent repetition from the experiment depicted
1217 in **Figure 4**. n: number of transgenic root systems or root pieces analysed. Numbers above
1218 the boxplots: the value of individual data points outside of the plotting area. The applied
1219 statistical method was Fisher's exact test: $*p < 0.05$; $***p < 0.001$; n.s.: not significant. Bars,
1220 (**A and C**) 100 µm; (**B and D**) 50 µm.

1221 **Source Data:**

1222 **Figure 1 – Source Data 1 |** Summary of the bioinformatic analysis resulting in the discovery
1223 of *PACE* using 37 species.

1224 **Figure 1 – Source Data 2 |** Results of the FIMO analysis of *PACE* in 163 species.

1225 **Figure 1 – Source Data 3 |** Status of *PACE* and *NIN* in non-nodulating FaFaCuRo species.

1226 **Figure 3 – Source Data 1 |** Complementation results of the *L. japonicus* Gifu *nin-2* and *nin-*
1227 *15* mutant lines.

1228

1229 **Additional Files:**

1230 **Supplementary file 1 |** List of plant genomes used for the search of conserved motifs within
1231 *ERN1* and *RAM1* promoters.

1232 **Supplementary file 2 |** List of seed bags, bacterial strains and incubation times.

1233 **Supplementary file 3 |** List of plasmids used.

1234 **Supplementary file 4 |** Sequences and IDs of oligonucleotides (DNA) used.

1235 **Supplementary file 5 |** Microscope/scanner settings and image analysis.

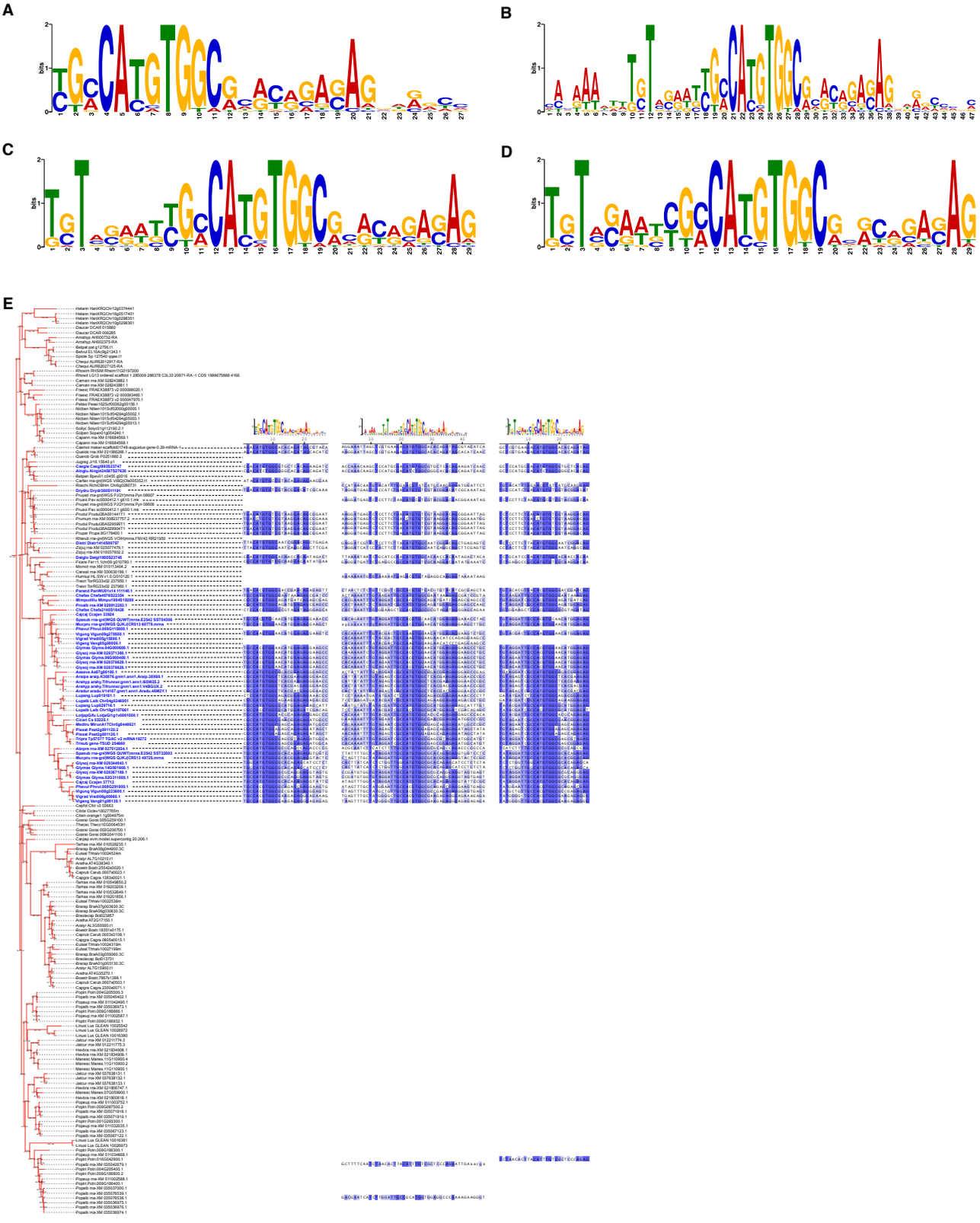


Figure 1 – figure supplement 1 | Discovery of *PACE* by MEME analyses. (**A - D**) Consensus sequence of the Position Weight Matrix identified by MEME analyses using the regions upstream of the translational start site (ATG) thus representing the promoters and 5'UTRs of *NIN* and *NLP* genes from 37 angiosperm species (see Methods): (**A**) in a discriminative search for a motif that is present in 3 kb upstream regions of the *NIN* genes from nodulating FaFaCuRo species, but absent in 3 kb upstream regions of the *NIN* genes from species outside of the FaFaCuRo clade and absent in the 3 kb upstream regions of *NLP* genes; (**B**) in an independent, non-discriminative search in the 3 kb upstream regions of the *NIN* genes from nodulating species revealing that the most conserved nucleotides span a larger region than (**A**); (**C**) the resulted most conserved 29 nucleotides derived from upstream regions of *NIN* genes of nodulating FaFaCuRo species; (**D**) the resulted most conserved 29 nucleotides (*PACE*) derived from the upstream region of one representative *NIN* gene per species of nodulating FaFaCuRo species. (**E**) Motif analyses by FIMO using the upstream regions of the *NIN* and *NLP* genes from an expanded list of 163 species (see Methods). Left: pruned tree from the whole *NLP* tree (demarcated in the black rectangle in **Figure 1 – figure supplement 2**) corresponding to the *NIN* orthologs. Right: three versions of consensus sequence (left to right, (**A, B and D**), respectively) were used to retrieve motifs from the 5 kb upstream region of *NIN* genes *via* FIMO search, the output of which is displayed underneath the consensus sequences. Sequence names of nodulating species are coloured in blue. Blank lines represent the absence of significant motifs.

;2:0.067,0.025,0.933,0.975;3:0.067,0.023,0.933,0.977;4:0.067,0.020,0.933,0.980;5:0.067,0.020,0.933,0.980;6:0.067,0.020,0.933,0.980;7:0.067,0.020,0.933,0.980;8:0.067,0.020,0.933,0.980;9:0.067,0.020,0.933,0.980"

Figure 1 – figure supplement 2 I *PACE* is present exclusively in the FaFaCuRo clade. Motif analyses by FIMO using the using the regions upstream of the translational start site (ATG) thus representing the promoters and 5'UTRs of the *NIN* and *NLP* genes from an expanded list of 163 species (see Methods). Left: Maximum likelihood tree of *NLP* family (model: GTR+F+R10) of 163 species (**Figure 1 – Source Data 2**). Tree was rooted on the Charophytes algae *Klebsormidium nitens*. Black box demarcates the area that is displayed in **Figure 1 – figure supplement 1E**. Blank lines represent the absence of significant motifs. Sequence names of nodulating species are coloured in blue. The different NLP clades are indicated by branch colours and named according to the *Arabidopsis thaliana* nomenclature. Right: three versions of consensus sequence (left to right, **Figure 1 – figure supplement 1A, 1B and 1D**, respectively) were used to retrieve motifs from the 5 kb upstream region of *NIN* genes via FIMO search, the output of which is displayed underneath the consensuses sequences.

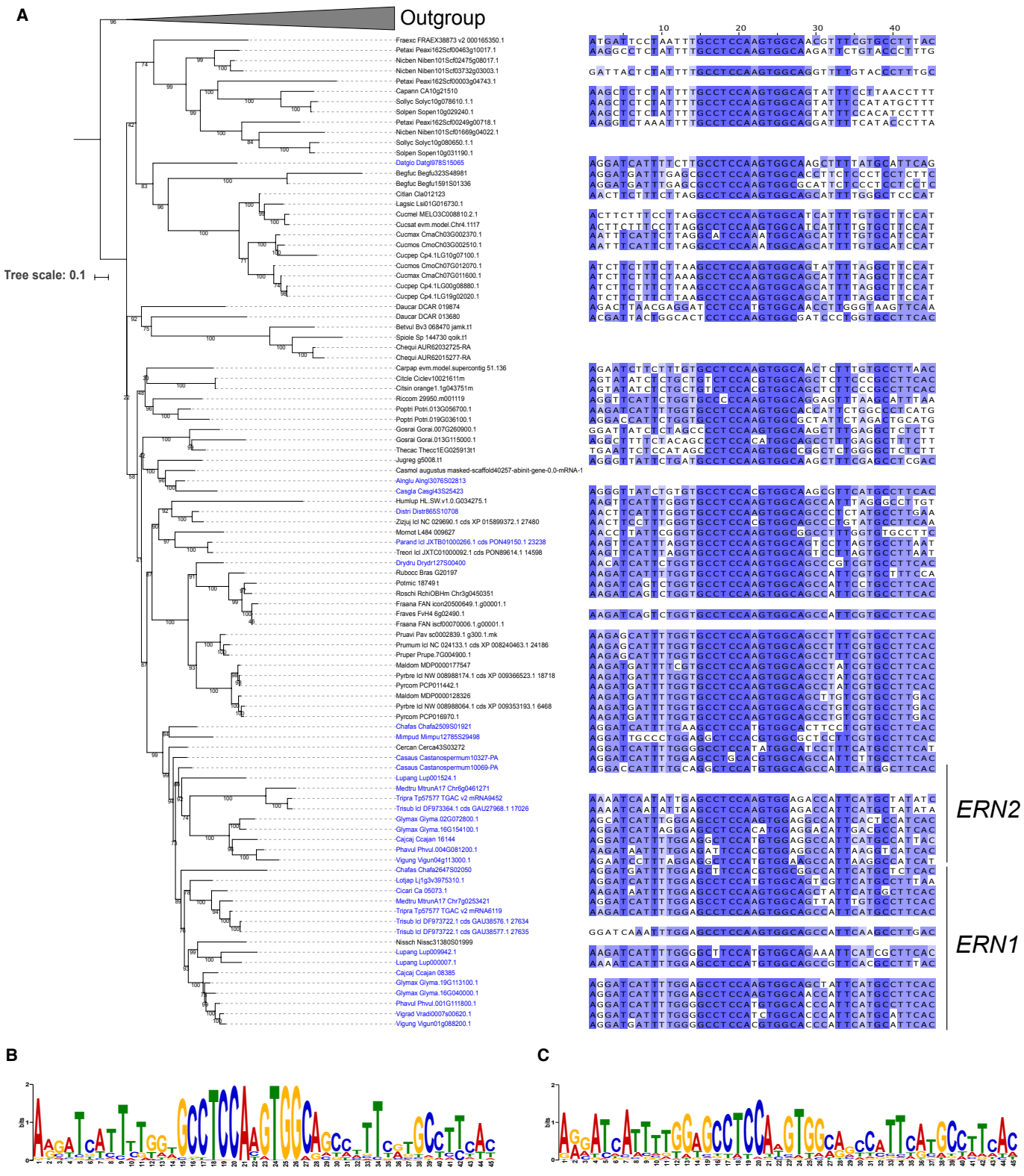


Figure 1 – figure supplement 3 | The presence of a motif encompassing the previously identified Cyclops binding site within the regions upstream of the translational start site (ATG) thus representing the promoters and 5'UTRs of *ERN1* genes extends beyond the FaFaCuRo clade. (**A**) Maximum likelihood of *ERN1/ERN2* (model: JTT+F+R5) from 87 species (**Supplementary file 1**) based on protein sequences. Tree was rooted using sequences from the *ERN1/ERN2* paralog *ERN3* (outgroup clade collapse and indicated by grey triangle marked "Outgroup"). Species belonging to the monophyletic FaFaCuRo clade (Stevens, 2017) are indicated by thicker black branches. *ERN1* and *ERN2* subclades are indicated. Sequence names of nodulating species are coloured in blue. These conserved motifs identified by MEME analyses in the 3 kb upstream regions were aligned and displayed next to species names. While the presence of a conserved Cyclops binding site containing motif in species outside of the FaFaCuRo clade is consistent with the role of *RAM1* in arbuscular mycorrhizal symbiosis, a function of *ERN1* in arbuscular mycorrhizal symbiosis has not been described and suggests a yet unidentified function of *ERN1* outside of the FaFaCuRo clade. (**B**) Consensus sequence of the Position Weight Matrix identified by MEME analyses of all the species. (**C**) Consensus sequence of the Position Weight Matrix identified by MEME analyses when analysing only the nodulating species.

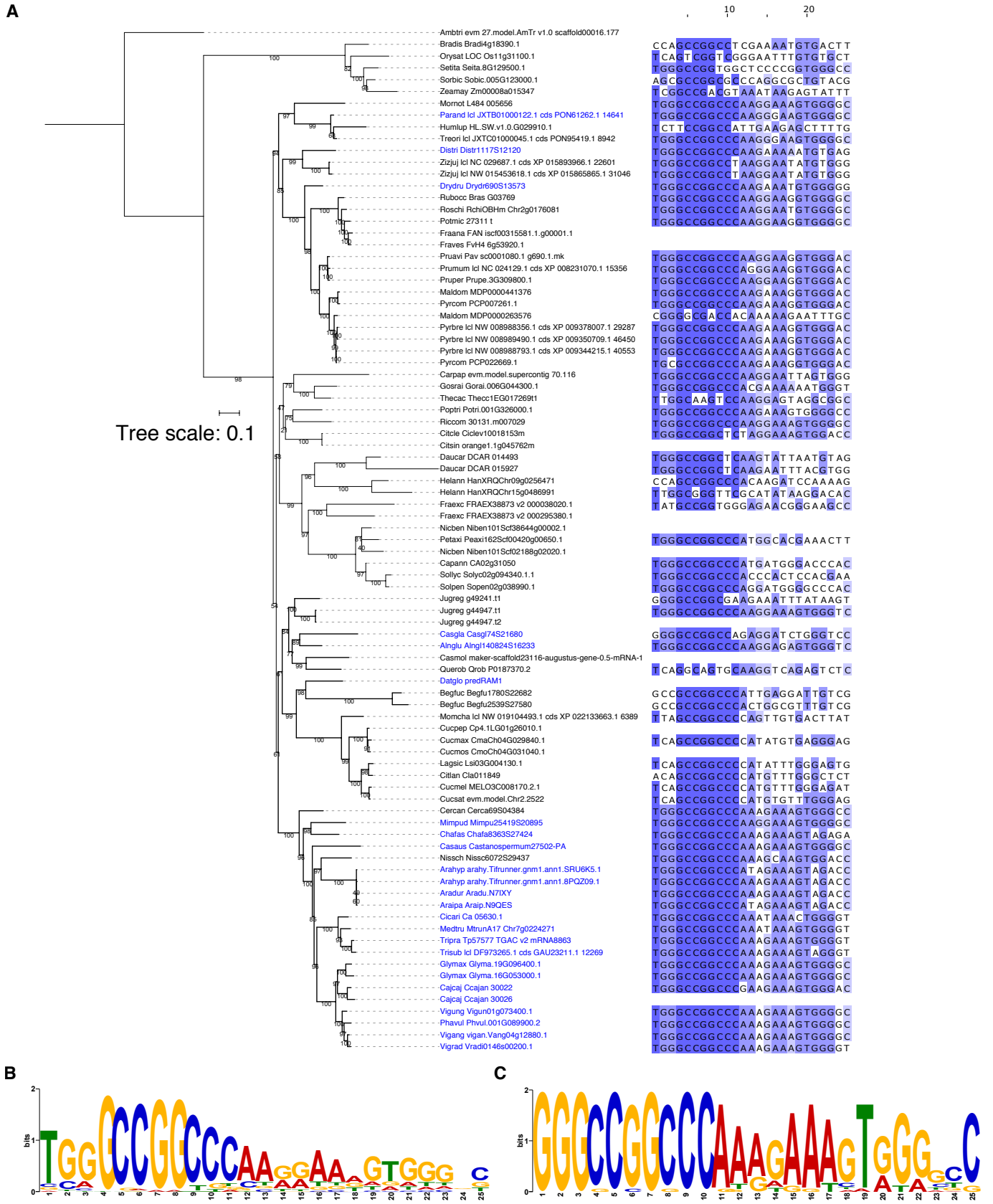
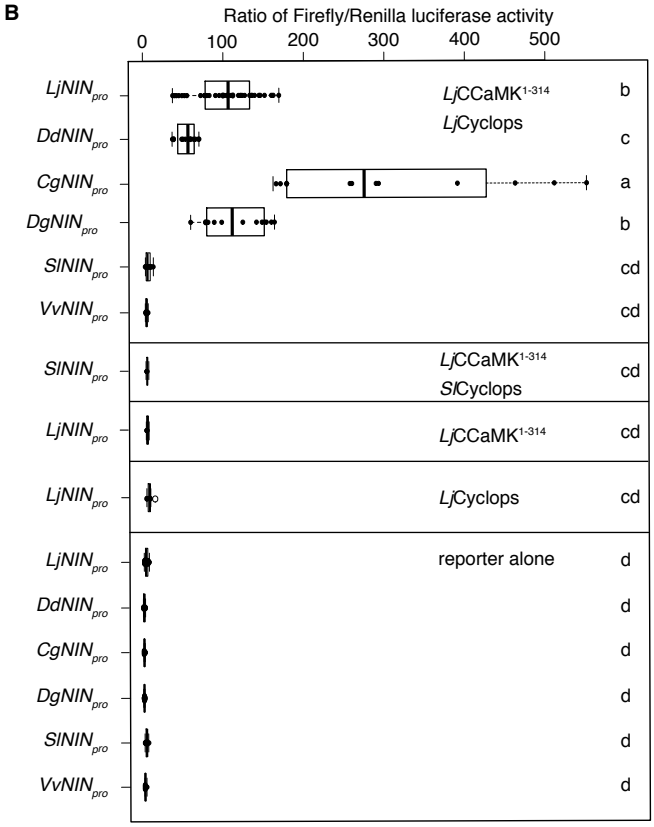


Figure 1 – figure supplement 4 I The presence of a motif encompassing the previously identified Cyclops binding site within the regions upstream of the translational start site (ATG) thus representing the promoters and 5'UTRs of *RAM1* genes extends beyond the FaFaCuRo clade. (**A**) Maximum likelihood of *RAM1* (model: JTT+F+R4) from 87 species (**Supplementary file 1**) engaging in arbuscular mycorrhizal symbiosis. Tree was rooted on the early-diverging angiosperm *Amborella trichopoda*. Species belonging to the monophyletic FaFaCuRo clade (Stevens, 2017) are indicated by thicker black branches. Sequence names of nodulating species are coloured in blue. These conserved motifs identified by MEME analyses in the 3 kb upstream regions are aligned and displayed next to species names. Blank lines represent the absence of identified motif due to genome contiguity issues. (**B**) Consensus sequence of the Position Weight Matrix identified by MEME analyses when analysing all the mycorrhizal species. (**C**) Consensus sequence of the Position Weight Matrix identified by MEME analyses when analysing only the nodulating species.

A

Abbre-viation	Species	Plant Order	RNS	FaFaCuRo member
Lj	*Lotus japonicus*	Fabales	Yes	Yes
Cg	*Casuarina glauca*	Fagales	Yes	Yes
Jr	*Juglans regia*	Fagales	No	Yes
Dd	*Dryas drummondii*	Rosales	Yes	Yes
Zj	*Ziziphus jujuba*	Rosales	No	Yes
Pp	*Prunus persica*	Rosales	No	Yes
Dg	*Datisca glomerata*	Cucurbitales	Yes	Yes
Sl	*Solanum lycopersicum* (tomato)	Solanales	No	No
Vv	*Vitis vinifera*	Vitales	No	No



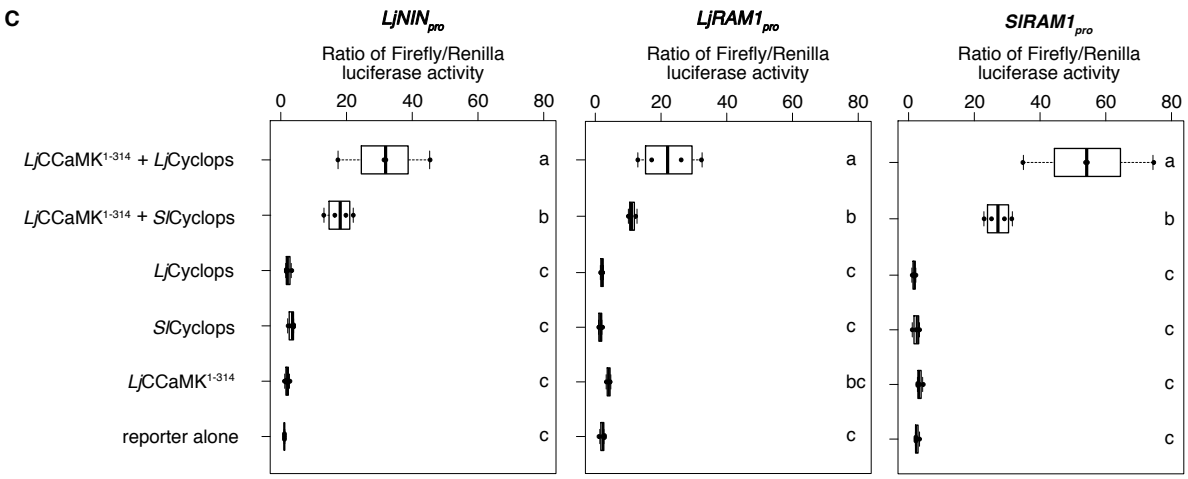


Figure 1 – figure supplement 5 | Transcriptional activation of *NIN promoter:Firefly luciferase* reporter gene by CCaMK$^{1-314}$/Cyclops is restricted to *NIN* promoters from species of the FaFaCuRo clade. *Nicotiana benthamiana* leaf cells were transformed with T-DNAs carrying a *Firefly luciferase* reporter gene driven by either of the indicated promoters in tandem with the *AtACT2$_{pro}$:Renilla luciferase* reporter fusion that provides a quantitative internal standard. (**A**) List of species within the FaFaCuRo clade (light red shade) and outside (light grey shade) and abbreviations. (**B**) Reporter gene activation by *L. japonicus* CCaMK$^{1-314}$/Cyclops via *NIN* promoters (*NIN$_{pro}$*) originating from listed species. (**C**) Comparison of the transactivation potential of Cyclops versions from *L. japonicus* and *S. lycopersicum*. Note that the expression of the *Firefly luciferase* reporter gene driven by *LjNIN$_{pro}$*, the *RAM1* promoters from *L. japonicus* and *S. lycopersicum* (*LjRAM1$_{pro}$* and *SlRAM1$_{pro}$*, respectively) was induced in the presence of CCaMK$^{1-314}$/Cyclops regardless of the origin of Cyclops. In contrast, the transactivation failed with the *SlNIN* promoter (panel (**A**)). The applied statistical method was ANOVA with *post hoc* Tukey: (**B**), $F_{14,214} = 71.07$, $p < 2\text{x}10^{-16}$; (**C**), plots from left to right: $F_{5,18} = 20.58$, $p = 7.14\text{x}10^{-7}$; $F_{5,18} = 25.38$, $p = 1.45\text{x}10^{-7}$ and $F_{5,18} = 40.49$, $p = 3.55\text{x}10^{-9}$, respectively. Different small letters indicate significant differences.

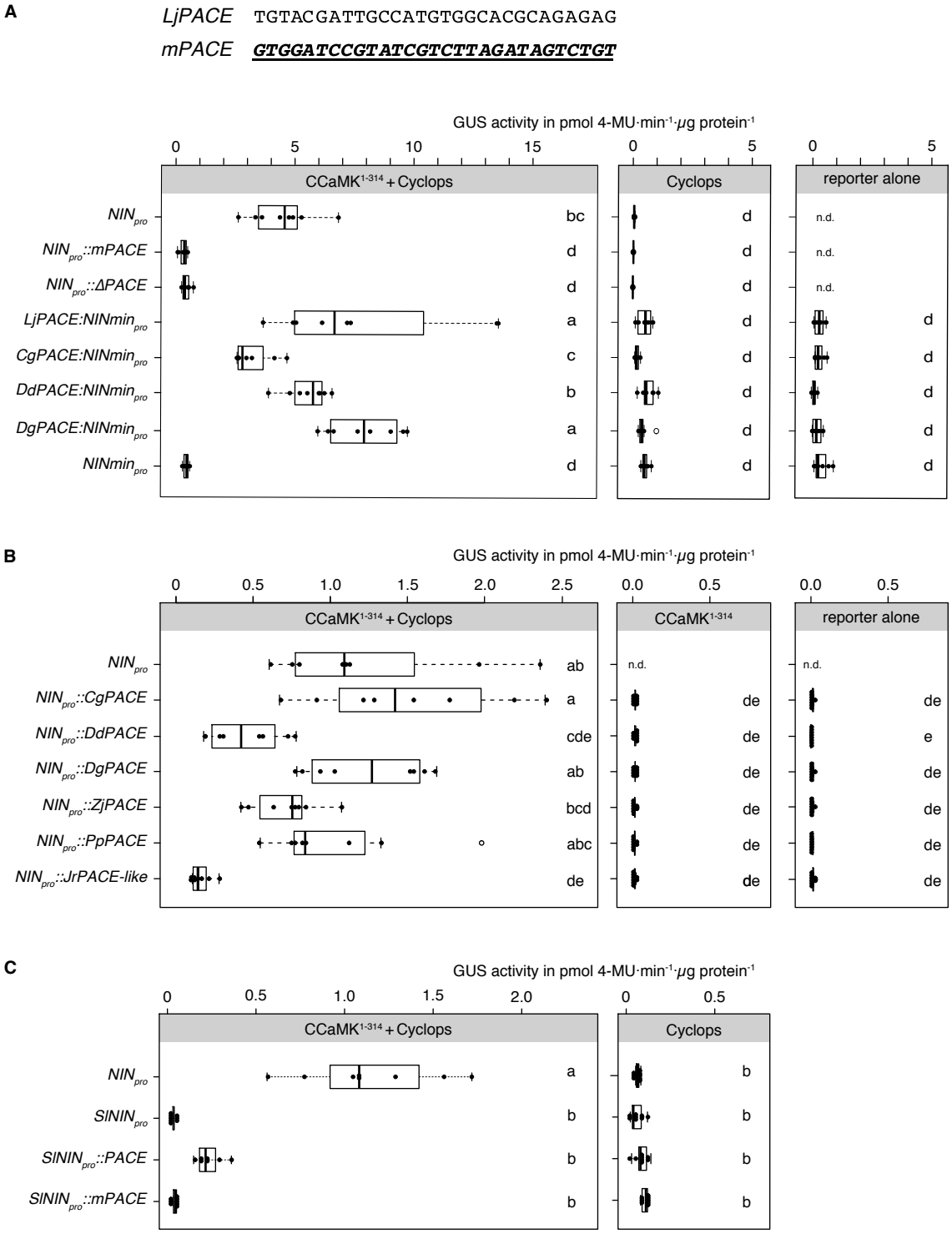
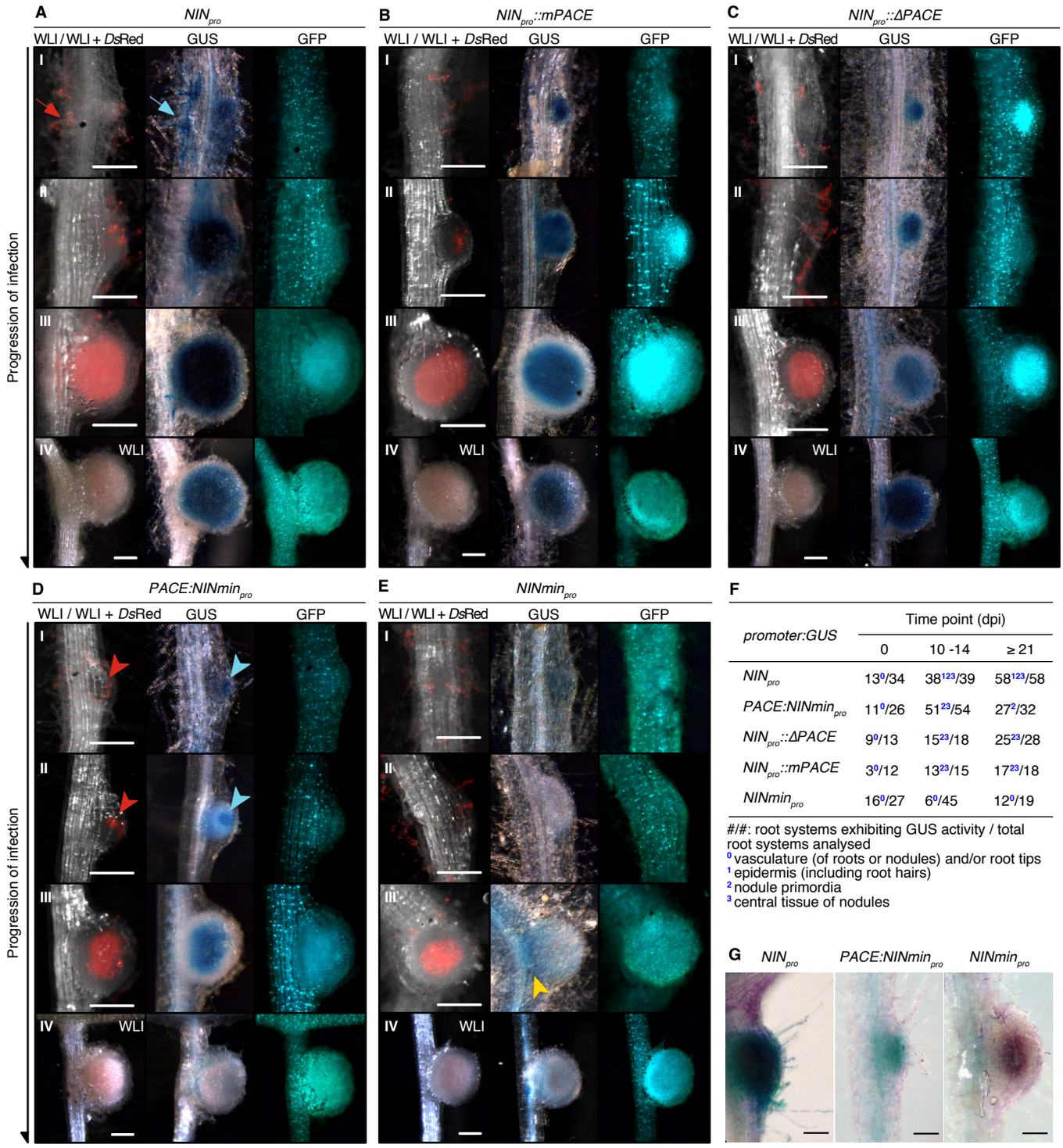


Figure 1 – figure supplement 6 | *PACE* sequence variants from species across the FaFaCuRo clade were able to functionally replace *L. japonicus PACE* in a *LjNIN$_{pro}$:GUS* reporter fusion. *N. benthamiana* leaf cells were transformed with T-DNAs carrying a *GUS* reporter gene driven by either of the indicated promoters: (**A**) the *L. japonicus NIN* promoter (*NIN$_{pro}$*), the *LjNIN* promoter with *PACE* mutated or deleted (*NIN$_{pro}$::mPACE* and *NIN$_{pro}$::ΔPACE*, respectively), or *PACE* sequence variants from the nodulating FaFaCuRo species fused to the *LjNIN* minimal promoter (*NINmin$_{pro}$*); (**B**) chimeric promoters where *LjPACE* in the *LjNIN* promoter was replaced with either one of the *PACE* variants from species tested in (**A**) or from non-nodulating FaFaCuRo species including the *Juglans regia PACE*-like motif (*JrPACE-like*); (**C**) the *S. lycopersicum NIN* promoter (*SlNIN$_{pro}$*), the *SlNIN* promoter with *LjPACE* (*SlNIN$_{pro}$::PACE*) or *mPACE* (*SlNIN$_{pro}$::mPACE*) inserted. For species abbreviations see **Figure 1 – figure supplement 5A**. Note in (**A**) that the deletion or mutation of *PACE* in *LjNIN* promoter resulted in a drastic reduction in reporter gene expression and in (**C**) insertion of *LjPACE* but not *mPACE* into the *S. lycopersicum* promoter confers transactivation by CCaMK$^{1-314}$/Cyclops. The applied statistical method was ANOVA with *post hoc* Tukey: (**A**) $F_{20,144} = 51.38$, $p < 2\text{x}10^{-16}$; (**B**), $F_{18,166} = 149.1$, $p < 2\text{x}10^{-16}$; (**C**) $F_{7,62} = 30.5$, $p = 7.02\text{x}10^{-7}$. Different small letters indicate significant difference. n.d., not determined.



F

promoter:GUS	Time point (dpi)		
	0	10 -14	≥ 21
NIN_{pro}	13^{0}/34	38^{123}/39	58^{123}/58
$PACE:NINmin_{pro}$	11^{0}/26	51^{23}/54	27^{2}/32
$NIN_{pro}::\Delta PACE$	9^{0}/13	15^{23}/18	25^{23}/28
$NIN_{pro}::mPACE$	3^{0}/12	13^{23}/15	17^{23}/18
$NINmin_{pro}$	16^{0}/27	6^{0}/45	12^{0}/19

#/#: root systems exhibiting GUS activity / total root systems analysed
[0] vasculature (of roots or nodules) and/or root tips
[1] epidermis (including root hairs)
[2] nodule primordia
[3] central tissue of nodules

Figure 2 – figure supplement 1 I Spatio-temporal *GUS* expression driven by *PACE* and the *NIN* promoter in *L. japonicus* roots during the bacterial infection process. *L. japonicus* wild-type hairy roots were transformed with T-DNAs carrying a $Ubq10_{pro}$*:NLS-GFP* transformation marker together with a *GUS* reporter gene driven by either of the indicated promoters: (**A**) the 3 kb *LjNIN* promoter (NIN_{pro}); the *LjNIN* promoter with *PACE* (**B**) mutated ($LjNIN_{pro}$*::mPACE*) or (**C**) deleted (NIN_{pro}*::ΔPACE*); (**D**) *PACE* fused to the *LjNIN* minimal promoter ($PACE:NINmin_{pro}$) or (**E**) the *LjNIN* minimal promoter ($NINmin_{pro}$). The progression of bacterial infection was determined by the *Ds*Red signal 10 - 14 days post inoculation (dpi) with *M. loti Ds*Red. Nodules undergoing different stages of infection (panels I to IV) were stained with X-Gluc to reveal the *GUS* expression pattern. Note the overlapping bacterial invasion zone and $PACE:NINmin_{pro}$*:GUS* expression in early infection stages (red and blue arrowheads in (**D**)) as well as the differences between $PACE:NINmin_{pro}$*:GUS* and the much broader NIN_{pro}*:GUS* expression at that stage (red and blue arrows in (**A**)). Red arrow and arrowheads: *M. loti Ds*Red. Blue arrow and arrowheads: GUS activity in root hairs bearing ITs and nodule primordia, respectively. The $NINmin_{pro}$*:GUS* fusion gave only rarely detectable signal, and if so in the vasculature (yellow arrowhead in (**E**)). Only pictures taken under white light illumination (WLI) are displayed for nodules in panel VI to reveal the pink colour of leghemoglobin, characteristic for mature and fully infected nodules. Note that $PACE:NINmin_{pro}$*:GUS* expression was absent at this stage, whereas the NIN_{pro}*:GUS* resulted in strong blue staining in the nodule regardless of the presence of *PACE* (compare panel IV in (**D**) and (**A - C**)). (**F**) Quantification of transgenic root systems exhibiting *GUS* expression in different cell types and tissues exemplarily displayed in (**A - E**). (**G**) *PACE* drove *GUS* reporter gene expression in the central tissue of primordia and nodules, but was not sufficient for expression in root hairs. Transgenic roots carrying promoter:GUS fusions same as in (**A, D and E**) were inoculated with *M. loti lacZ* and dual-stained with X-Gluc and Magenta-Gal. Purple: *M. loti lacZ*. Blue: GUS activity. Note the co-existence of blue and purple staining in root hairs on roots transformed by NIN_{pro}*:GUS*, but not that transformed by $PACE:NINmin_{pro}$*:GUS*. Bars, 250 µm.

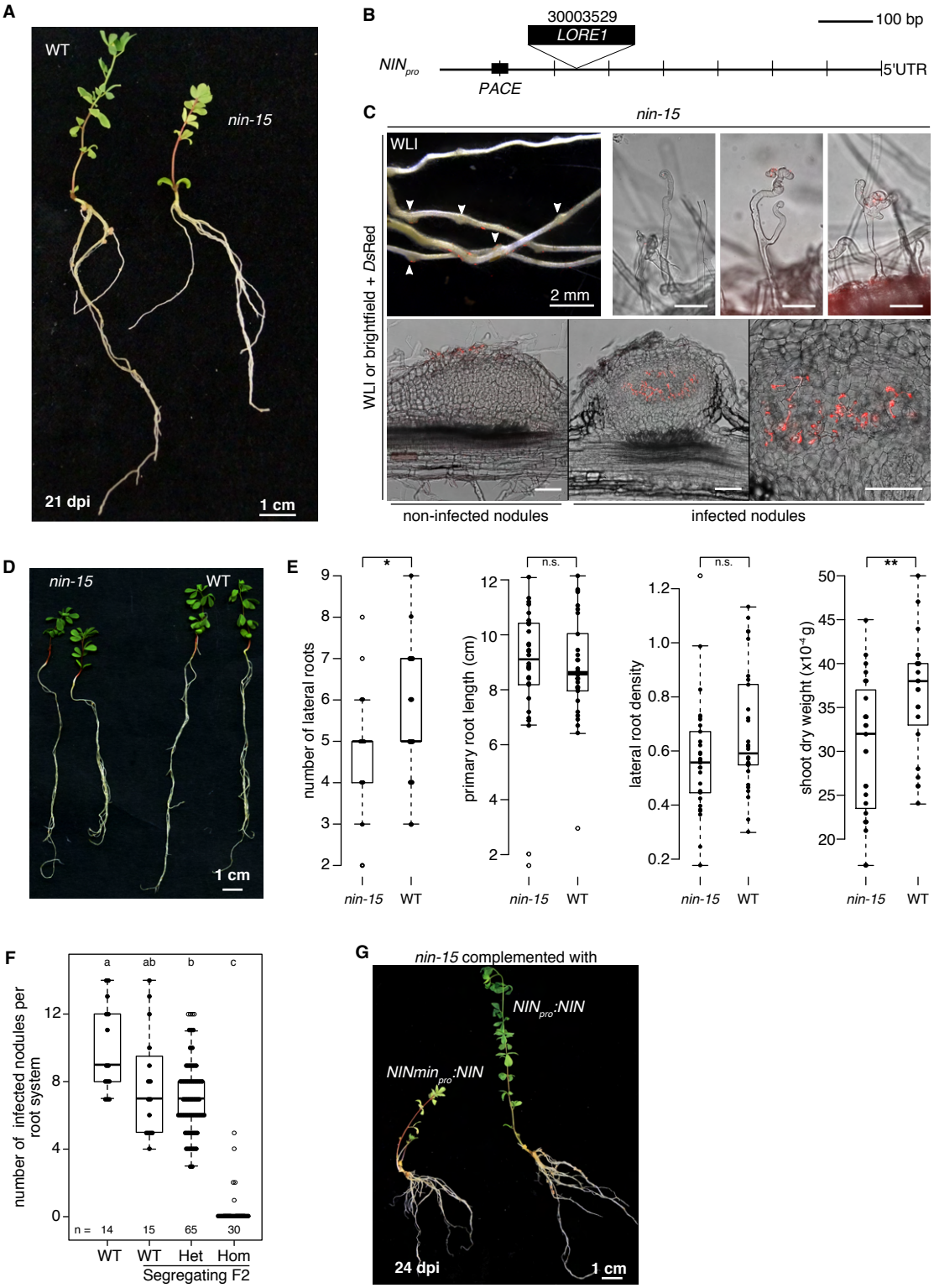


Figure 3 – figure supplement 1 I *L. japonicus nin-15* mutant phenotype. (**A**) A representative picture of *L. japonicus* wild-type (WT, left) and *nin-15* (right) plants 21 dpi with *M. loti Ds*Red. (**B**) Position of the *Lotus Retrotransposon 1* (*LORE1*) insertion within the *NIN* promoter in the *nin-15* mutant. (**C**) Representative pictures of *nin-15* root hairs and nodule sections 21 dpi with *M. loti Ds*Red. Forty-nine plants with a total number of 436 nodules were analysed: only four plants bore one or two IT(s) within root hairs and seven plants bore one or two infected nodule(s). Deformed or curled root hairs in the presence of *M. loti Ds*Red were abundant but infection threads were rarely found. Arrowheads: uninfected nodules. Unlabelled bars, 100 µm. (**D - E**) Phenotype of *nin-15* in the presence of a symbiosis-independent nitrogen source (15 mM KNO_3) for 28 days. (**D**) Pictures documenting the healthy status of *L. japonicus* WT and *nin-15* plants (compare (**D**) and (**A**)) and (**E**) quantitative assessment of parameters displayed in boxplots. Thirty plants per genotype were analysed. Each dot represents one plant. Lateral root density: number of lateral roots/primary root length (cm). The applied statistical method was pairwise *t*-test: $^{*}p < 0.05$; $^{**}p < 0.01$; n.s.: not significant. (**F**) Segregation analysis of *nin-15*. The applied statistical method was ANOVA with *post hoc* Tukey: $F_{3,120} = 84.1$, $p = 2\text{x}10^{-16}$. Different small letters indicate significant difference. (**G**) Representative pictures of *nin-15* plants with hairy roots transformed with the *NIN* gene driven by the *L. japonicus NIN* minimal promoter ($NINmin_{pro}$) or the 3 kb *NIN* promoter (NIN_{pro}) 24 dpi with *M. loti Ds*Red. WLI: white light illumination.

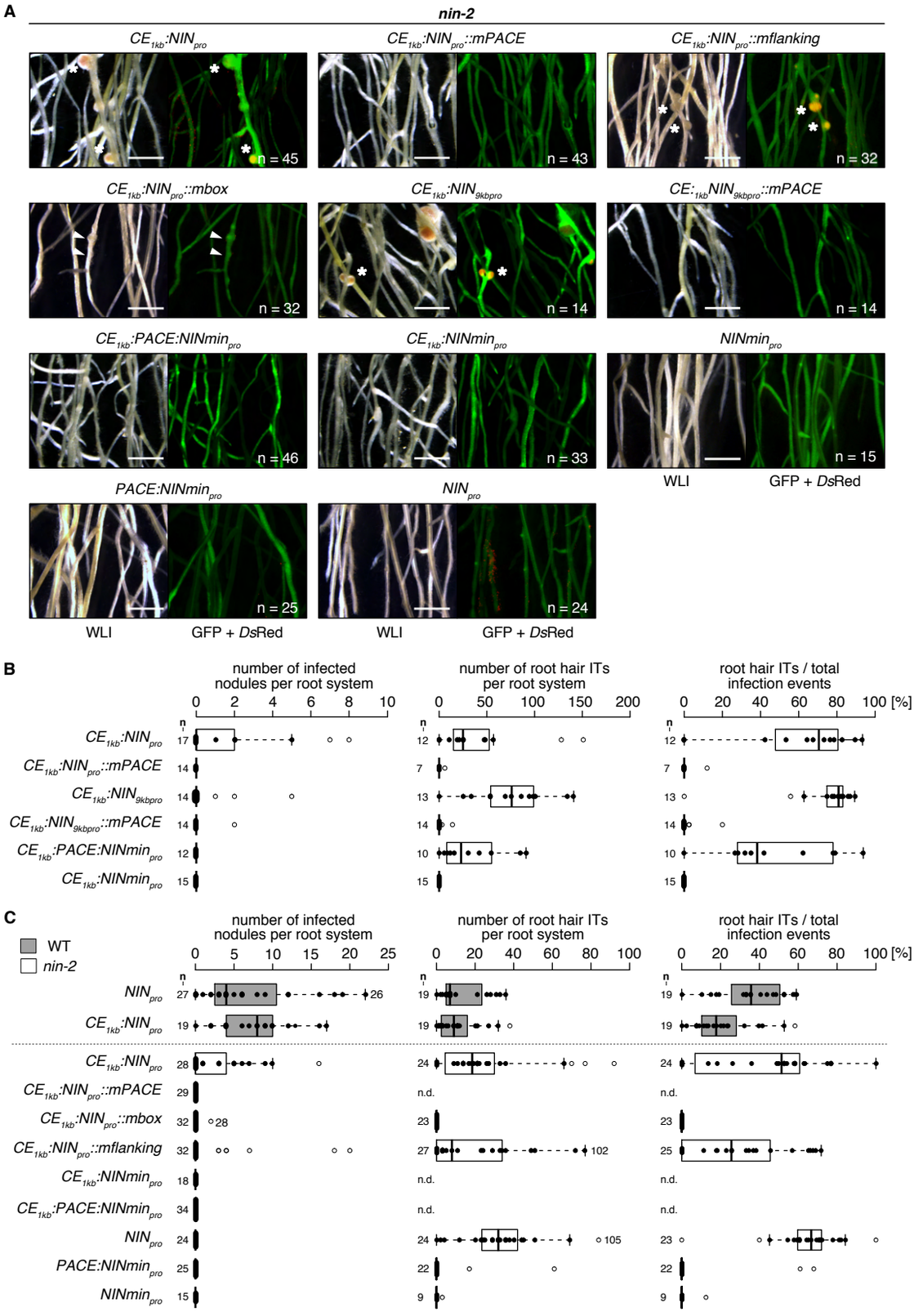


Figure 3 – figure supplement 2 | The *CYC-box* and flanking sequences of *PACE* are required for the complete restoration of the bacterial infection process in the *L. japonicus nin-2* mutant. *nin-2* roots were transformed with T-DNAs carrying a $Ubq10_{pro}$:*NLS-GFP* transformation marker in tandem with the *LjNIN* gene driven by either of the following promoter versions: the cytokinin element-containing region of 1 kb (CE_{1kb}) fused to the 3 kb or 9 kb *LjNIN* promoter (CE_{1kb}:NIN_{pro} or CE_{1kb}:NIN_{9kbpro}, respectively); CE_{1kb}:NIN_{pro} or CE_{1kb}:NIN_{9kbpro} with *PACE* mutated (CE_{1kb}:NIN_{pro}::*mPACE* or CE_{1kb}:NIN_{9kbpro}::*mPACE*, respectively); CE_{1kb}:NIN_{pro} carrying a mutated Cyclops binding site (*CYC-box*) (CE_{1kb}:NIN_{pro}::*mbox*); CE_{1kb}:NIN_{pro} carrying mutated sequences flanking the *CYC-box* in *PACE* (CE_{1kb}:NIN_{pro}::*mflanking*); CE_{1kb} fused to the *LjNIN* minimal promoter (CE_{1kb}:$NINmin_{pro}$); CE_{1kb} fused to *PACE* and to $NINmin_{pro}$ (CE_{1kb}:*PACE*:$NINmin_{pro}$); NIN_{pro}, *PACE*:$NINmin_{pro}$ or $NINmin_{pro}$. (**A**) Representative overview pictures of transgenic root systems. Roots were analysed 21 dpi with *M. loti Ds*Red. White asterisks and arrowheads: infected and non-infected nodules, respectively. Bars, 2 mm. (**B - C**) Boxplots displaying the number of root hair ITs or infected nodules and the percentage of root hair ITs among total infection events (sum of bacterial entrapments and ITs). Each dot represents one transgenic *nin-2* root system or root piece. *L. japonicus* WT roots transformed with NIN_{pro}:*NIN* or CE_{1kb}:NIN_{pro}:*NIN* were included as controls. Note the loss of restoration of nodules and IT formation associated with the mutation of *PACE* or only the *CYC-box* in *PACE*; and the reduction of same when sequences flanking the *CYC-box* in *PACE* were mutated. n: number of transgenic root systems or root pieces analysed. Numbers above the boxplots: the value of individual data points outside of the plotting area. n.d.: not determined. WLI: white light illumination.

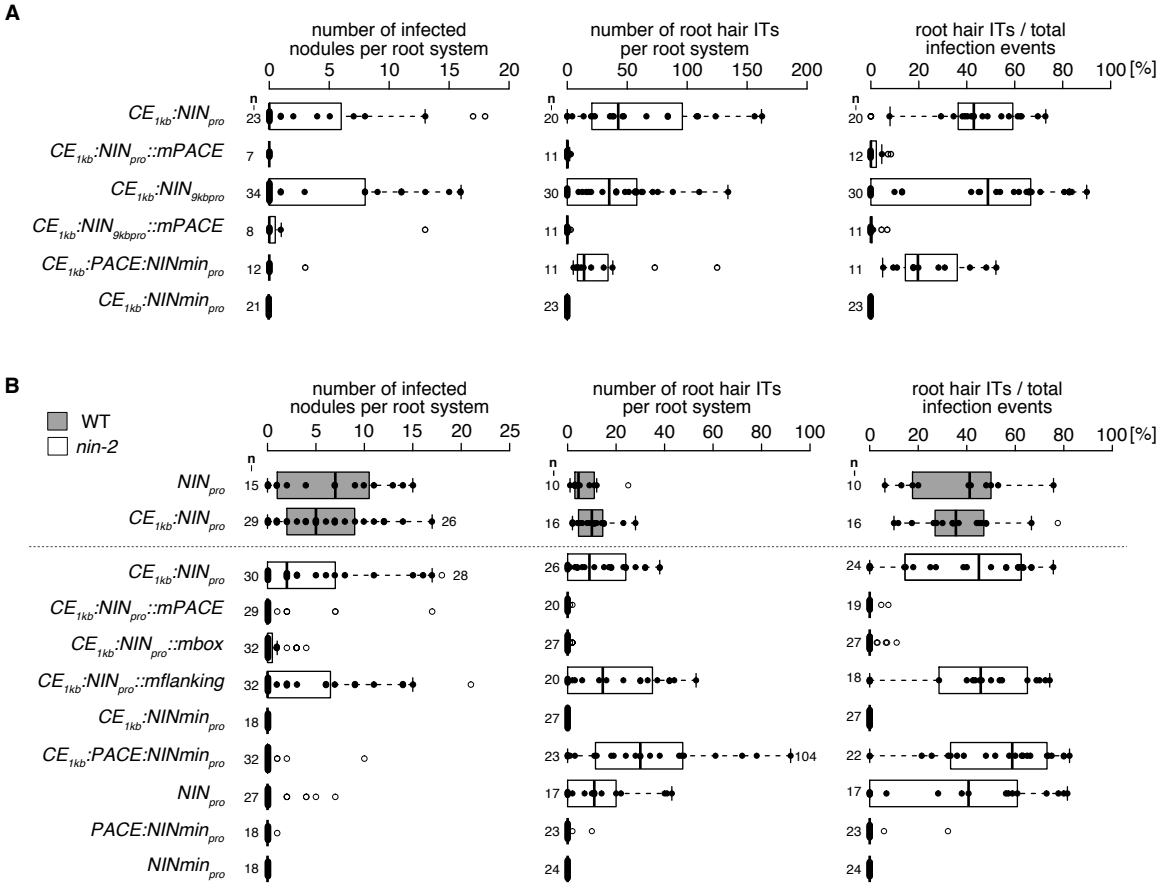


Figure 3 – figure supplement 3 | The *CYC-box* and flanking sequences of *PACE* are required for the complete restoration of the bacterial infection process in the *L. japonicus nin-2* mutant. Roots were from a subset of plants from the same experiment depicted in **Figure 3 – figure supplement 2** but analysed 35 dpi with *M. loti Ds*Red. (**A - B**) Boxplots displaying the number of root hair ITs or infected nodules and the percentage of root hair ITs among total infection events (sum of bacterial entrapments and ITs). Each dot represents one transgenic *nin-2* root system or root piece. *L. japonicus* WT roots transformed with NIN_{pro}:*NIN* or CE_{1kb}:NIN_{pro}:*NIN* were included as controls. Note that the results follow the same trend as those obtained 21 dpi with *M. loti Ds*Red (**Figure 3 – figure supplement 2**). n: number of transgenic root systems or root pieces analysed. Numbers above the boxplots: the value of individual data points outside of the plotting area.

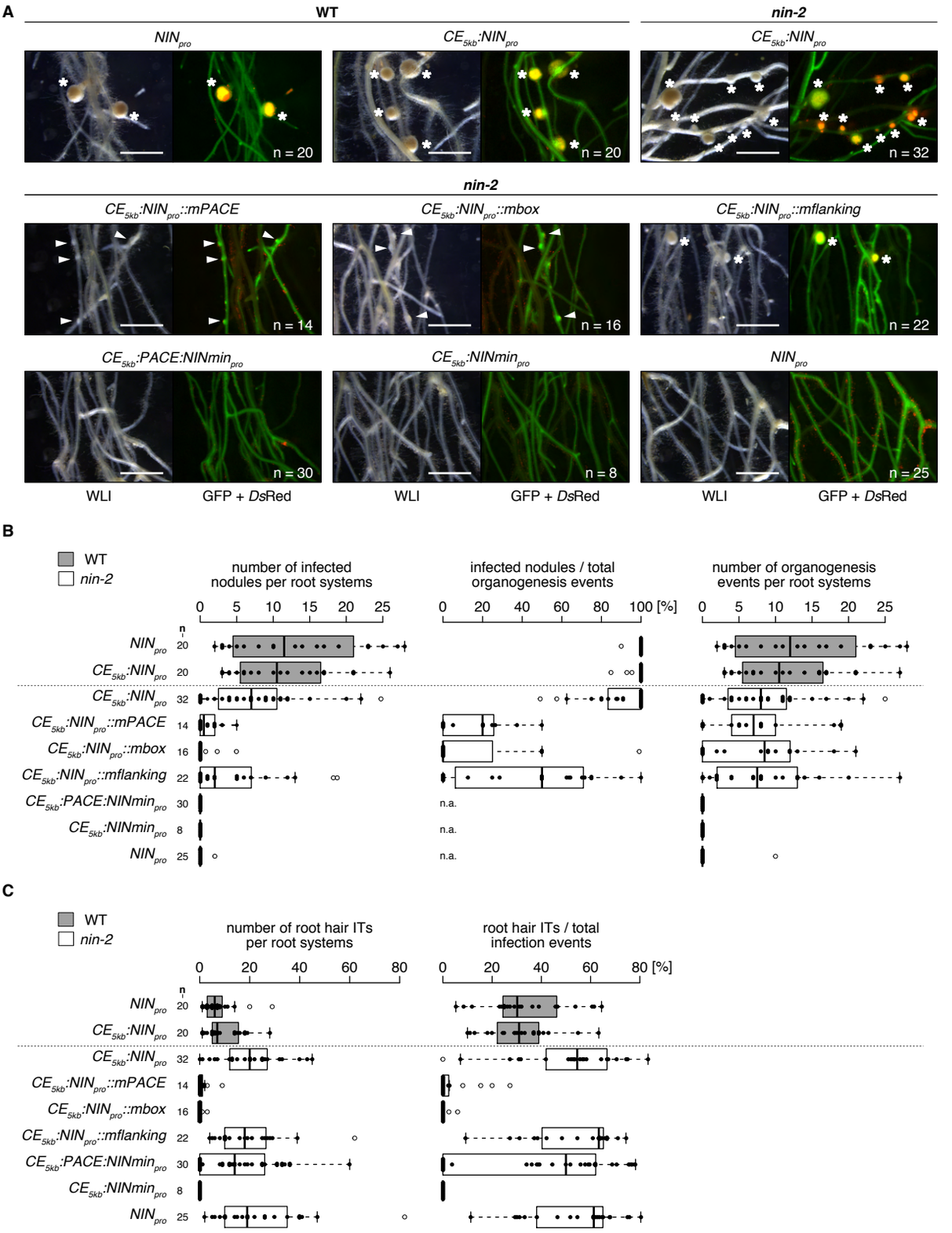


Figure 3 – figure supplement 4 | The *CYC-box* and flanking sequences of *PACE* are required for the complete restoration of the bacterial infection process but are dispensable for the nodule organogenesis process in the *L. japonicus nin-2* mutant. *nin-2* roots were transformed with T-DNAs carrying a $Ubq10_{pro}$:*NLS-GFP* transformation marker in tandem with the *LjNIN* gene driven by either of the following promoter versions: the cytokinin element-containing region of 5 kb (CE_{5kb}) fused to the 3 kb *LjNIN* promoter (CE_{5kb}:NIN_{pro}); CE_{5kb}:NIN_{pro} with *PACE* mutated (CE_{5kb}:NIN_{pro}::*mPACE*); CE_{5kb}:NIN_{pro} carrying a mutated Cyclops binding site (*CYC-box*) (CE_{5kb}:NIN_{pro}::*mbox*); CE_{5kb}:NIN_{pro} carrying mutated sequences flanking the *CYC-box* in *PACE* (CE_{5kb}:NIN_{pro}::*mflanking*); CE_{5kb} fused to the *LjNIN* minimal promoter (CE_{5kb}:$NINmin_{pro}$); CE_{5kb} fused to *PACE* and to $NINmin_{pro}$ (CE_{5kb}:*PACE*:$NINmin_{pro}$) or NIN_{pro}. (**A**) Representative overview pictures of transgenic root systems. Roots were analysed 21 dpi with *M. loti Ds*Red. White asterisks and arrowheads: infected and non-infected nodules, respectively. Bars, 2 mm. (**B**) Boxplots displaying the number of infected nodules, the percentage of infected nodules among total organogenesis events (sum of infected and non-infected nodules) and the number of organogenesis events. (**C**) Boxplots displaying the number of root hair ITs and the percentage of root hair ITs among total infection events (sum of bacterial entrapments and ITs). Each dot represents one transgenic *nin-2* root system or root piece. *L. japonicus* WT roots transformed with NIN_{pro}:*NIN* or CE_{5kb}:NIN_{pro}:*NIN* were included as controls. Note that the mutation of *PACE* or only the *CYC-box* in *PACE* led to an almost complete loss of IT formation and infected nodules per root system while nodule organogenesis was not significantly reduced; and that mutation of sequences flanking the *CYC-box* in *PACE* led to a reduction of the number of infected nodules per root systems. n: number of transgenic root systems or root pieces analysed. Numbers above the boxplots: the value of individual data points outside of the plotting area. n.a.: not applicable. WLI: white light illumination.

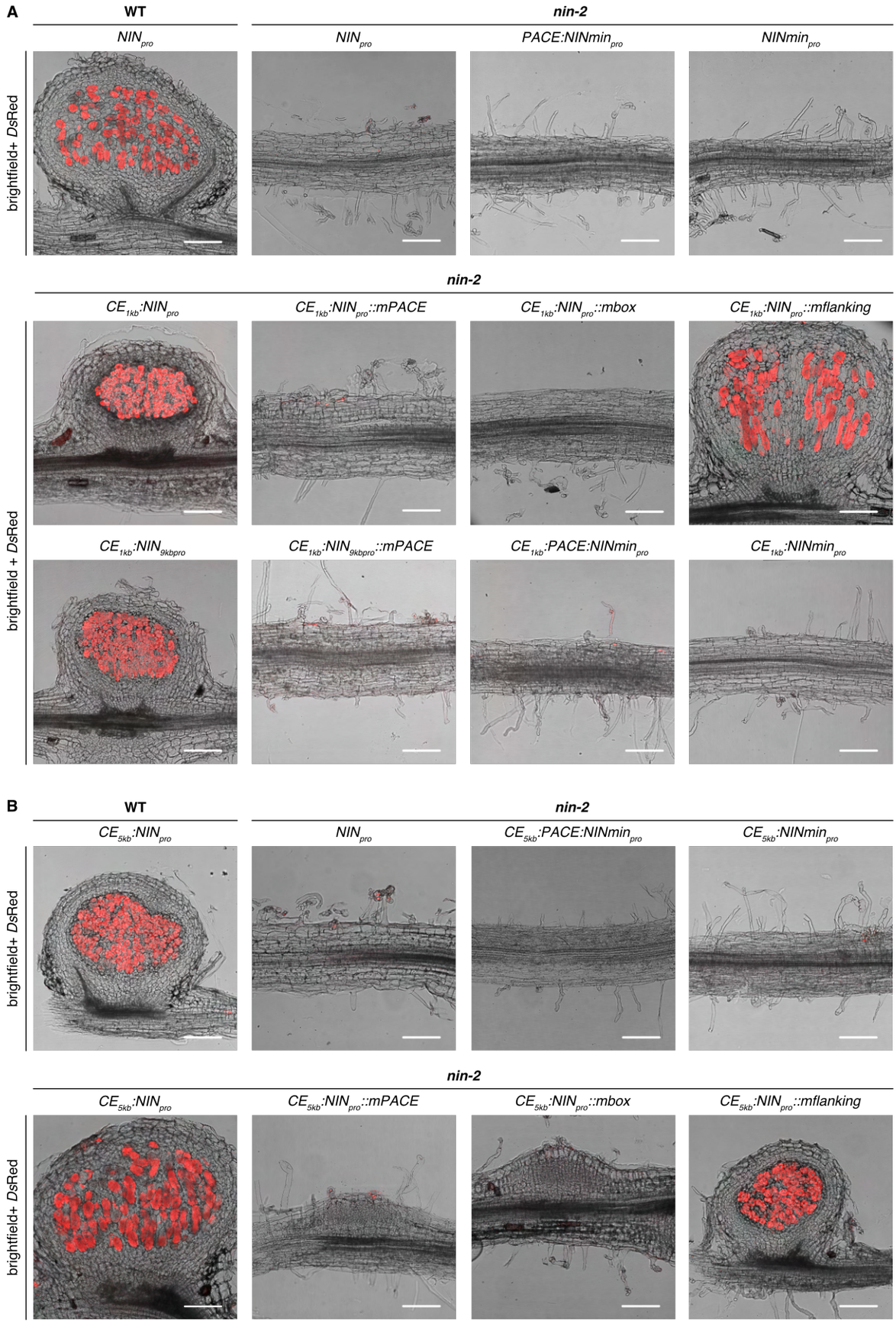


Figure 3 – figure supplement 5 | The *CYC-box* and flanking sequences of *PACE* are required for the complete restoration of the bacterial infection process but are dispensable for the nodule organogenesis process in the *L. japonicus nin-2* mutant. Pictures of nodule sections or roots from *L. japonicus nin-2* roots 21 dpi with *M. loti Ds*Red from the same experiments depicted in **Figure 3 – figure supplement 2** (**A**) and **Figure 3 – figure supplement 4** (**B**). Nodule sections from *L. japonicus* WT roots transformed with NIN_{pro}:*NIN* and $CE_{5kb}NIN_{pro}$:*NIN* were included for comparison. Note that when the cytokinin element-containing region of 1 kb was fused to NIN_{pro} nodule organogenesis was abolished by mutation of *PACE* or only the *CYC-box* in *PACE* and that these mutations did not abolish organogenesis when the cytokinin element-containing region of 5 kb was fused to NIN_{pro}. Bars, 100 µm.

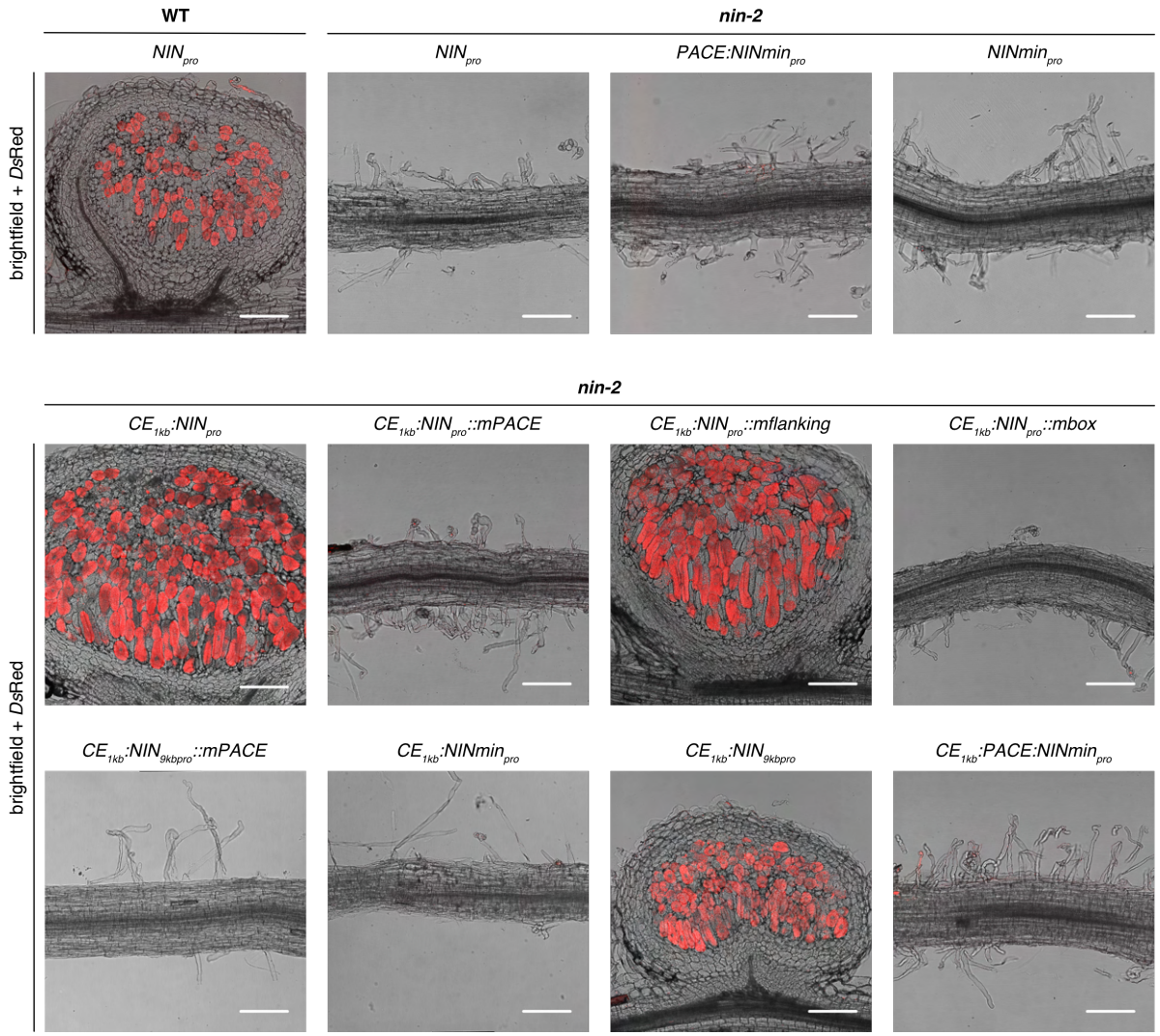


Figure 3 – figure supplement 6 I The *CYC-box* and flanking sequences of *PACE* are required for the complete restoration of the bacterial infection process in the *L. japonicus nin-2* mutant. Pictures of nodule sections or roots from *L. japonicus nin-2* roots 35 dpi with *M. loti Ds*Red from the same experiments depicted in **Figure 3 – figure supplement 3**. Upper left corner: a nodule section from a *L. japonicus* WT root transformed with NIN_{pro}:*NIN* was included for comparison. Bars, 100 µm.

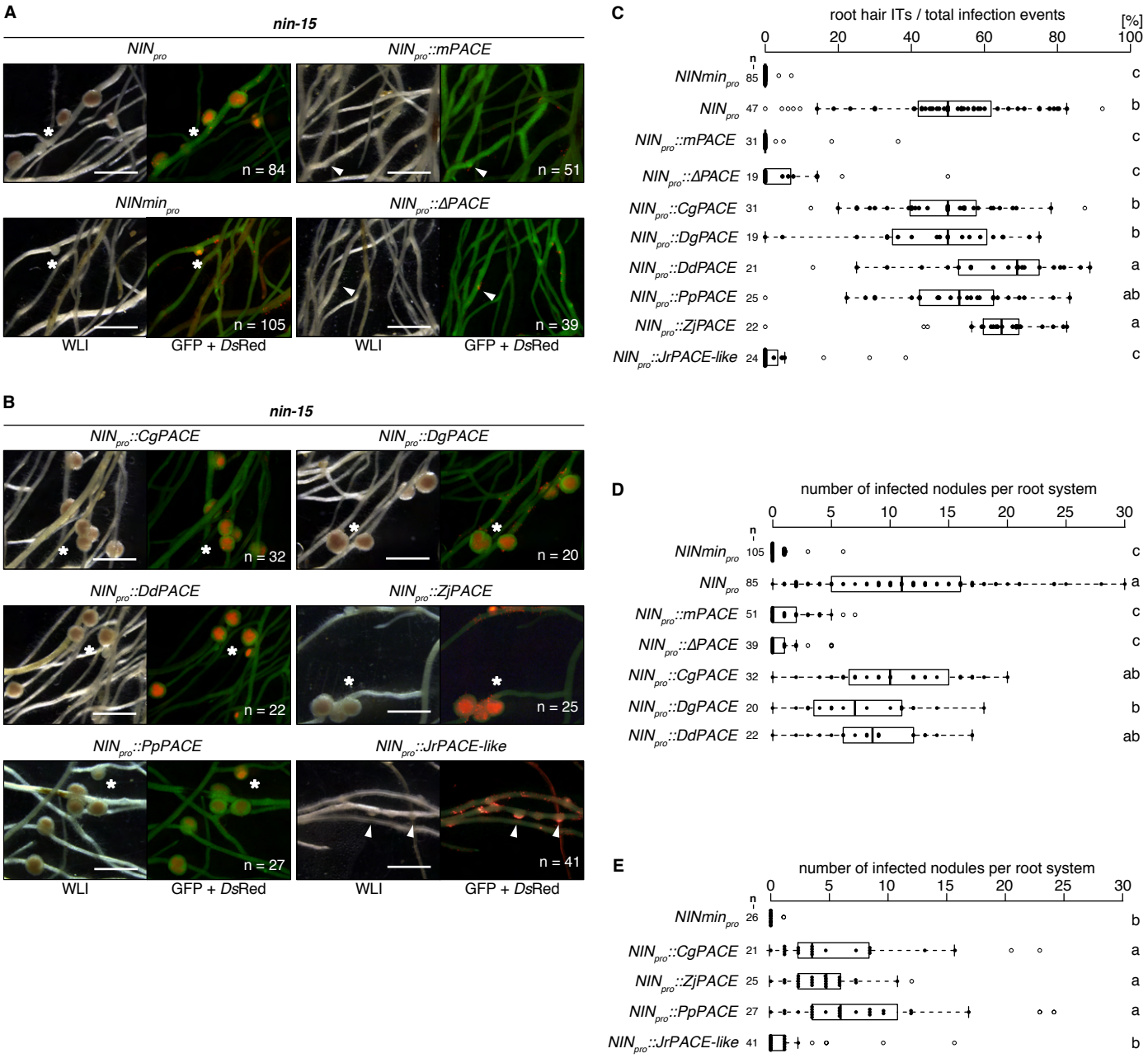
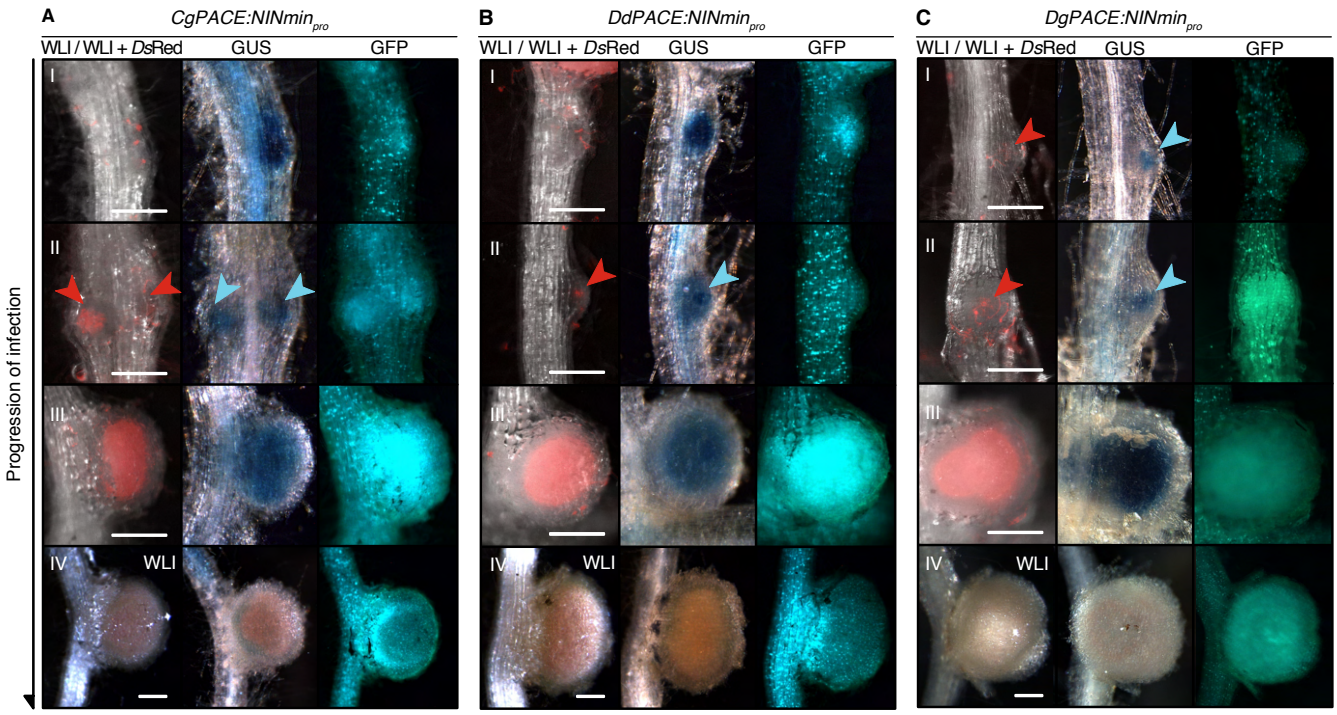


Figure 3 – figure supplement 7 | *PACEs* from FaFaCuRo species are functionally equivalent in restoring bacterial infection in the *L. japonicus nin-15* mutant. *L. japonicus nin-15* roots were transformed with T-DNAs carrying a *Ubq10*$_{pro}$*:NLS-GFP* transformation marker in tandem with the *LjNIN* gene driven by either of the following promoters: (**A**) the 3 kb *LjNIN* promoter (*NIN*$_{pro}$), the *LjNIN* minimal promoter (*NINmin*$_{pro}$), the 3 kb *LjNIN* promoter with *PACE* deleted (*NIN*$_{pro}$*::ΔPACE*) or mutated (*NIN*$_{pro}$*::mPACE*); (**B**) the 3 kb *LjNIN* promoter with *LjPACE* replaced with either of the *PACE* sequence variants from nodulating or non-nodulating FaFaCuRo species and analysed 21 dpi with *M. loti Ds*Red. (**A - B**) Representative overview pictures of *nin-15* transgenic roots systems. Sections of representative nodules are displayed in **Figure 3**. Note the drastic reduction of restoration of infection in nodules and root hairs associated with the mutation or deletion of *PACE* as well as the replacement of *PACE* with *JrPACE-like* in the context of the *LjNIN* promoter. White asterisks and arrowheads: infected and non-infected nodules, respectively. (**C - E**) Boxplots displaying **c,** the percentage of root hair ITs among total infection events (sum of bacterial entrapments and ITs) and (**D - E**) the number of infected nodules from two independent experiments. Each dot in (**D - E**) represents one *nin-15* transgenic root system. (**C**) displays merged data from all experiments as the percentage represents a normalised value calculated for each root piece (see **Figure 3 – Source Data 1**). n: number of transgenic root systems or root pieces analysed. For species abbreviations see **Figure 1 – figure supplement 5A**. The applied statistical method was ANOVA with *post hoc* Tukey: (**C**) $F_{9,313} = 106.7$, $p < 2\times10^{-16}$; (**D**) $F_{6,346} = 82.89$, $p < 2\times10^{-16}$; (**E**) $F_{4,135} = 20.18$, $p = 4.76\times10^{-13}$. Different small letters indicate significant differences. Bars, 2 mm. WLI: white light illumination.



D

promoter:GUS	Time point (dpi)		
	0	10 -14	≥ 21
CgPACE:NINmin$_{pro}$	7[0]/7	14[23]/19	15[2]/20
DdPACE:NINmin$_{pro}$	8[0]/8	15[23]/18	17[2]/22
DgPACE:NINmin$_{pro}$	n.d.	35[23]/46	28[2]/38

#/#: root systems exhibiting GUS activity / total root systems analysed
[0] vasculature (of roots or nodules) and/or root tips
[1] epidermis (including root hairs)
[2] nodule primordia
[3] central tissue of nodules

Figure 3 – figure supplement 8 I Spatio-temporal *GUS* expression driven by *PACE* variants in *L. japonicus* roots during the bacterial infection process. *L. japonicus* WT roots were transformed with T-DNAs carrying a *Ubq10*$_{pro}$*:NLS-GFP* transformation marker together with a *GUS* reporter gene driven by either of the *PACE* variants from nodulating FaFaCuRo species fused to the *LjNIN* minimal promoter (*NINmin*$_{pro}$). For species abbreviations and experimental details see **Figure 1 – figure supplement 5A and Figure 2 – figure supplement 1**, respectively. Note the overlapping bacterial invasion zone and *PACE:NINmin*$_{pro}$*:GUS* expression in early infection stages (red and blue arrowheads in (**A - C**)). Red arrowheads: *M. loti Ds*Red. Blue arrowheads: GUS activity in nodule primordia. Only pictures taken under white light illumination (WLI) are displayed for nodules in panel VI to reveal the pink colour of leghemoglobin, characteristic for mature and fully infected nodules. Note that like *LjPACE*, the *PACE* variants-driven *GUS* expressions were absent at this stage (panel IV in (**A - C**) and panel IV in **Figure 2 – figure supplement 1D**). (**D**) Quantification of transgenic root systems exhibiting *GUS* expression in different cell types and tissues exemplarily displayed in (**A - C**). n.d.: not determined. Bars, 250 µm.

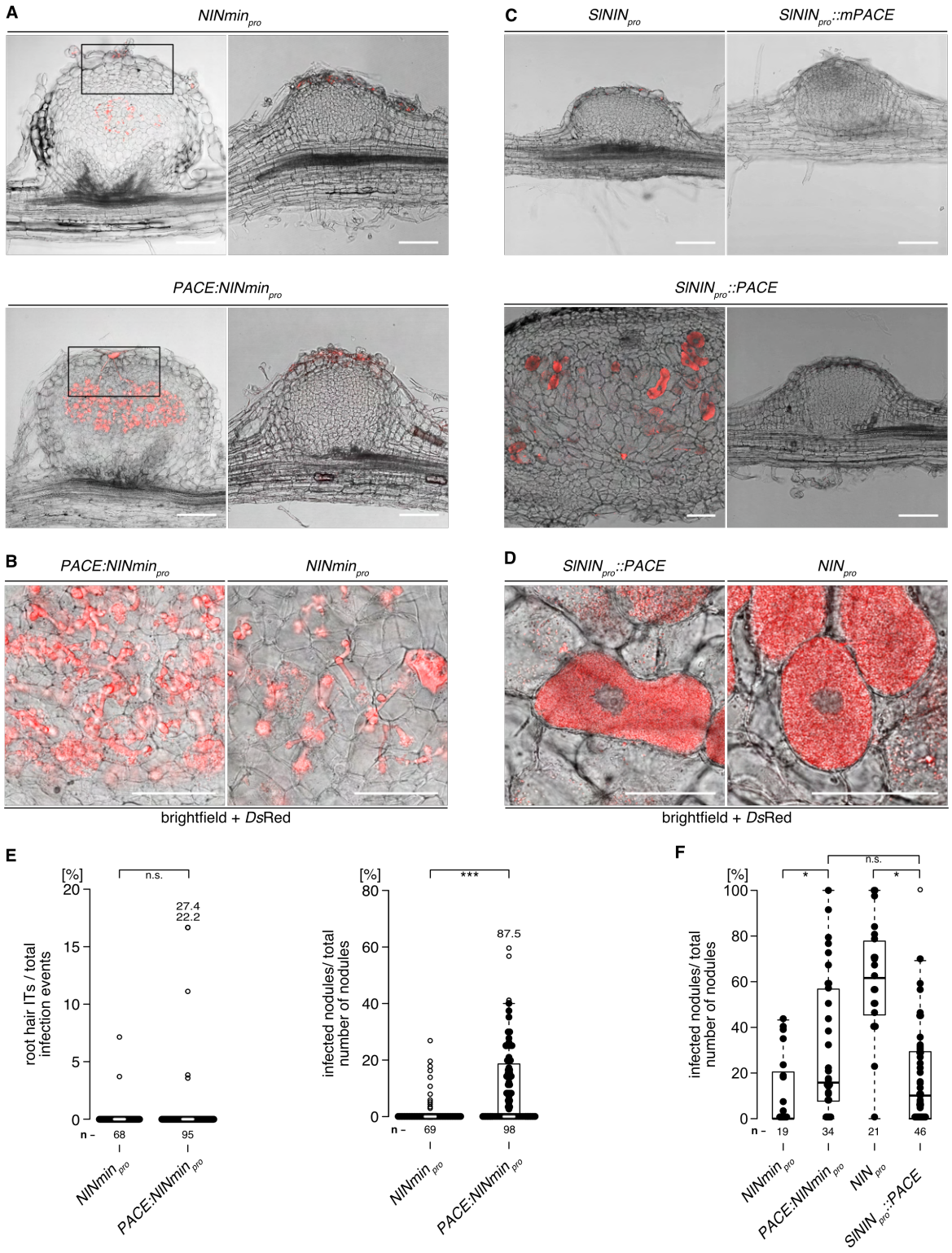


Figure 4 – figure supplement 1 | *PACE* alone or in the context of the *S. lycopersicum NIN* promoter (a species outside of the FaFaCuRo clade) enables IT formation in the cortex. (**A - D**) Representative pictures of sections of nodules formed on *L. japonicus nin-15* roots transformed with T-DNAs carrying a *Ubq10*$_{pro}$*:NLS-GFP* transformation marker together with the *LjNIN* gene driven by either of the following promoters: (**A - B**) the *L. japonicus NIN* minimal promoter (*NINmin*$_{pro}$) or *PACE* fused to *NINmin*$_{pro}$ (*PACE:NINmin*$_{pro}$); (**C - D**) the 3 kb *S. lycopersicum NIN* promoter (*SlNIN*$_{pro}$), the 3 kb *SlNIN* promoter with mutated *PACE* (*SlNIN*$_{pro}$*::mPACE*) or with *L. japonicus PACE* inserted (*SlNIN*$_{pro}$*::PACE*), 21 dpi with *M. loti Ds*Red (from the same experiments depicted in **Figure 4**). Black rectangles in (**A**) demarcate the enlarged area displayed in **Figure 4A and 4B** to focus on the initial infection structures. Note the absence of cells filled with symbiosomes in nodules transformed with the *LjNIN* gene driven by *PACE:NINmin*$_{pro}$ or *NINmin*$_{pro}$. By contrast, infected cells were often filled with symbiosomes in the *SlNIN*$_{pro}$*::PACE:NIN*-transformed nodules, like those resulted by *NIN*$_{pro}$*:NIN* (see (**C**) and compare the two sections in (**D**)). (**E - F**) Boxplots displaying the percentage of root hair ITs among total infection events (sum of bacterial entrapments and ITs) or the percentage of infected nodules among total number of nodules (**E**) 21 dpi and (**F**) 35 dpi with *M. loti Ds*Red, respectively. Each dot represents one *nin-15* transgenic root system or root piece. (**E**) displays results from an independent repetition from the experiment depicted in **Figure 4**. n: number of transgenic root systems or root pieces analysed. Numbers above the boxplots: the value of individual data points outside of the plotting area. The applied statistical method was Fisher's exact test: *$p < 0.05$; ***$p < 0.001$; n.s.: not significant. Bars, (**A and C**) 100 µm; (**B and D**) 50 µm.

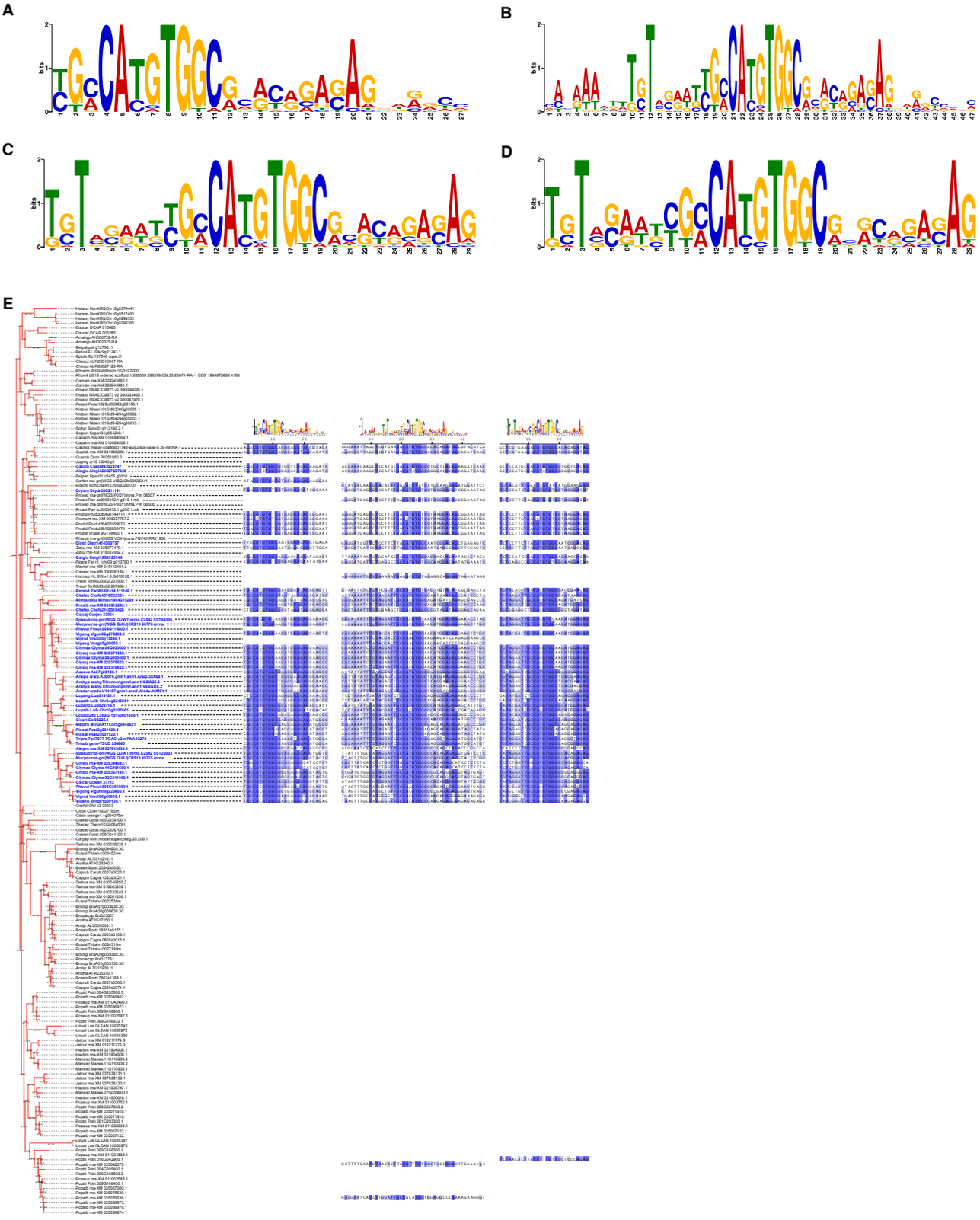


Figure 1 – figure supplement 1 | Discovery of *PACE* by MEME analyses. (**A - D**) Consensus sequence of the Position Weight Matrix identified by MEME analyses using the regions upstream of the translational start site (ATG) thus representing the promoters and 5'UTRs of *NIN* and *NLP* genes from 37 angiosperm species (see Methods): (**A**) in a discriminative search for a motif that is present in 3 kb upstream regions of the *NIN* genes from nodulating FaFaCuRo species, but absent in 3 kb upstream regions of the *NIN* genes from species outside of the FaFaCuRo clade and absent in the 3 kb upstream regions of *NLP* genes; (**B**) in an independent, non-discriminative search in the 3 kb upstream regions of the *NIN* genes from nodulating species revealing that the most conserved nucleotides span a larger region than (**A**); (**C**) the resulted most conserved 29 nucleotides derived from upstream regions of *NIN* genes of nodulating FaFaCuRo species; (**D**) the resulted most conserved 29 nucleotides (*PACE*) derived from the upstream region of one representative *NIN* gene per species of nodulating FaFaCuRo species. (**E**) Motif analyses by FIMO using the upstream regions of the *NIN* and *NLP* genes from an expanded list of 163 species (see Methods). Left: pruned tree from the whole *NLP* tree (demarcated in the black rectangle in **Figure 1 – figure supplement 2**) corresponding to the *NIN* orthologs. Right: three versions of consensus sequence (left to right, (**A, B and D**), respectively) were used to retrieve motifs from the 5 kb upstream region of *NIN* genes *via* FIMO search, the output of which is displayed underneath the consensus sequences. Sequence names of nodulating species are coloured in blue. Blank lines represent the absence of significant motifs.

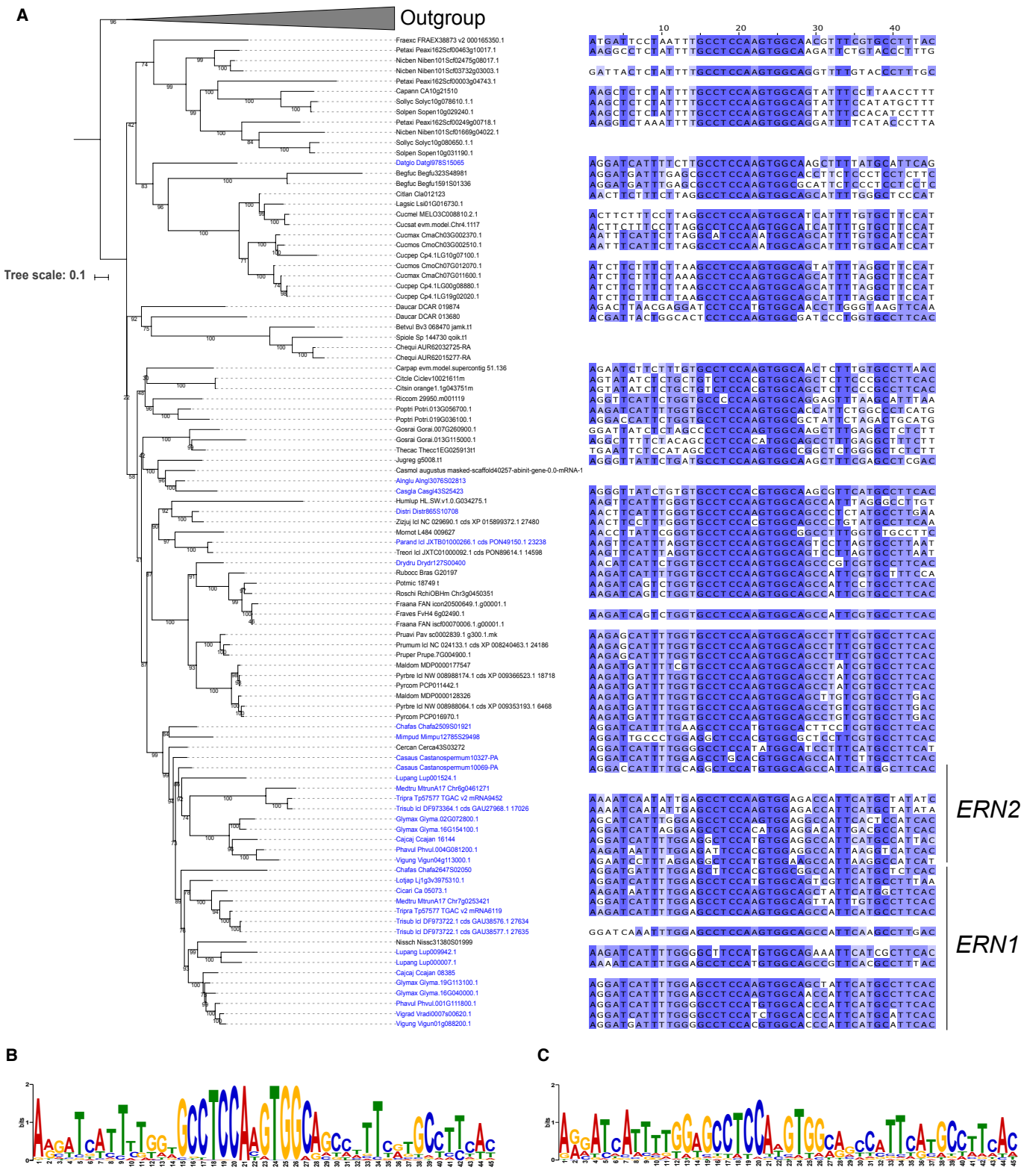


Figure 1 – figure supplement 3 | The presence of a motif encompassing the previously identified Cyclops binding site within the regions upstream of the translational start site (ATG) thus representing the promoters and 5'UTRs of *ERN1* genes extends beyond the FaFaCuRo clade. (**A**) Maximum likelihood of *ERN1/ERN2* (model: JTT+F+R5) from 87 species (**Supplementary file 1**) based on protein sequences. Tree was rooted using sequences from the *ERN1/ERN2* paralog *ERN3* (outgroup clade collapse and indicated by grey triangle marked "Outgroup"). Species belonging to the monophyletic FaFaCuRo clade (Stevens, 2017) are indicated by thicker black branches. *ERN1* and *ERN2* subclades are indicated. Sequence names of nodulating species are coloured in blue. These conserved motifs identified by MEME analyses in the 3 kb upstream regions were aligned and displayed next to species names. While the presence of a conserved Cyclops binding site containing motif in species outside of the FaFaCuRo clade is consistent with the role of *RAM1* in arbuscular mycorrhizal symbiosis, a function of *ERN1* in arbuscular mycorrhizal symbiosis has not been described and suggests a yet unidentified function of *ERN1* outside of the FaFaCuRo clade. (**B**) Consensus sequence of the Position Weight Matrix identified by MEME analyses of all the species. (**C**) Consensus sequence of the Position Weight Matrix identified by MEME analyses when analysing only the nodulating species.

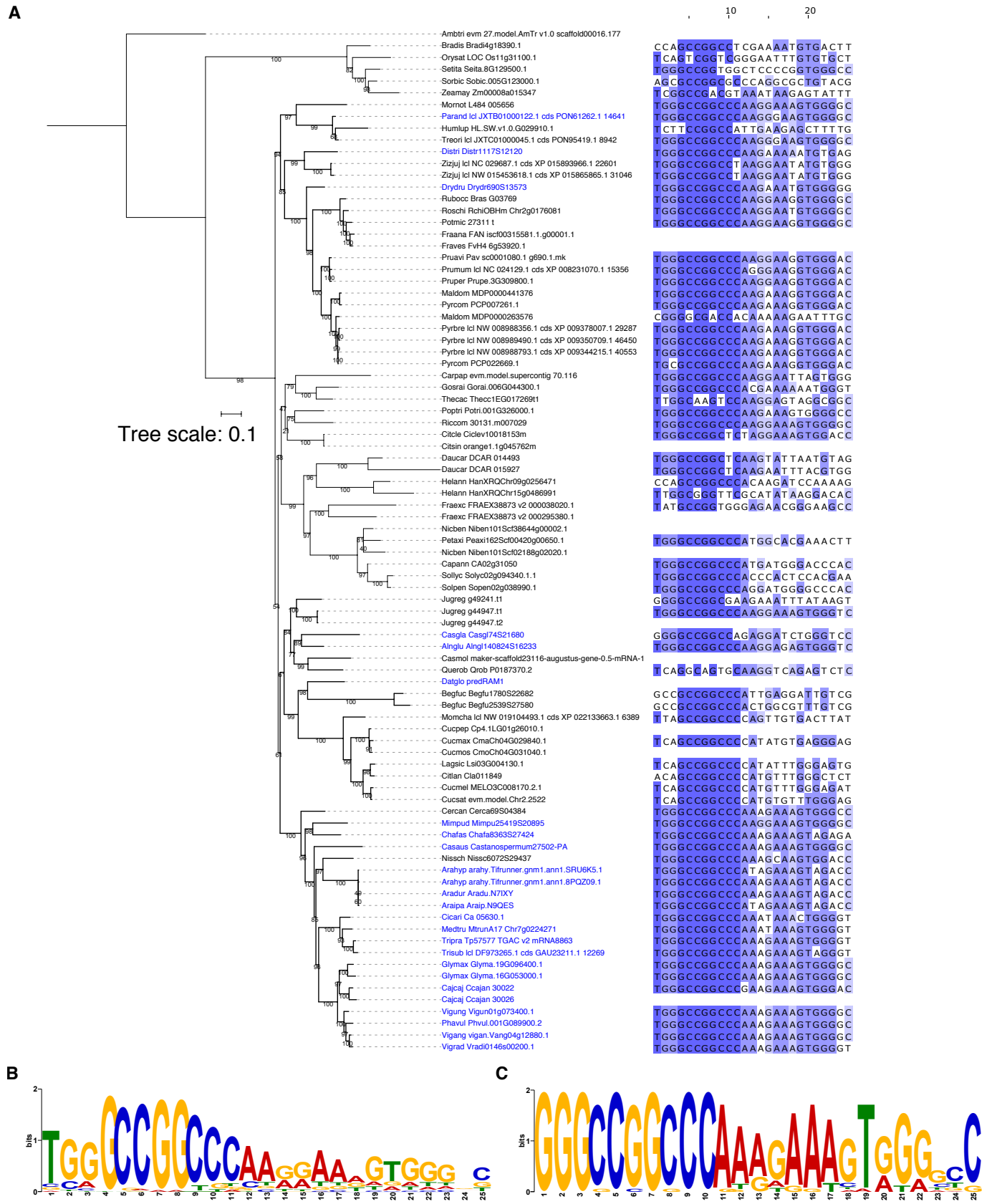
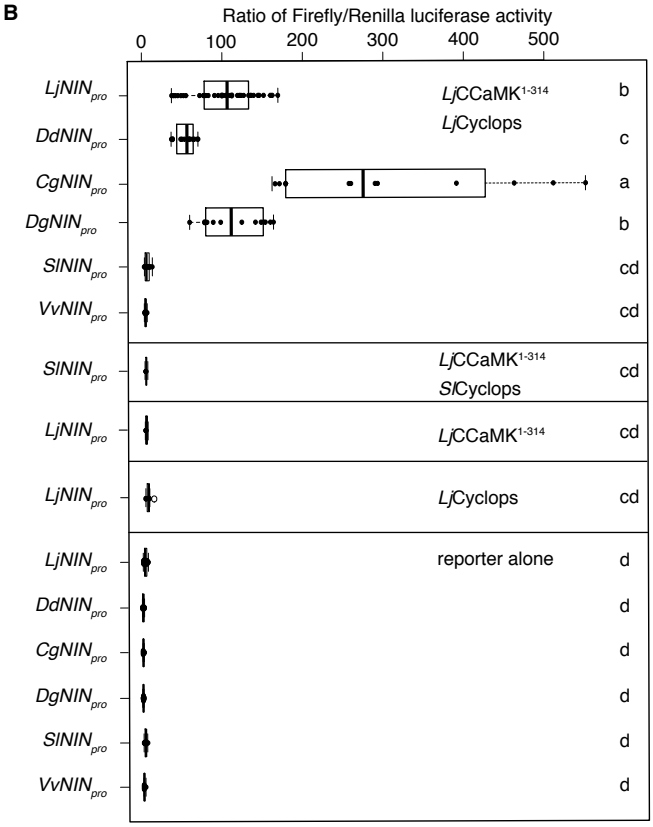


Figure 1 – figure supplement 4 I The presence of a motif encompassing the previously identified Cyclops binding site within the regions upstream of the translational start site (ATG) thus representing the promoters and 5'UTRs of *RAM1* genes extends beyond the FaFaCuRo clade. (**A**) Maximum likelihood of *RAM1* (model: JTT+F+R4) from 87 species (**Supplementary file 1**) engaging in arbuscular mycorrhizal symbiosis. Tree was rooted on the early-diverging angiosperm *Amborella trichopoda*. Species belonging to the monophyletic FaFaCuRo clade (Stevens, 2017) are indicated by thicker black branches. Sequence names of nodulating species are coloured in blue. These conserved motifs identified by MEME analyses in the 3 kb upstream regions are aligned and displayed next to species names. Blank lines represent the absence of identified motif due to genome contiguity issues. (**B**) Consensus sequence of the Position Weight Matrix identified by MEME analyses when analysing all the mycorrhizal species. (**C**) Consensus sequence of the Position Weight Matrix identified by MEME analyses when analysing only the nodulating species.

A

Abbre-viation	Species	Plant Order	RNS	FaFaCuRo member
Lj	*Lotus japonicus*	Fabales	Yes	Yes
Cg	*Casuarina glauca*	Fagales	Yes	Yes
Jr	*Juglans regia*	Fagales	No	Yes
Dd	*Dryas drummondii*	Rosales	Yes	Yes
Zj	*Ziziphus jujuba*	Rosales	No	Yes
Pp	*Prunus persica*	Rosales	No	Yes
Dg	*Datisca glomerata*	Cucurbitales	Yes	Yes
Sl	*Solanum lycopersicum* (tomato)	Solanales	No	No
Vv	*Vitis vinifera*	Vitales	No	No



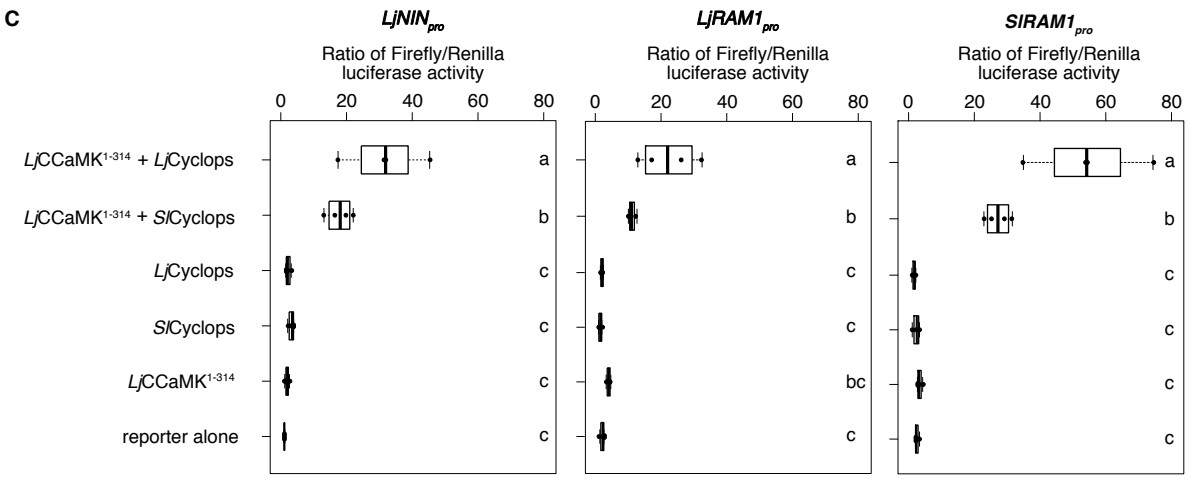


Figure 1 – figure supplement 5 | Transcriptional activation of *NIN promoter:Firefly luciferase* reporter gene by CCaMK$^{1-314}$/Cyclops is restricted to *NIN* promoters from species of the FaFaCuRo clade. *Nicotiana benthamiana* leaf cells were transformed with T-DNAs carrying a *Firefly luciferase* reporter gene driven by either of the indicated promoters in tandem with the *AtACT2$_{pro}$:Renilla luciferase* reporter fusion that provides a quantitative internal standard. (**A**) List of species within the FaFaCuRo clade (light red shade) and outside (light grey shade) and abbreviations. (**B**) Reporter gene activation by *L. japonicus* CCaMK$^{1-314}$/Cyclops via *NIN* promoters (*NIN$_{pro}$*) originating from listed species. (**C**) Comparison of the transactivation potential of Cyclops versions from *L. japonicus* and *S. lycopersicum*. Note that the expression of the *Firefly luciferase* reporter gene driven by *LjNIN$_{pro}$*, the *RAM1* promoters from *L. japonicus* and *S. lycopersicum* (*LjRAM1$_{pro}$* and *SlRAM1$_{pro}$*, respectively) was induced in the presence of CCaMK$^{1-314}$/Cyclops regardless of the origin of Cyclops. In contrast, the transactivation failed with the *SlNIN* promoter (panel (**A**)). The applied statistical method was ANOVA with *post hoc* Tukey: (**B**), $F_{14,214} = 71.07$, $p < 2\text{x}10^{-16}$; (**C**), plots from left to right: $F_{5,18} = 20.58$, $p = 7.14\text{x}10^{-7}$; $F_{5,18} = 25.38$, $p = 1.45\text{x}10^{-7}$ and $F_{5,18} = 40.49$, $p = 3.55\text{x}10^{-9}$, respectively. Different small letters indicate significant differences.

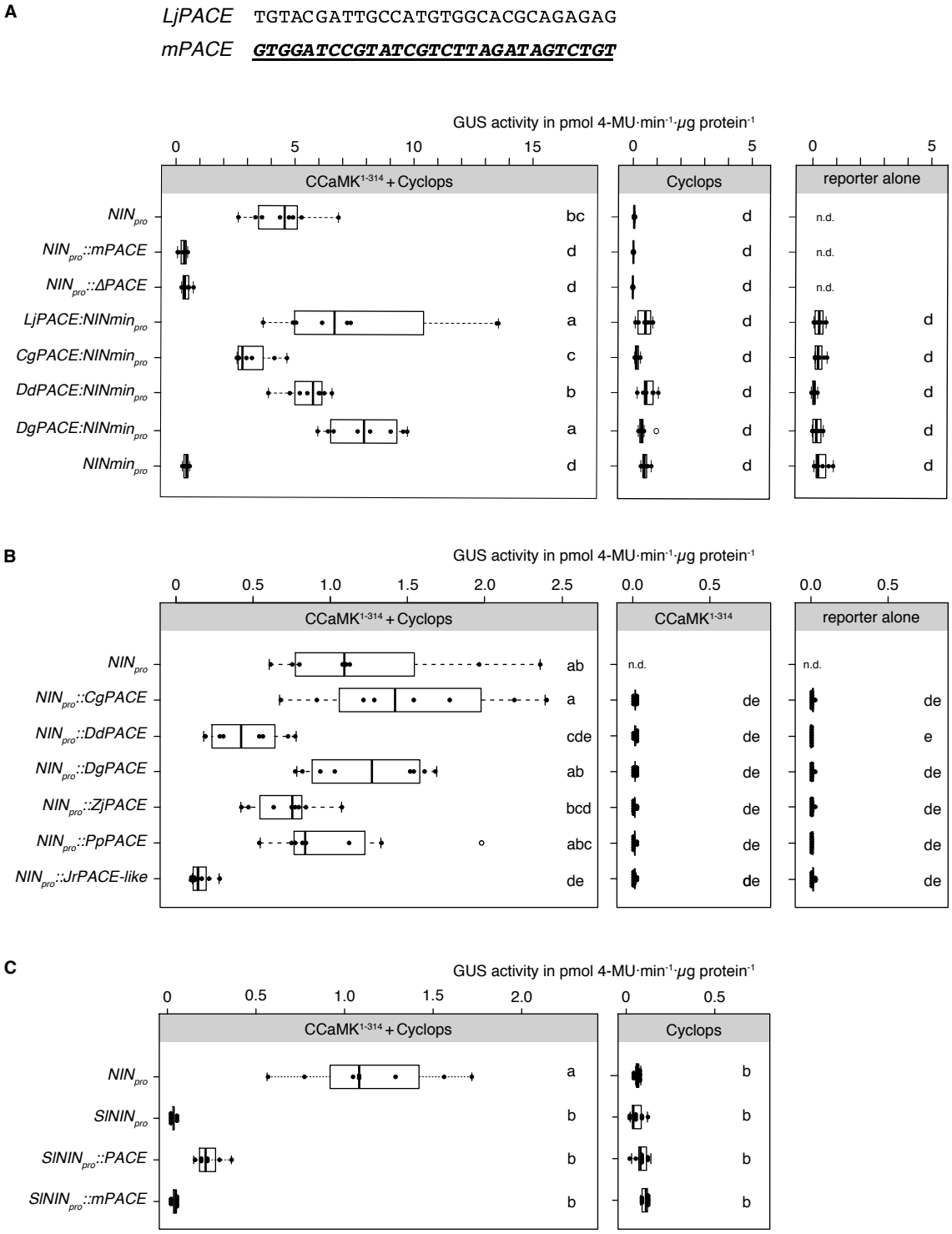
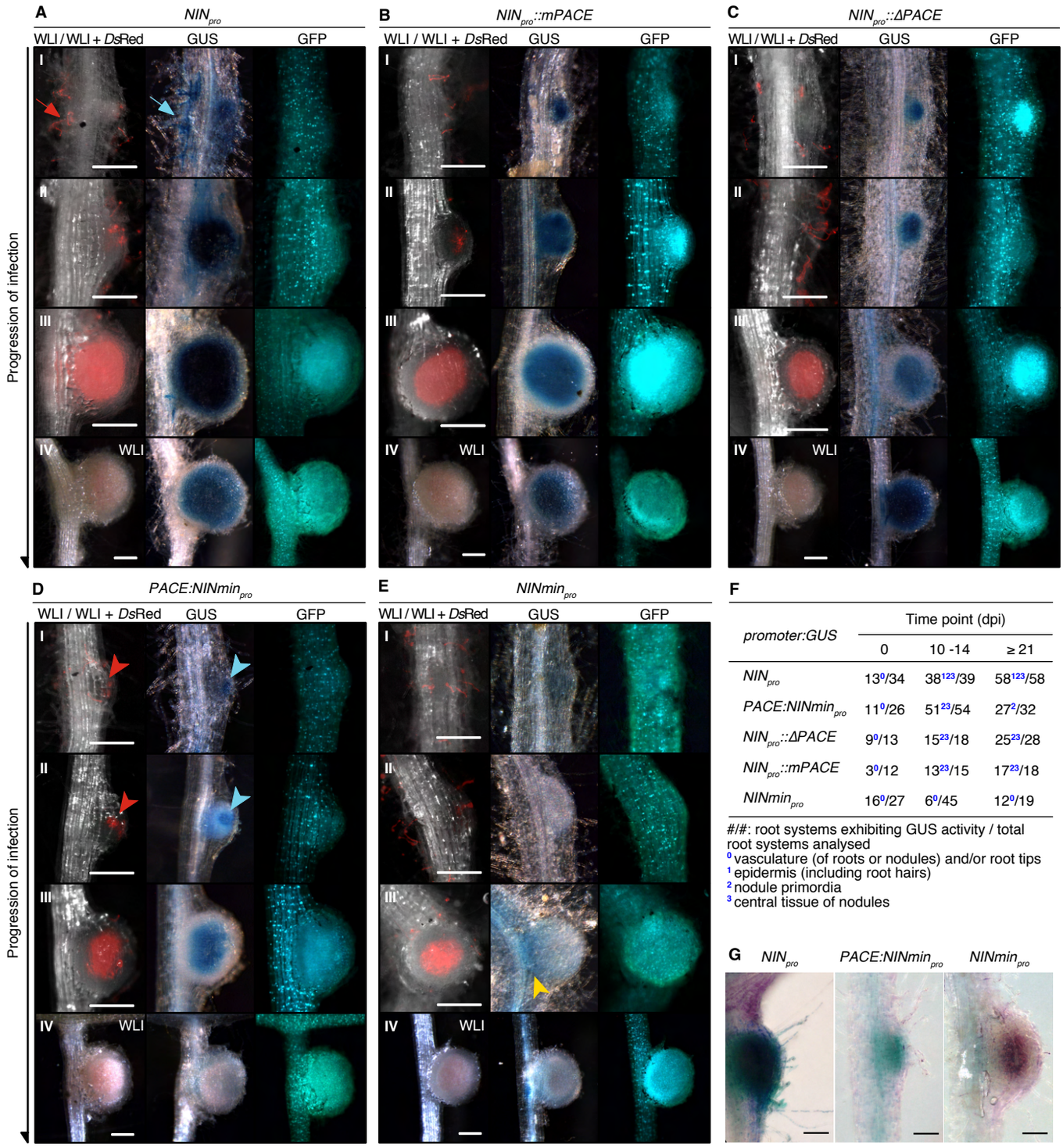


Figure 1 – figure supplement 6 | *PACE* sequence variants from species across the FaFaCuRo clade were able to functionally replace *L. japonicus PACE* in a *LjNIN$_{pro}$:GUS* reporter fusion. *N. benthamiana* leaf cells were transformed with T-DNAs carrying a *GUS* reporter gene driven by either of the indicated promoters: (**A**) the *L. japonicus NIN* promoter (*NIN$_{pro}$*), the *LjNIN* promoter with *PACE* mutated or deleted (*NIN$_{pro}$::mPACE* and *NIN$_{pro}$::ΔPACE*, respectively), or *PACE* sequence variants from the nodulating FaFaCuRo species fused to the *LjNIN* minimal promoter (*NINmin$_{pro}$*); (**B**) chimeric promoters where *LjPACE* in the *LjNIN* promoter was replaced with either one of the *PACE* variants from species tested in (**A**) or from non-nodulating FaFaCuRo species including the *Juglans regia PACE*-like motif (*JrPACE-like*); (**C**) the *S. lycopersicum NIN* promoter (*SlNIN$_{pro}$*), the *SlNIN* promoter with *LjPACE* (*SlNIN$_{pro}$::PACE*) or *mPACE* (*SlNIN$_{pro}$::mPACE*) inserted. For species abbreviations see **Figure 1 – figure supplement 5A**. Note in (**A**) that the deletion or mutation of *PACE* in *LjNIN* promoter resulted in a drastic reduction in reporter gene expression and in (**C**) insertion of *LjPACE* but not *mPACE* into the *S. lycopersicum* promoter confers transactivation by CCaMK$^{1-314}$/Cyclops. The applied statistical method was ANOVA with *post hoc* Tukey: (**A**) $F_{20,144} = 51.38$, $p < 2\text{x}10^{-16}$; (**B**), $F_{18,166} = 149.1$, $p < 2\text{x}10^{-16}$; (**C**) $F_{7,62} = 30.5$, $p = 7.02\text{x}10^{-7}$. Different small letters indicate significant difference. n.d., not determined.



F

promoter:GUS	Time point (dpi)		
	0	10 -14	≥ 21
NIN_{pro}	13^{0}/34	38^{123}/39	58^{123}/58
$PACE:NINmin_{pro}$	11^{0}/26	51^{23}/54	27^{2}/32
$NIN_{pro}::\Delta PACE$	9^{0}/13	15^{23}/18	25^{23}/28
$NIN_{pro}::mPACE$	3^{0}/12	13^{23}/15	17^{23}/18
$NINmin_{pro}$	16^{0}/27	6^{0}/45	12^{0}/19

#/#: root systems exhibiting GUS activity / total root systems analysed
[0] vasculature (of roots or nodules) and/or root tips
[1] epidermis (including root hairs)
[2] nodule primordia
[3] central tissue of nodules

Figure 2 – figure supplement 1 I Spatio-temporal *GUS* expression driven by *PACE* and the *NIN* promoter in *L. japonicus* roots during the bacterial infection process. *L. japonicus* wild-type hairy roots were transformed with T-DNAs carrying a $Ubq10_{pro}$*:NLS-GFP* transformation marker together with a *GUS* reporter gene driven by either of the indicated promoters: (**A**) the 3 kb *LjNIN* promoter (NIN_{pro}); the *LjNIN* promoter with *PACE* (**B**) mutated ($LjNIN_{pro}$*::mPACE*) or (**C**) deleted (NIN_{pro}*::ΔPACE*); (**D**) *PACE* fused to the *LjNIN* minimal promoter ($PACE:NINmin_{pro}$) or (**E**) the *LjNIN* minimal promoter ($NINmin_{pro}$). The progression of bacterial infection was determined by the *Ds*Red signal 10 - 14 days post inoculation (dpi) with *M. loti Ds*Red. Nodules undergoing different stages of infection (panels I to IV) were stained with X-Gluc to reveal the *GUS* expression pattern. Note the overlapping bacterial invasion zone and $PACE:NINmin_{pro}$*:GUS* expression in early infection stages (red and blue arrowheads in (**D**)) as well as the differences between $PACE:NINmin_{pro}$*:GUS* and the much broader NIN_{pro}*:GUS* expression at that stage (red and blue arrows in (**A**)). Red arrow and arrowheads: *M. loti Ds*Red. Blue arrow and arrowheads: GUS activity in root hairs bearing ITs and nodule primordia, respectively. The $NINmin_{pro}$*:GUS* fusion gave only rarely detectable signal, and if so in the vasculature (yellow arrowhead in (**E**)). Only pictures taken under white light illumination (WLI) are displayed for nodules in panel VI to reveal the pink colour of leghemoglobin, characteristic for mature and fully infected nodules. Note that $PACE:NINmin_{pro}$*:GUS* expression was absent at this stage, whereas the NIN_{pro}*:GUS* resulted in strong blue staining in the nodule regardless of the presence of *PACE* (compare panel IV in (**D**) and (**A - C**)). (**F**) Quantification of transgenic root systems exhibiting *GUS* expression in different cell types and tissues exemplarily displayed in (**A - E**). (**G**) *PACE* drove *GUS* reporter gene expression in the central tissue of primordia and nodules, but was not sufficient for expression in root hairs. Transgenic roots carrying promoter:GUS fusions same as in (**A, D and E**) were inoculated with *M. loti lacZ* and dual-stained with X-Gluc and Magenta-Gal. Purple: *M. loti lacZ*. Blue: GUS activity. Note the co-existence of blue and purple staining in root hairs on roots transformed by NIN_{pro}*:GUS*, but not that transformed by $PACE:NINmin_{pro}$*:GUS*. Bars, 250 µm.

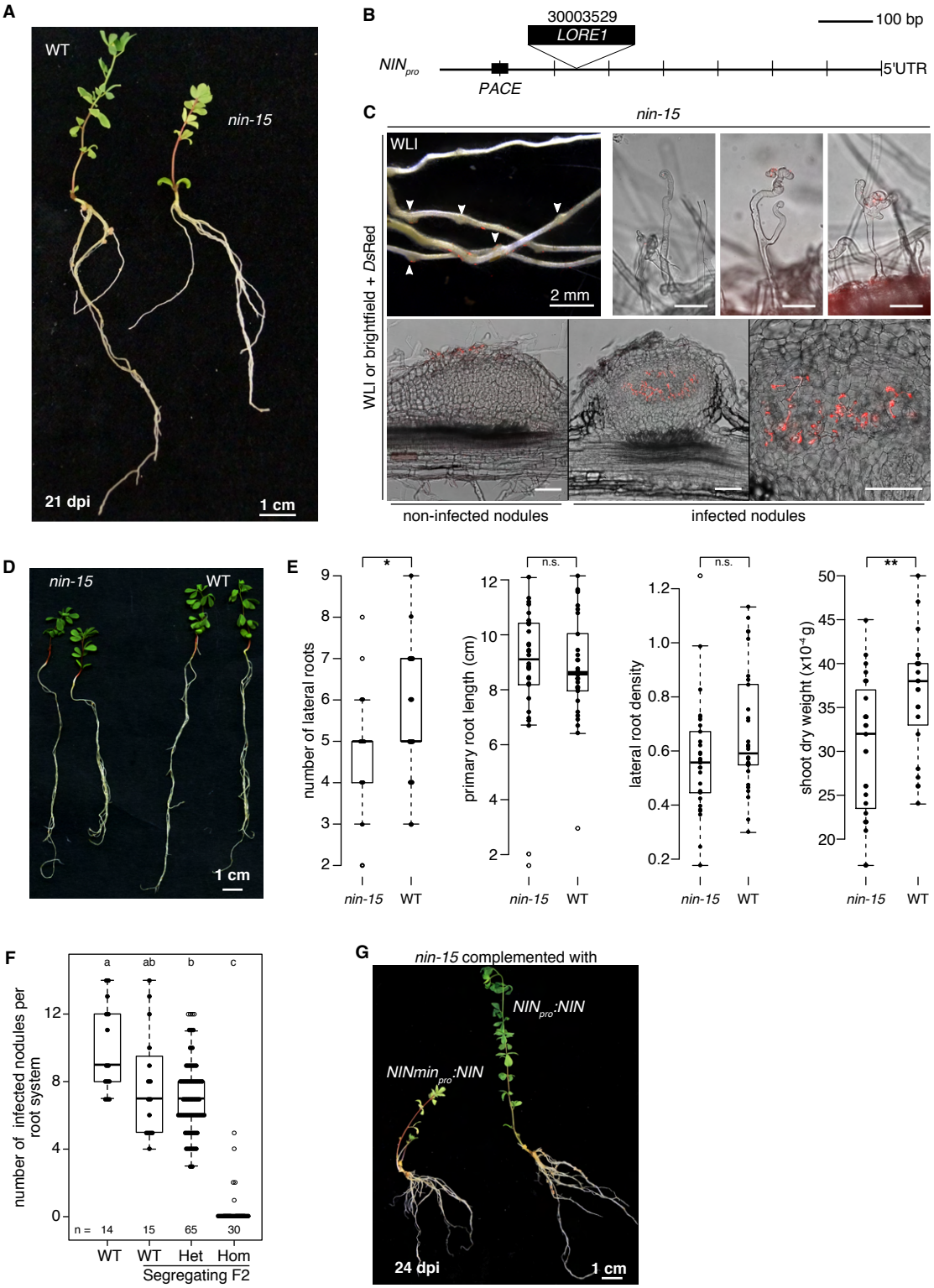


Figure 3 – figure supplement 1 I *L. japonicus nin-15* mutant phenotype. (**A**) A representative picture of *L. japonicus* wild-type (WT, left) and *nin-15* (right) plants 21 dpi with *M. loti Ds*Red. (**B**) Position of the *Lotus Retrotransposon 1* (*LORE1*) insertion within the *NIN* promoter in the *nin-15* mutant. (**C**) Representative pictures of *nin-15* root hairs and nodule sections 21 dpi with *M. loti Ds*Red. Forty-nine plants with a total number of 436 nodules were analysed: only four plants bore one or two IT(s) within root hairs and seven plants bore one or two infected nodule(s). Deformed or curled root hairs in the presence of *M. loti Ds*Red were abundant but infection threads were rarely found. Arrowheads: uninfected nodules. Unlabelled bars, 100 µm. (**D - E**) Phenotype of *nin-15* in the presence of a symbiosis-independent nitrogen source (15 mM KNO_3) for 28 days. (**D**) Pictures documenting the healthy status of *L. japonicus* WT and *nin-15* plants (compare (**D**) and (**A**)) and (**E**) quantitative assessment of parameters displayed in boxplots. Thirty plants per genotype were analysed. Each dot represents one plant. Lateral root density: number of lateral roots/primary root length (cm). The applied statistical method was pairwise *t*-test: $^{*}p < 0.05$; $^{**}p < 0.01$; n.s.: not significant. (**F**) Segregation analysis of *nin-15.* The applied statistical method was ANOVA with *post hoc* Tukey: $F_{3,120} = 84.1$, $p = 2\text{x}10^{-16}$. Different small letters indicate significant difference. (**G**) Representative pictures of *nin-15* plants with hairy roots transformed with the *NIN* gene driven by the *L. japonicus NIN* minimal promoter ($NINmin_{pro}$) or the 3 kb *NIN* promoter (NIN_{pro}) 24 dpi with *M. loti Ds*Red. WLI: white light illumination.

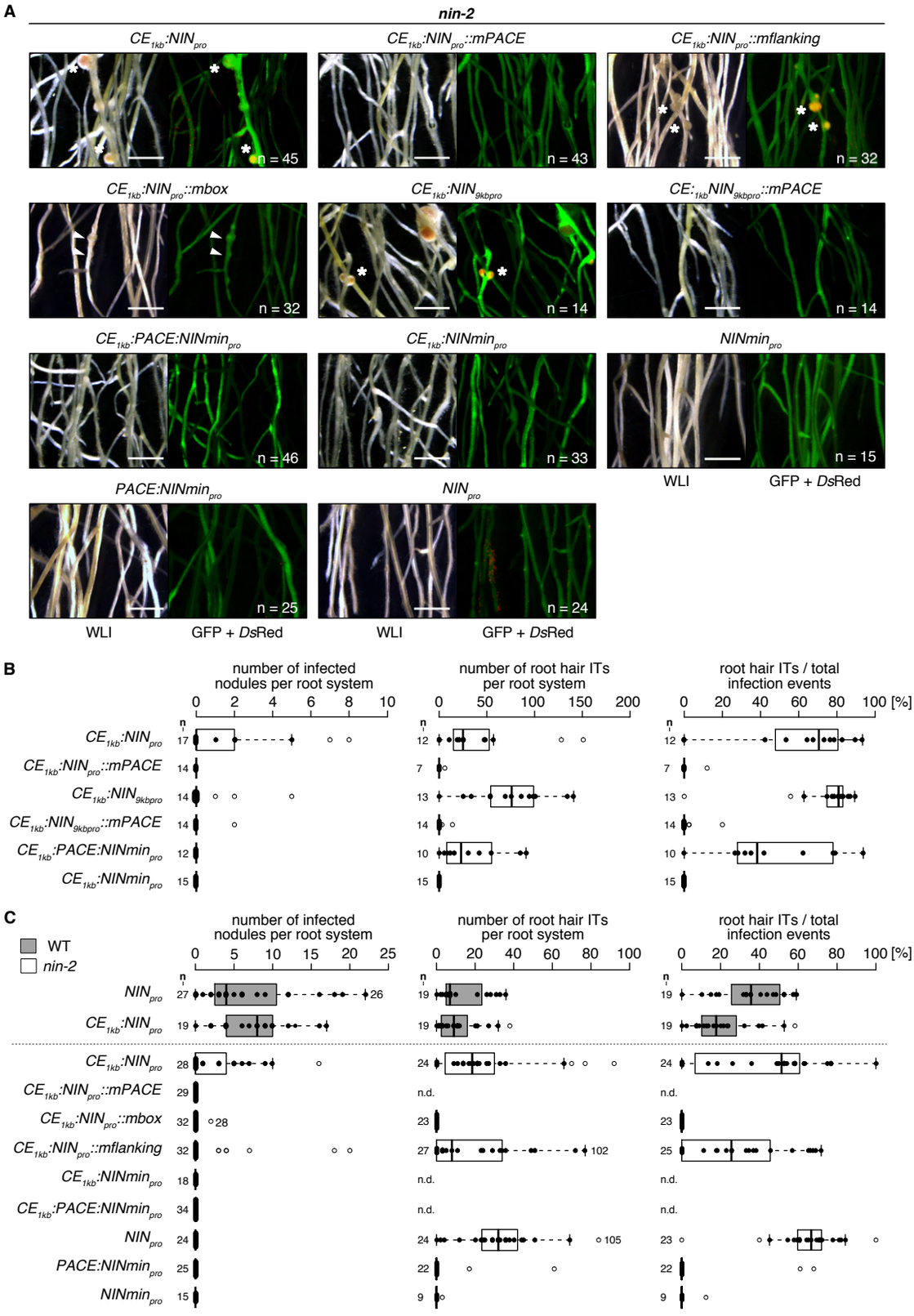


Figure 3 – figure supplement 2 | The *CYC-box* and flanking sequences of *PACE* are required for the complete restoration of the bacterial infection process in the *L. japonicus nin-2* mutant. *nin-2* roots were transformed with T-DNAs carrying a $Ubq10_{pro}$:*NLS-GFP* transformation marker in tandem with the *LjNIN* gene driven by either of the following promoter versions: the cytokinin element-containing region of 1 kb (CE_{1kb}) fused to the 3 kb or 9 kb *LjNIN* promoter (CE_{1kb}:NIN_{pro} or CE_{1kb}:NIN_{9kbpro}, respectively); CE_{1kb}:NIN_{pro} or CE_{1kb}:NIN_{9kbpro} with *PACE* mutated (CE_{1kb}:NIN_{pro}::*mPACE* or CE_{1kb}:NIN_{9kbpro}::*mPACE*, respectively); CE_{1kb}:NIN_{pro} carrying a mutated Cyclops binding site (*CYC-box*) (CE_{1kb}:NIN_{pro}::*mbox*); CE_{1kb}:NIN_{pro} carrying mutated sequences flanking the *CYC-box* in *PACE* (CE_{1kb}:NIN_{pro}::*mflanking*); CE_{1kb} fused to the *LjNIN* minimal promoter (CE_{1kb}:$NINmin_{pro}$); CE_{1kb} fused to *PACE* and to $NINmin_{pro}$ (CE_{1kb}:*PACE*:$NINmin_{pro}$); NIN_{pro}, *PACE*:$NINmin_{pro}$ or $NINmin_{pro}$. (**A**) Representative overview pictures of transgenic root systems. Roots were analysed 21 dpi with *M. loti Ds*Red. White asterisks and arrowheads: infected and non-infected nodules, respectively. Bars, 2 mm. (**B - C**) Boxplots displaying the number of root hair ITs or infected nodules and the percentage of root hair ITs among total infection events (sum of bacterial entrapments and ITs). Each dot represents one transgenic *nin-2* root system or root piece. *L. japonicus* WT roots transformed with NIN_{pro}:*NIN* or CE_{1kb}:NIN_{pro}:*NIN* were included as controls. Note the loss of restoration of nodules and IT formation associated with the mutation of *PACE* or only the *CYC-box* in *PACE*; and the reduction of same when sequences flanking the *CYC-box* in *PACE* were mutated. n: number of transgenic root systems or root pieces analysed. Numbers above the boxplots: the value of individual data points outside of the plotting area. n.d.: not determined. WLI: white light illumination.

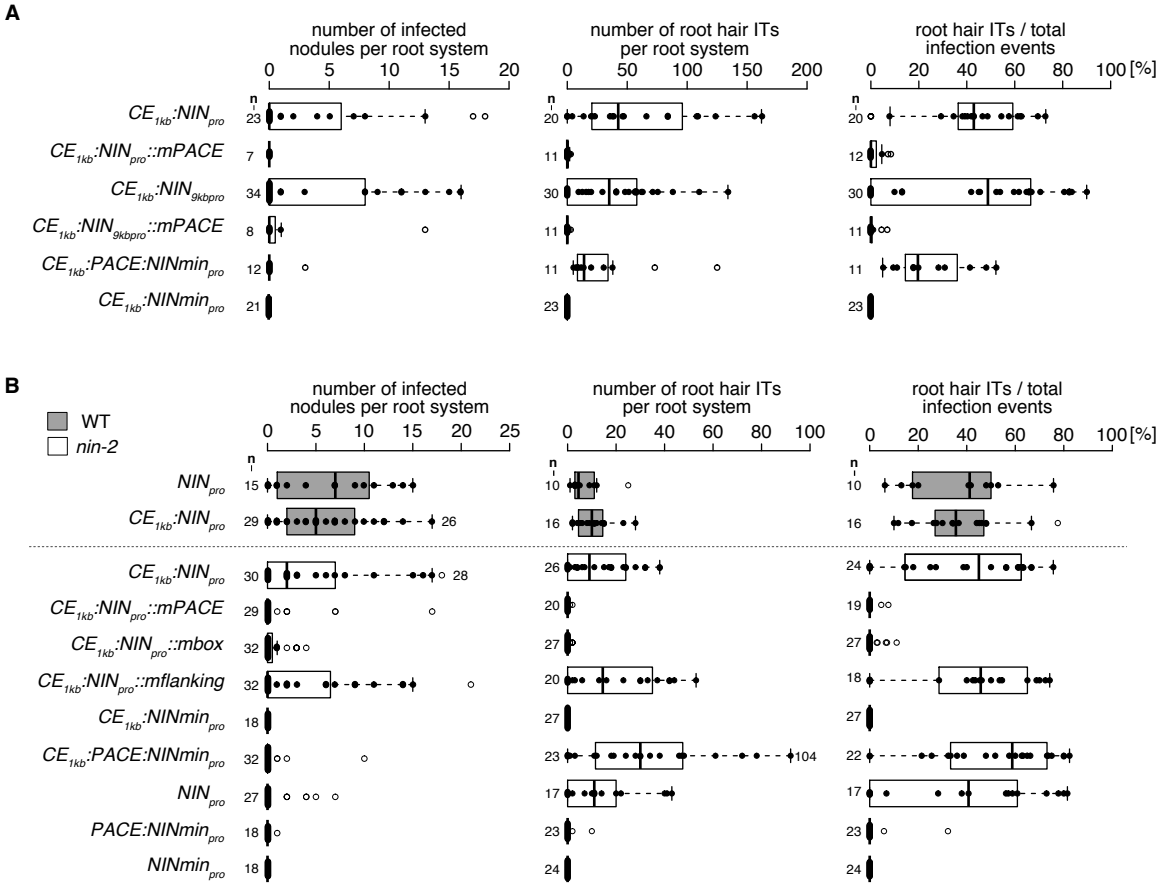


Figure 3 – figure supplement 3 | The *CYC-box* and flanking sequences of *PACE* are required for the complete restoration of the bacterial infection process in the *L. japonicus nin-2* mutant. Roots were from a subset of plants from the same experiment depicted in **Figure 3 – figure supplement 2** but analysed 35 dpi with *M. loti Ds*Red. (**A - B**) Boxplots displaying the number of root hair ITs or infected nodules and the percentage of root hair ITs among total infection events (sum of bacterial entrapments and ITs). Each dot represents one transgenic *nin-2* root system or root piece. *L. japonicus* WT roots transformed with NIN_{pro}:*NIN* or CE_{1kb}:NIN_{pro}:*NIN* were included as controls. Note that the results follow the same trend as those obtained 21 dpi with *M. loti Ds*Red (**Figure 3 – figure supplement 2**). n: number of transgenic root systems or root pieces analysed. Numbers above the boxplots: the value of individual data points outside of the plotting area.

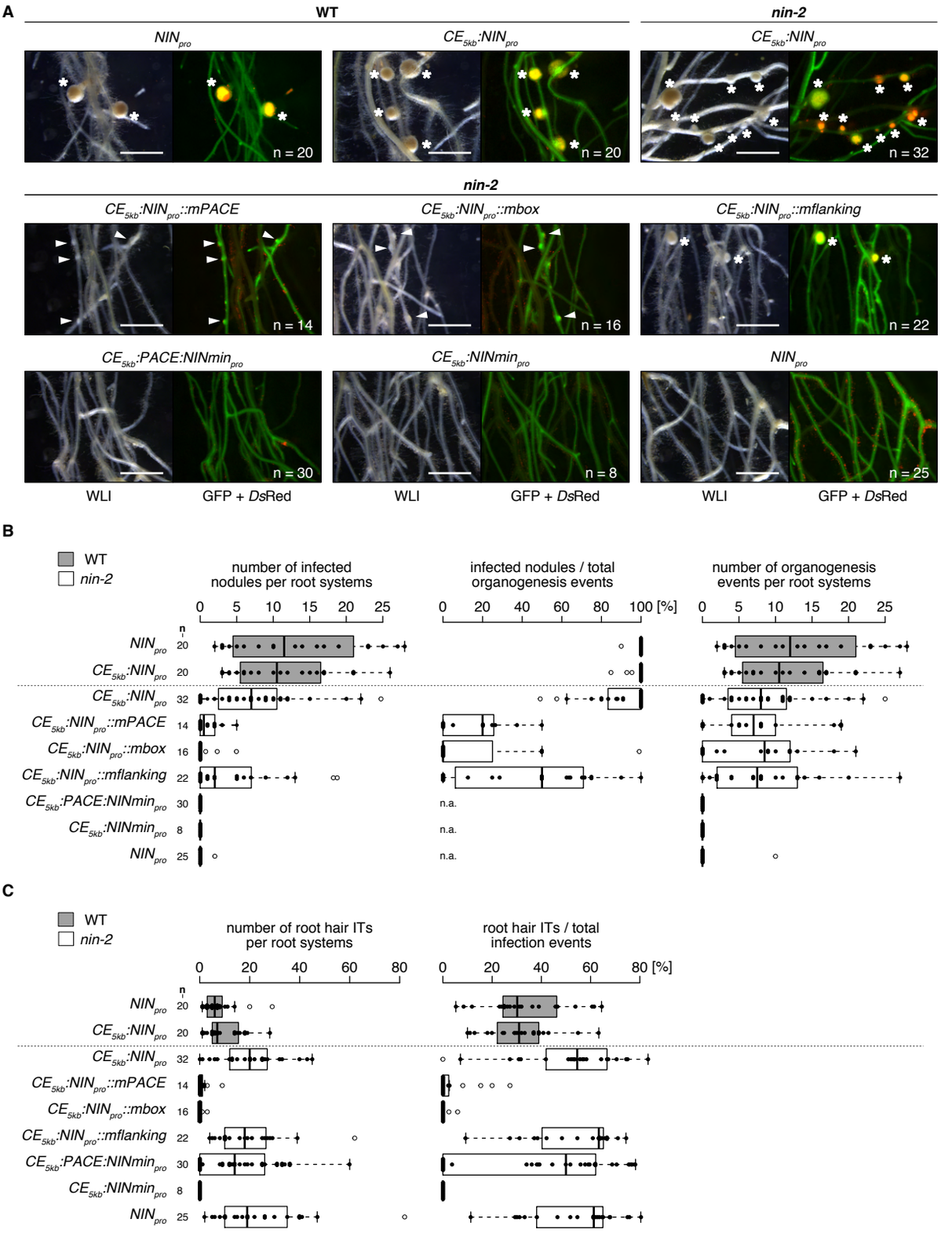


Figure 3 – figure supplement 4 | The *CYC-box* and flanking sequences of *PACE* are required for the complete restoration of the bacterial infection process but are dispensable for the nodule organogenesis process in the *L. japonicus nin-2* mutant. *nin-2* roots were transformed with T-DNAs carrying a $Ubq10_{pro}$:*NLS-GFP* transformation marker in tandem with the *LjNIN* gene driven by either of the following promoter versions: the cytokinin element-containing region of 5 kb (CE_{5kb}) fused to the 3 kb *LjNIN* promoter (CE_{5kb}:NIN_{pro}); CE_{5kb}:NIN_{pro} with *PACE* mutated (CE_{5kb}:NIN_{pro}::*mPACE*); CE_{5kb}:NIN_{pro} carrying a mutated Cyclops binding site (*CYC-box*) (CE_{5kb}:NIN_{pro}::*mbox*); CE_{5kb}:NIN_{pro} carrying mutated sequences flanking the *CYC-box* in *PACE* (CE_{5kb}:NIN_{pro}::*mflanking*); CE_{5kb} fused to the *LjNIN* minimal promoter (CE_{5kb}:$NINmin_{pro}$); CE_{5kb} fused to *PACE* and to $NINmin_{pro}$ (CE_{5kb}:*PACE*:$NINmin_{pro}$) or NIN_{pro}. (**A**) Representative overview pictures of transgenic root systems. Roots were analysed 21 dpi with *M. loti Ds*Red. White asterisks and arrowheads: infected and non-infected nodules, respectively. Bars, 2 mm. (**B**) Boxplots displaying the number of infected nodules, the percentage of infected nodules among total organogenesis events (sum of infected and non-infected nodules) and the number of organogenesis events. (**C**) Boxplots displaying the number of root hair ITs and the percentage of root hair ITs among total infection events (sum of bacterial entrapments and ITs). Each dot represents one transgenic *nin-2* root system or root piece. *L. japonicus* WT roots transformed with NIN_{pro}:*NIN* or CE_{5kb}:NIN_{pro}:*NIN* were included as controls. Note that the mutation of *PACE* or only the *CYC-box* in *PACE* led to an almost complete loss of IT formation and infected nodules per root system while nodule organogenesis was not significantly reduced; and that mutation of sequences flanking the *CYC-box* in *PACE* led to a reduction of the number of infected nodules per root systems. n: number of transgenic root systems or root pieces analysed. Numbers above the boxplots: the value of individual data points outside of the plotting area. n.a.: not applicable. WLI: white light illumination.

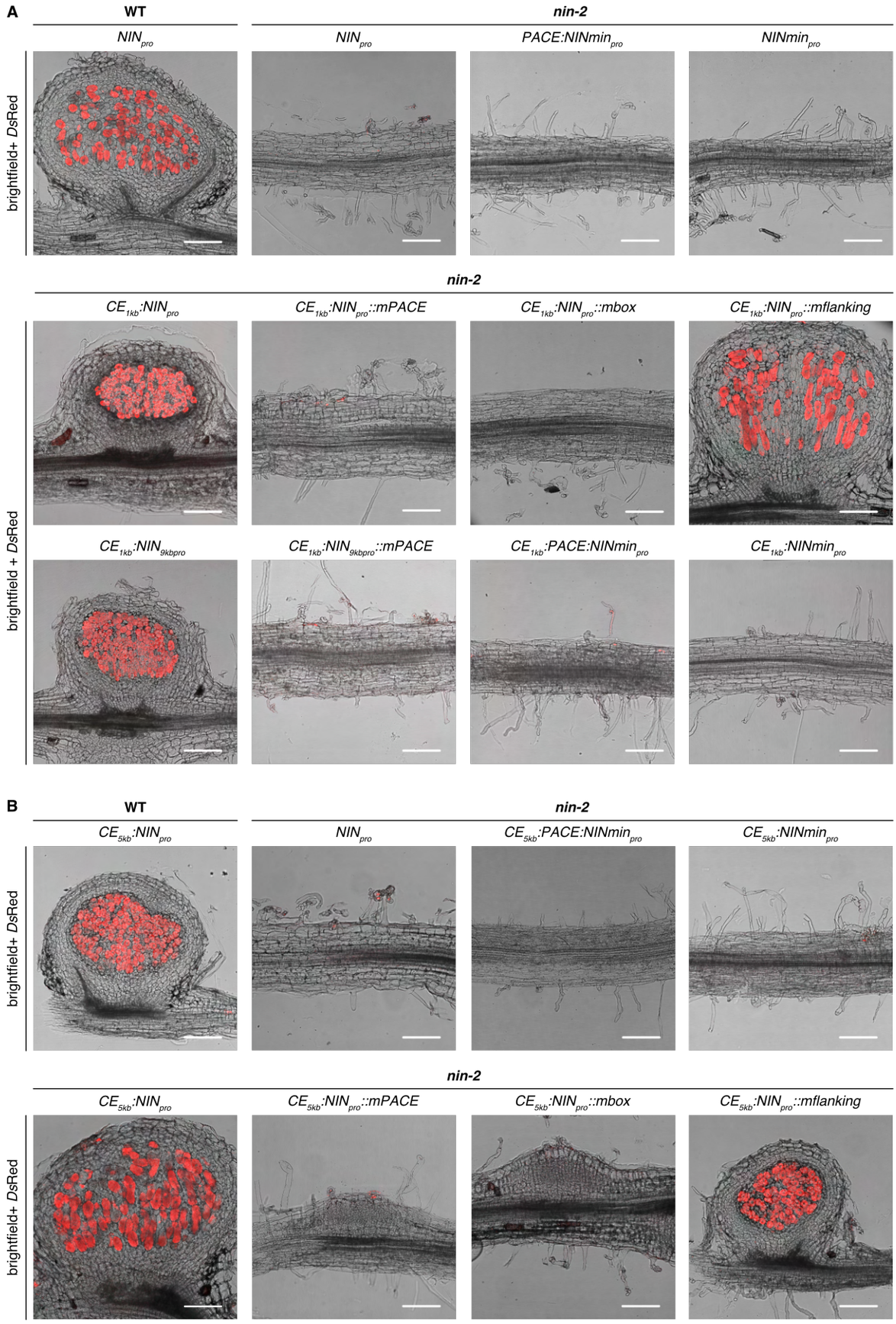


Figure 3 – figure supplement 5 | The *CYC-box* and flanking sequences of *PACE* are required for the complete restoration of the bacterial infection process but are dispensable for the nodule organogenesis process in the *L. japonicus nin-2* mutant. Pictures of nodule sections or roots from *L. japonicus nin-2* roots 21 dpi with *M. loti Ds*Red from the same experiments depicted in **Figure 3 – figure supplement 2** (**A**) and **Figure 3 – figure supplement 4** (**B**). Nodule sections from *L. japonicus* WT roots transformed with NIN_{pro}:*NIN* and $CE_{5kb}NIN_{pro}$:*NIN* were included for comparison. Note that when the cytokinin element-containing region of 1 kb was fused to NIN_{pro} nodule organogenesis was abolished by mutation of *PACE* or only the *CYC-box* in *PACE* and that these mutations did not abolish organogenesis when the cytokinin element-containing region of 5 kb was fused to NIN_{pro}. Bars, 100 µm.

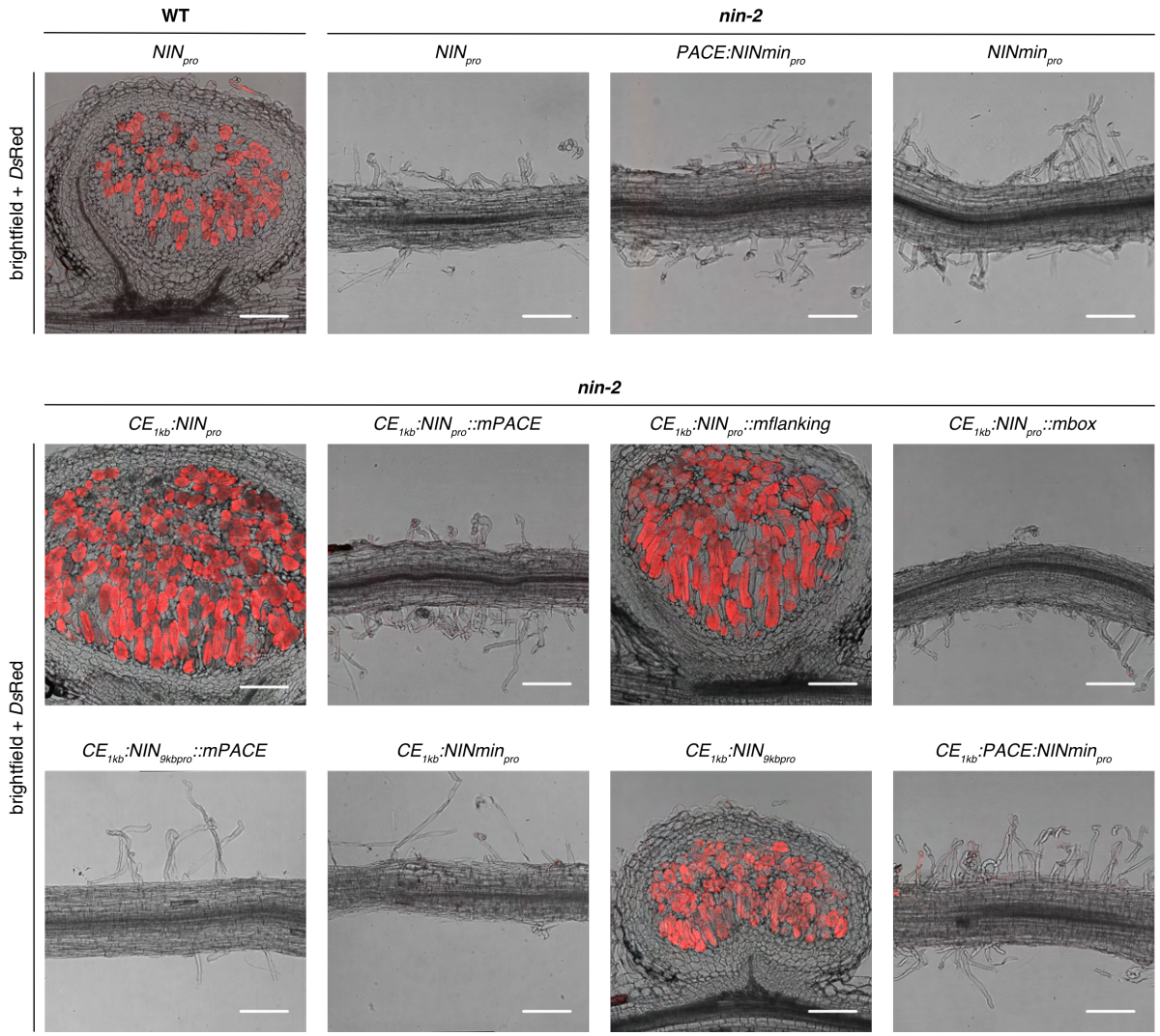


Figure 3 – figure supplement 6 | The *CYC-box* and flanking sequences of *PACE* are required for the complete restoration of the bacterial infection process in the *L. japonicus nin-2* mutant. Pictures of nodule sections or roots from *L. japonicus nin-2* roots 35 dpi with *M. loti Ds*Red from the same experiments depicted in **Figure 3 – figure supplement 3**. Upper left corner: a nodule section from a *L. japonicus* WT root transformed with NIN_{pro}:*NIN* was included for comparison. Bars, 100 µm.

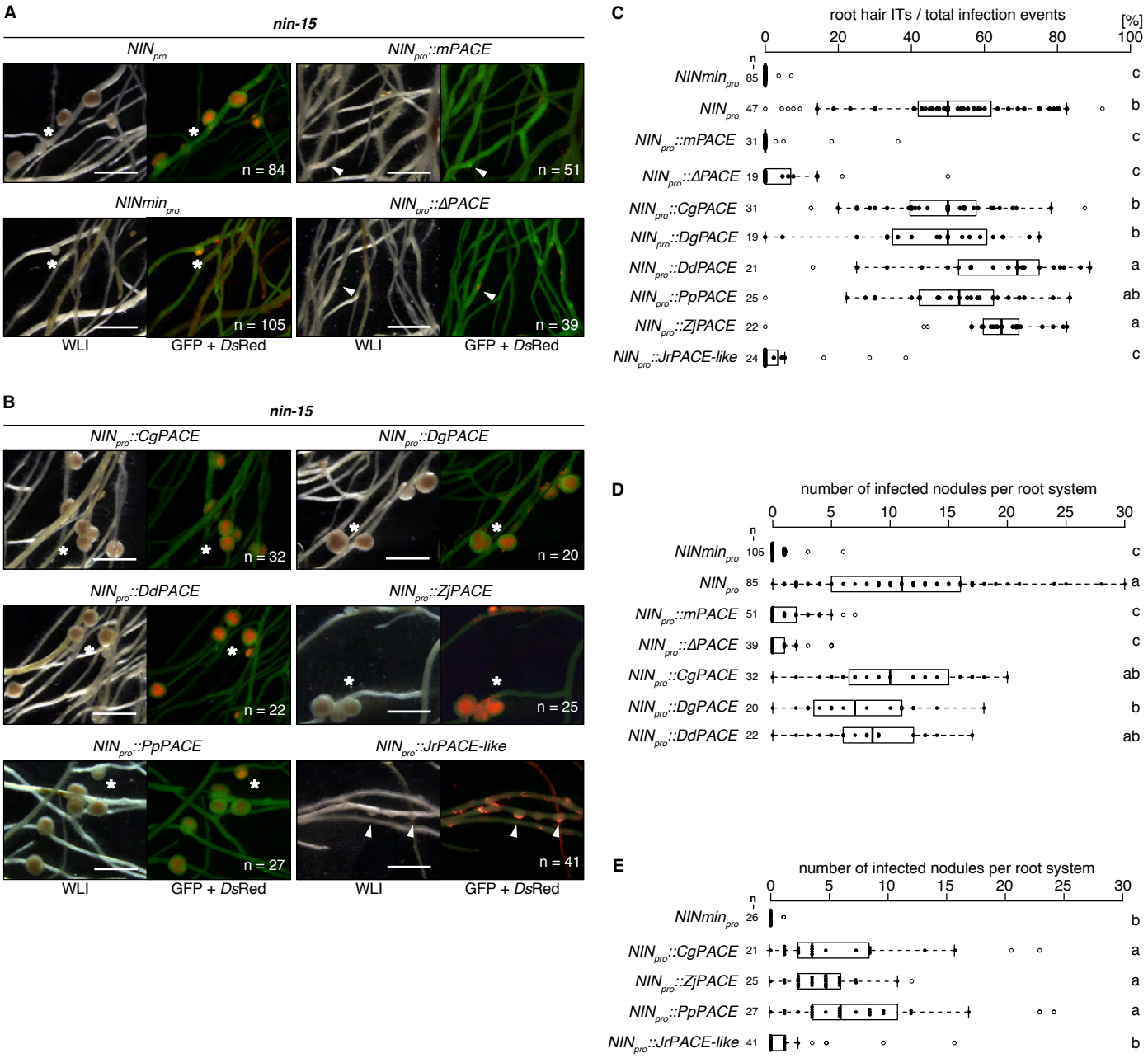
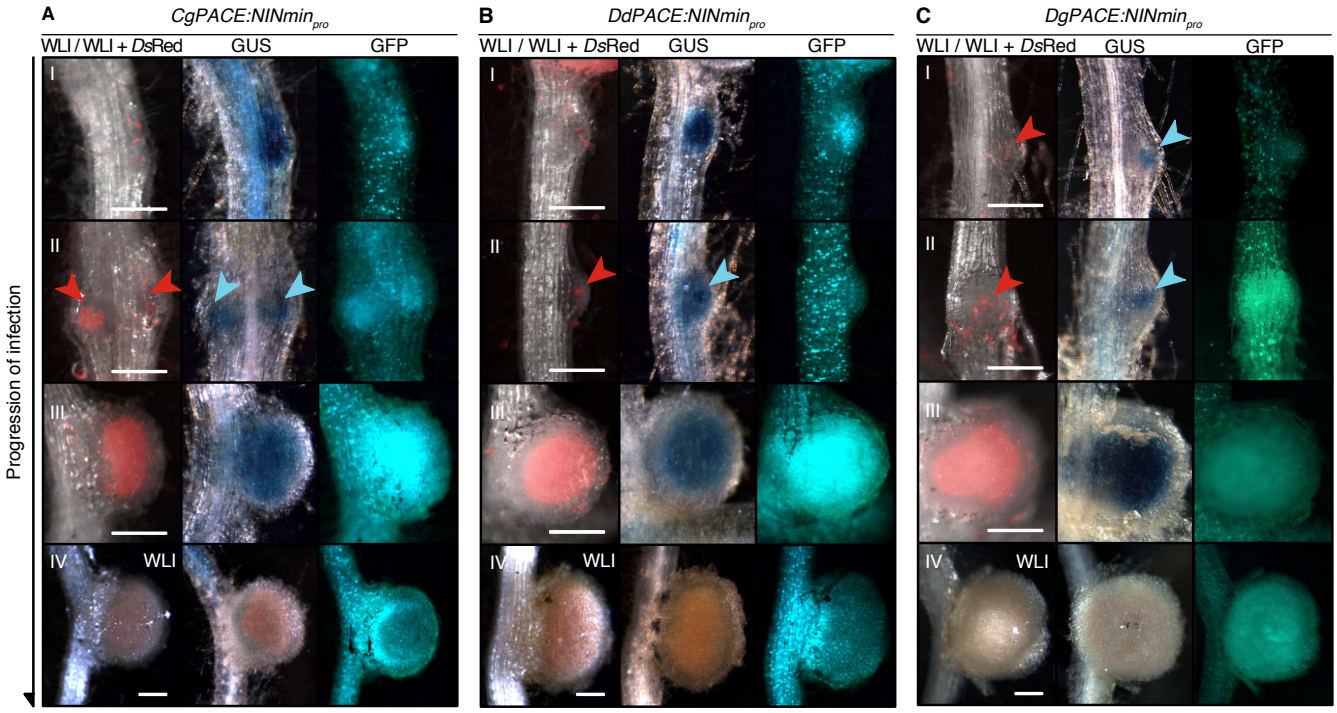


Figure 3 – figure supplement 7 | *PACEs* from FaFaCuRo species are functionally equivalent in restoring bacterial infection in the *L. japonicus nin-15* mutant. *L. japonicus nin-15* roots were transformed with T-DNAs carrying a *Ubq10*$_{pro}$*:NLS-GFP* transformation marker in tandem with the *LjNIN* gene driven by either of the following promoters: (**A**) the 3 kb *LjNIN* promoter (*NIN*$_{pro}$), the *LjNIN* minimal promoter (*NINmin*$_{pro}$), the 3 kb *LjNIN* promoter with *PACE* deleted (*NIN*$_{pro}$*::ΔPACE*) or mutated (*NIN*$_{pro}$*::mPACE*); (**B**) the 3 kb *LjNIN* promoter with *LjPACE* replaced with either of the *PACE* sequence variants from nodulating or non-nodulating FaFaCuRo species and analysed 21 dpi with *M. loti Ds*Red. (**A - B**) Representative overview pictures of *nin-15* transgenic roots systems. Sections of representative nodules are displayed in **Figure 3**. Note the drastic reduction of restoration of infection in nodules and root hairs associated with the mutation or deletion of *PACE* as well as the replacement of *PACE* with *JrPACE-like* in the context of the *LjNIN* promoter. White asterisks and arrowheads: infected and non-infected nodules, respectively. (**C - E**) Boxplots displaying **c,** the percentage of root hair ITs among total infection events (sum of bacterial entrapments and ITs) and (**D - E**) the number of infected nodules from two independent experiments. Each dot in (**D - E**) represents one *nin-15* transgenic root system. (**C**) displays merged data from all experiments as the percentage represents a normalised value calculated for each root piece (see **Figure 3 – Source Data 1**). n: number of transgenic root systems or root pieces analysed. For species abbreviations see **Figure 1 – figure supplement 5A**. The applied statistical method was ANOVA with *post hoc* Tukey: (**C**) $F_{9,313} = 106.7$, $p < 2\times10^{-16}$; (**D**) $F_{6,346} = 82.89$, $p < 2\times10^{-16}$; (**E**) $F_{4,135} = 20.18$, $p = 4.76\times10^{-13}$. Different small letters indicate significant differences. Bars, 2 mm. WLI: white light illumination.



D

promoter:GUS	Time point (dpi)		
	0	10 -14	≥ 21
CgPACE:NINmin$_{pro}$	7[0]/7	14[23]/19	15[2]/20
DdPACE:NINmin$_{pro}$	8[0]/8	15[23]/18	17[2]/22
DgPACE:NINmin$_{pro}$	n.d.	35[23]/46	28[2]/38

#/#: root systems exhibiting GUS activity / total root systems analysed
[0] vasculature (of roots or nodules) and/or root tips
[1] epidermis (including root hairs)
[2] nodule primordia
[3] central tissue of nodules

Figure 3 – figure supplement 8 I Spatio-temporal *GUS* expression driven by *PACE* variants in *L. japonicus* roots during the bacterial infection process. *L. japonicus* WT roots were transformed with T-DNAs carrying a *Ubq10*$_{pro}$*:NLS-GFP* transformation marker together with a *GUS* reporter gene driven by either of the *PACE* variants from nodulating FaFaCuRo species fused to the *LjNIN* minimal promoter (*NINmin*$_{pro}$). For species abbreviations and experimental details see **Figure 1 – figure supplement 5A and Figure 2 – figure supplement 1**, respectively. Note the overlapping bacterial invasion zone and *PACE:NINmin*$_{pro}$*:GUS* expression in early infection stages (red and blue arrowheads in (**A - C**)). Red arrowheads: *M. loti Ds*Red. Blue arrowheads: GUS activity in nodule primordia. Only pictures taken under white light illumination (WLI) are displayed for nodules in panel VI to reveal the pink colour of leghemoglobin, characteristic for mature and fully infected nodules. Note that like *LjPACE*, the *PACE* variants-driven *GUS* expressions were absent at this stage (panel IV in (**A - C**) and panel IV in **Figure 2 – figure supplement 1D**). (**D**) Quantification of transgenic root systems exhibiting *GUS* expression in different cell types and tissues exemplarily displayed in (**A - C**). n.d.: not determined. Bars, 250 µm.

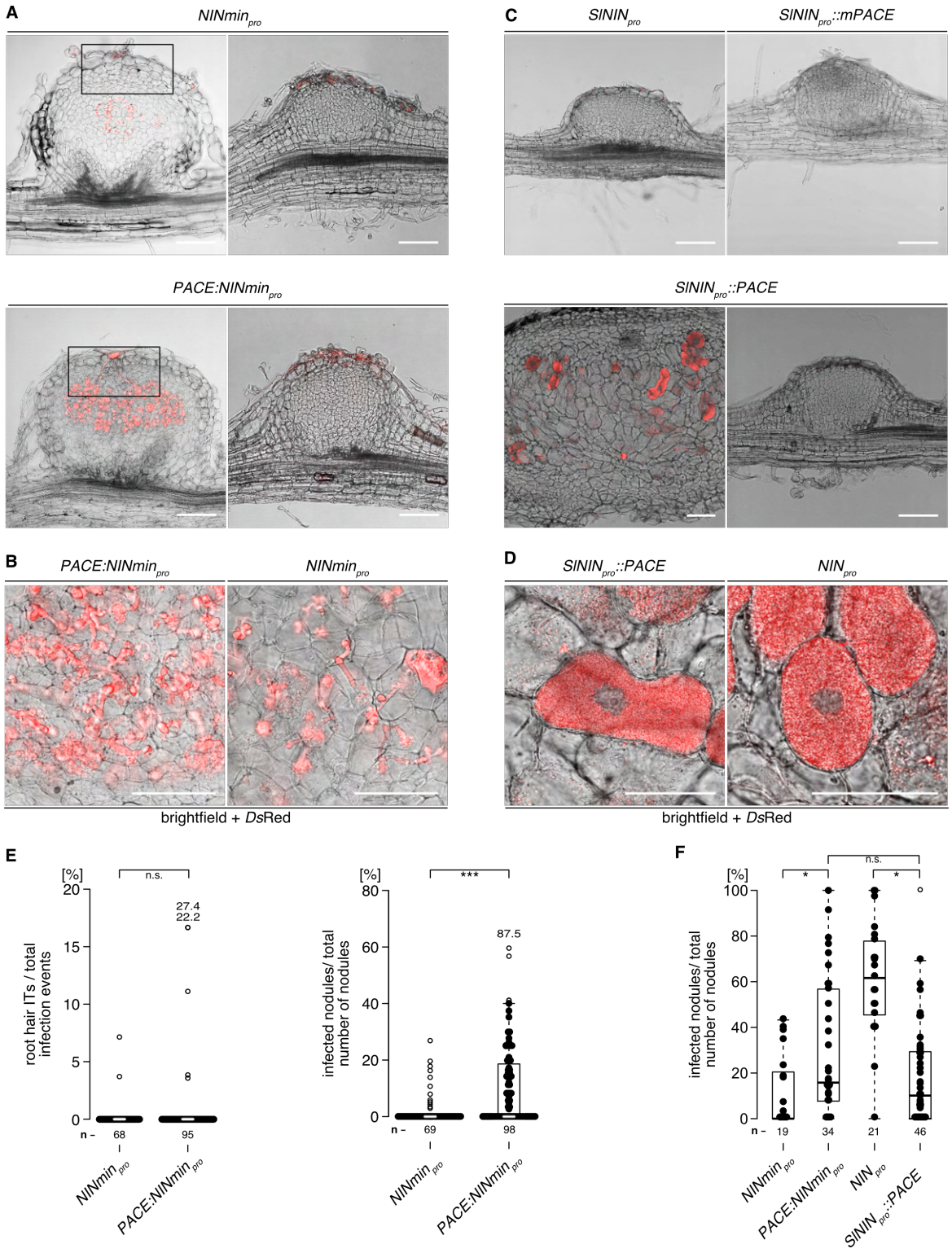


Figure 4 – figure supplement 1 | *PACE* alone or in the context of the *S. lycopersicum NIN* promoter (a species outside of the FaFaCuRo clade) enables IT formation in the cortex. (**A - D**) Representative pictures of sections of nodules formed on *L. japonicus nin-15* roots transformed with T-DNAs carrying a *Ubq10*$_{pro}$*:NLS-GFP* transformation marker together with the *LjNIN* gene driven by either of the following promoters: (**A - B**) the *L. japonicus NIN* minimal promoter (*NINmin*$_{pro}$) or *PACE* fused to *NINmin*$_{pro}$ (*PACE:NINmin*$_{pro}$); (**C - D**) the 3 kb *S. lycopersicum NIN* promoter (*SlNIN*$_{pro}$), the 3 kb *SlNIN* promoter with mutated *PACE* (*SlNIN*$_{pro}$*::mPACE*) or with *L. japonicus PACE* inserted (*SlNIN*$_{pro}$*::PACE*), 21 dpi with *M. loti Ds*Red (from the same experiments depicted in **Figure 4**). Black rectangles in (**A**) demarcate the enlarged area displayed in **Figure 4A and 4B** to focus on the initial infection structures. Note the absence of cells filled with symbiosomes in nodules transformed with the *LjNIN* gene driven by *PACE:NINmin*$_{pro}$ or *NINmin*$_{pro}$. By contrast, infected cells were often filled with symbiosomes in the *SlNIN*$_{pro}$*::PACE:NIN*-transformed nodules, like those resulted by *NIN*$_{pro}$*:NIN* (see (**C**) and compare the two sections in (**D**)). (**E - F**) Boxplots displaying the percentage of root hair ITs among total infection events (sum of bacterial entrapments and ITs) or the percentage of infected nodules among total number of nodules (**E**) 21 dpi and (**F**) 35 dpi with *M. loti Ds*Red, respectively. Each dot represents one *nin-15* transgenic root system or root piece. (**E**) displays results from an independent repetition from the experiment depicted in **Figure 4**. n: number of transgenic root systems or root pieces analysed. Numbers above the boxplots: the value of individual data points outside of the plotting area. The applied statistical method was Fisher's exact test: *$p < 0.05$; ***$p < 0.001$; n.s.: not significant. Bars, (**A and C**) 100 µm; (**B and D**) 50 µm.