A shade-responsive microProtein in the Arabidopsis *ATHB2* gene regulates elongation growth and root development

Ashleigh Edwards[1,2#], Maurizio Junior Chiurazzi[1,2#], Anko Blaakmeer[1,3#], Ylenia Vittozzi[1,2], Ashish Sharma[1,4], Sanne Matton[1,5], Valdeko Kruusvee[1,2], Daniel Straub[1,6], Giovanna Sessa[7], Monica Carabelli[7], Giorgio Morelli[8], Stephan Wenkel[1,2]*

[1]University of Copenhagen, Department of Plant and Environmental Sciences, Thorvaldsensvej 40, 1871 Frederiksberg, Denmark

[2]Department of Plant Physiology, Umeå Plant Science Centre, Umeå University, SE-901 87, Umeå, Sweden

[3]current address: Carlsberg Research Laboratory, J.C. Jacobsens Gade 4, Copenhagen V DK-1799, Denmark

[4]current address: CSIR, Central Institute of Medicinal and Aromatic Plants, Lucknow, India

[5]current address: Department of Plant Science, Wagening University, The Netherlands

[6]current address: Quantitative Biology Center (QBiC), University of Tübingen, Auf der Morgenstelle, Tübingen, Germany

[7]Institute of Molecular Biology and Pathology, National Research Council, P.le A. Moro 5, 00185 Rome, Italy

[8]Research Centre for Genomics and Bioinformatics, Council for Agricultural Research and Economics (CREA), Via Ardeatina 546, 00178 Rome, Italy

[#]these authors contributed equally

*corresponding author: stephan.wenkel@umu.se

Abstract

The ability of plants to thrive under suboptimal light conditions, such as shade, is crucial for their overall survival and reproductive success. Here, we show that Arabidopsis seedlings produce a large number of alternative transcripts when exposed to shade. Notably, one of the identified transcript candidates, which was upregulated in shade conditions, was found to be an alternative transcript of the *ATHB2* gene. ATHB2 belongs to the HD-ZIPII class of transcription factors and is a well-established regulator of the shade avoidance response. The function of the alternative transcript and the small leucine zipper protein encoded by it, ATHB2miP, was investigated. We found that ATHB2miP is primarily expressed in the shoot meristem and interacts with full-length ATHB2 protein to inhibit its activity through a negative feedback mechanism. Deletion of the genomic region encoding the leucine zipper domain of the *ATHB2* gene using CRISPR, resulted in plants exhibiting altered shade avoidance responses and root development. We show that the leucine zipper domain is required for dimerising and localising to nuclear photobodies. There is a significant overlap in deregulated genes between plants ectopically expressing ATHB2miP and *athb2* mutant plants. The analysis of gene ontology and clustering revealed that the most affected processes are auxin synthesis and signaling, root development, and iron homeostasis. Shade growth experiments at different iron concentrations revealed a role for *ATHB2* in regulating iron uptake and showed that iron availability affects shade growth in an *ATHB2*-dependent manner. This study identifies ATHB2miP as a novel regulator of shade avoidance responses in Arabidopsis, highlighting the intricate transcriptional regulation underlying these processes.

Introduction

Since the proposition of the "one gene - one enzyme hypothesis" by George W. Beadle and Edward L. Tatum in 1941 (Beadle & Tatum, 1941), it has become increasingly evident that the complexity of the proteome is more intricate than initially anticipated. It is firmly established that alternative splicing can give rise to multiple mRNA isoforms (Berget *et al*, 1977), and in certain instances, these isoforms can be translated into different protein variants. Alongside alternative splicing events, such as exon-skipping, which can potentially result in protein isoforms with distinct structures, the alternative use of transcription start sites can also lead to protein variants lacking specific domains that exhibit contrasting functions compared to the full-length variant (Svoboda *et al*, 1988).

Recent endeavors in uncovering the small peptidome, employing riboproteomics, a combination of Ribo-seq and proteomics, have unveiled thousands of novel protein-coding genes in humans (Mudge *et al*, 2022). Some of these short open reading frames (sORFs) originate from their corresponding loci in an overlapping manner and are frequently encoded

in an alternate reading frame. Certain upstream sORFs can possess a regulatory role in controlling the larger downstream ORF at the transcriptional or translational level (Luo *et al*, 1995; Orbach *et al*, 1990). Furthermore, long non-coding RNAs (lncRNAs), which arise from pervasive transcription of the seemingly non-coding regions of the genome, represent another abundant source of small and potentially emerging protein-coding genes. It has been suggested that translated lncRNAs may even serve as a source for *de novo* gene birth (Ruiz-Orera *et al*, 2020).

Small proteins are abundant in both prokaryotes and eukaryotes, and many of them exhibit dynamic expression patterns. Notably, sORFs have been demonstrated to possess significant regulatory roles in plant development (Hanada *et al*, 2013) as well as in animals (Kondo *et al*, 2010; Zhang *et al*, 2020). In the field of human biology, certain sORFs have been identified to exhibit disease-specific expression patterns, suggesting their potential utility as diagnostic markers or future therapeutic targets (Chong *et al*, 2020).

When compared to animals, plants display a greater degree of developmental plasticity and exhibit substantial changes in gene expression in response to environmental fluctuations. Due to their sessile nature, plants constantly monitor their surroundings, particularly in relation to light, which is their main source of energy. Plants possess various photoreceptor systems, such as phytochromes and cryptochromes, which enable them to continuously assess the quality of light and determine their growth conditions in competition with other plants (Halliday *et al*, 1994; Pedmale *et al*, 2016). Phytochrome photoreceptors respond to both red and far-red light in a toggle switch manner where red light activates PHYB and converts it to the far-red light-responsive Pfr form, while far-red light inactivates PHYB and converts it to the inactive Pr form (Quail, 2002). Shade-sensitive plants like Arabidopsis promote elongation growth in response to a low red-to-far-red ratio. This growth response is mediated by the stabilization of PHYTOCHROME INTERACTING FACTOR (PIF) transcription factors (Lorrain *et al*, 2008). Under conditions of high red light (no shade), PHYB exists in the active Pfr form, leading to the degradation of PIFs. When present, PIFs directly activate genes encoding YUCCA enzymes, resulting in increased production of the plant hormone auxin in response to shade (Müller-Moulé *et al*, 2016; Won *et al*, 2011).

In addition to the pivotal role played by PIFs in regulating growth in response to shade, the class II homeodomain leucine zipper (HD-ZIPII) family of transcription factors has long been recognized as being light-regulated (Carabelli *et al*, 1993). Among the 10 genes that encode HD-ZIPII transcription factors, for *HAT1*, *HAT2*, *HAT3*, *ATHB2*, and *ATHB4*, it has been demonstrated that their expression is transcriptionally upregulated in response to shade (Carabelli *et al*, 2013; Ciarbelli *et al*, 2008). The presence of an EAR-domain in all of these HD-ZIPII genes suggests their function as transcriptional repressors. Initial experiments showed that ATHB2 is capable of binding to a *cis*-element within its own promoter to suppress

its own expression (Ohgishi *et al*, 2001; Steindler *et al*, 1999). In addition to their role in self-regulation, HD-ZIPIIs have been found to influence the expression of genes encoding enzymes and signaling components involved in hormone pathways such as auxin, brassinosteroids, and gibberellic acid, all of which are known to promote growth (Sorin *et al*, 2009). Furthermore, HD-ZIPIIs have been shown to play a crucial role in leaf and meristem development, as loss-of-function mutants exhibit severe defects in maintaining meristems and developing leaves (Bou-Torrent *et al*, 2012; Turchi *et al*, 2015). Interestingly, REVOLUTA (REV), a member of the class III homeodomain leucine zipper (HD-ZIPIII) family, acts upstream of HD-ZIPII factors and also participates in the regulation of shade-induced growth by transcriptionally upregulating *HAT2*, *HAT3*, *ATHB2*, and *ATHB4* (Brandt *et al*, 2012). Furthermore, it has been discovered that a dimer composed of HD-ZIPII and HD-ZIPIII can bind to a conserved *cis*-element in the microRNA genes *MIR165/166*, leading to the repression of their expression specifically in the adaxial domain of developing leaves (Merelo *et al*, 2016). Studies have also revealed that shade influences the rate of leaf proliferation, which is controlled by ATHB2 and ATHB4 (Carabelli *et al*, 2018). Additionally, in the hypocotyl, REV has been found to regulate xylem proliferation in a shade-dependent manner, likely providing structural support to the elongating hypocotyl while increasing water flow (Botterweg-Paredes *et al*, 2020).

HD-ZIPIII factors act as key regulators of meristem maintenance and leaf development, and therefore their regulation occurs at multiple levels, including post-transcriptional and post-translational. At the post-transcriptional level, *microRNA165/166* modulates *HD-ZIPIII* expression, limiting it to the adaxial domain of developing leaves (Mallory *et al*, 2004). At the post-translational level, HD-ZIPIII factors are subject to regulation by LITTLE ZIPPER (ZPR) microProteins (Kim *et al*, 2008; Wenkel *et al*, 2007). MicroProteins are short regulatory proteins consisting of approximately 100 amino acids that share structural similarities with the proteins they regulate (Bhati *et al*, 2021; Eguen *et al*, 2015). The ZPR microProteins specifically possess a leucine zipper domain similar to that found in HD-ZIPIII factors, and they are directly and positively regulated by HD-ZIPIII transcription factors (Brandt *et al*, 2013). Evolutionarily, ZPRs are related to HD-ZIPIII factors, and it has been demonstrated that ZPRs originated from a duplication of an ancient HD-ZIPIII paralog in the euphyllophyte clade, followed by evolutionary trimming through the accumulation of degenerative mutations, resulting in the preservation of only the leucine zipper domain (Floyd *et al*, 2014).

Here, we investigate how shade affects the transcription landscape in Arabidopsis. We find that in response to shading alternative transcription start sites are more often employed, potentially, increasing proteome diversity. One of the alternative transcription start sites is located in the third exon of *ATHB2* and the alternative transcript encodes a microProtein

147 resembling the ZPRs that is devoid of the homeodomain DNA binding moiety but contains the
148 leucine zipper part, necessary to interact with HD-ZIPII proteins. Ectopic expression of
149 ATHB2miP, the microProtein isoform, inactivates ATHB2 by engaging it in a non-productive
150 heterodimer and thereby affects growth of the hypocotyl and the root. By using ATHB2miP as
151 a tool, we can effectively inactivate ATHB2 functions. CRISPR-induced mutations that remove
152 the leucine zipper domain from the *ATHB2* gene result in defects in responding to shade and
153 additional defects in root development. The finding that ATHB2 localizes to nuclear
154 photobodies, mediated by the leucine zipper domain, supports a fundamental role for this
155 domain in the formation of higher-order protein aggregates and signaling. Transgenic plants
156 that ectopically express ATHB2miP exhibit overlapping transcriptome changes with *athb2*
157 mutant plants. Gene ontology and cluster analysis reveal significant changes in genes
158 encoding components of auxin synthesis and signaling, as well as genes controlling root
159 development and nutrient uptake, specifically iron. After discovering that genes related to iron
160 uptake were upregulated in *athb2* mutants, we measured the internal iron levels in both shoots
161 and roots. Our findings showed that the levels were increased in both *athb2* mutants and
162 transgenic plants that ectopically expressed ATHB2miP. Experiments on shade growth
163 conducted on low and high iron media revealed that the availability of iron affects shoot and
164 root growth, which is dependent on the activity of ATHB2. In summary, these findings
165 demonstrate that ATHB2 and ATHB2miP are key mediators between shoot and root growth in
166 response to environmental fluctuations and the nutritional status of the plant.
167

168 Results

169 The *ATHB2* gene harbors an alternative transcription start site in the third exon

170 To mine the transcriptome of *Arabidopsis thaliana* for novel microProteins regulating shade
171 avoidance, we used 5' PEAT sequencing, a method that enriches capped RNA transcripts and
172 results in reads that correspond to transcription start sites (Ni *et al*, 2010). Plants were grown
173 for ten days in a white light regime (WL; R:FR = 7.7) and then exposed to additional far-red
174 light, to simulate shade (WL+FR; R:FR = 0.13; deep shade). After shade treatment for 45- and
175 90-minutes samples, along with an untreated control (0 minutes shade) were collected.
176 Overall, we identified 12,398 transcription start sites (TSSs). The majority of these
177 corresponded to putative TSSs from TAIR10, but 40% (4,982) corresponded to sites that had
178 not previously been annotated (Fig. 1A). Most of the reads mapped to the 5'UTR of already
179 annotated genes, but a subset was found to map to regions within gene bodies. As we were
180 particularly interested in identifying novel transcripts that might encode microProteins (i.e.
181 overlapping gene products), we focused on the subset of genes encoding such alternative
182 microProteins. We therefore established a set of criteria in order to identify sites in our dataset

183 that might correspond to an alternative microProtein mRNA transcript. Firstly, we limited our
184 search to open reading frames (ORFs) that would encode a protein within the size range of a
185 typical microProtein of 5-15 kDa but would differ significantly in size to the overlapping full-
186 length gene product. Secondly, we selected for ORFs that contained a single protein-protein
187 interaction domain allowing the alternative proteins to interfere with the full-length isoforms by
188 forming heterodimers. Finally, we considered the biological relevance of each alternative
189 transcript: the number of reads must exceed a certain threshold to ensure robust expression
190 of the transcript isoform. By applying these criteria, we identified 377 potential microProtein
191 candidates, corresponding to 3% of the total number of TSSs identified in our dataset (Fig.1A).
192 One of the alternative microProtein candidates we identified overlaps with the end of *ATHB2*
193 (Fig. 1B), a gene encoding a class II HD-ZIP transcription factor that plays a central role in the
194 shade avoidance response (Ciarbelli *et al.*, 2008; Schena *et al*, 1993). Like the full-length
195 transcript, the alternative transcript appears to be shade-sensitive, with 5'PEAT-seq showing
196 that expression peaks at 45 minutes. The *ATHB2* alternative transcript was also identified by
197 5'RACE (Fig. S1A) and by sequencing the PCR product, we found that the alternative
198 transcript is spliced at the same site as the full-length transcript and starts around position 653
199 of the full-length transcript (Fig. S1B). In Arabidopsis, the average length of a 5'UTR is 155 bp
200 (Srivastava *et al*, 2018), therefore, we searched for alternative start codons no more than 300
201 bp downstream of the alternative TSS. We identified two near-cognate start codons with strong
202 Kozak sequence similarity (≥ 0.7 and < 0.8) using the online tool TISpredictor (Gleason *et al*,
203 2022). The two start codons are TTG and CTG, both in-frame with the full-length ATHB2
204 protein sequence and producing ORFs that are 126 and 102 amino acids long, respectively,
205 and consisting of the leucine zipper domain, and, in the case of the former, parts of the
206 homeodomain (Fig. S1B). It is unclear whether the alternative transcript we have identified is
207 the result of an independent transcription event or due to post-transcriptional processing of
208 the full-length *ATHB2* transcript involving downstream recapping (Ni *et al.*, 2010). To confirm
209 that ATHB2miP is the result of an independent transcription event, we sequenced full-length
210 cDNAs using PacBio isoseq. This analysis revealed isoforms that start in the third exon of
211 ATHB2, supporting an independent transcription event (Fig. 1C). Furthermore, we also
212 observed antisense transcripts covering the introns. These changes occurred for both isoseq
213 and the 5'PEAT-seq only at the 45-minute timepoint, indicating major chromatin reorganization
214 in response to short-term shading. At 90 minutes, the situation stabilized, and ATHB2 was
215 expressed at higher levels (Fig. 1C). To determine whether there are sequences in the gene
216 body of *ATHB2* that can promote transcription and translation, we fused a β-glucuronidase
217 *uidA* (GUS) reporter gene to different sequences from the *ATHB2* gene body (Fig. 1D). First,
218 we fused GUS to the sequence 1 kb upstream of the *ATHB2* full-length start codon ATG
219 (*pATHB2::GUS*). As a control, we fused GUS to the sequence immediately upstream of an

220 ATG located in the middle of intron two (*pATHB2control::GUS*). Finally, we fused GUS to the
221 sequence immediately upstream of the first alternative, in-frame codon that we had identified,
222 TTG (*pATHB2miP::GUS*). Whilst *pATHB2control::GUS* did not display any GUS staining upon
223 treatment with the substrate 5-bromo-4-chloro-3-indolyl ß-d-glucuronide (X-Gluc; Fig. 1F),
224 both *pATHB2::GUS* and *pATHB2miP::GUS* showed staining in the shoot apical meristem
225 (SAM) and the hypocotyl, though to a lesser extent in the latter in *pATHB2miP::GUS* (Fig. 1E,
226 G). This suggests that there are sequences within the second intron and third exon, that are
227 excluded from *pATHB2control::GUS*, that can promote both transcription and translation. This
228 might also lend support to our hypothesis that the alternative transcript arises through an
229 independent transcription event to the full-length *ATHB2* transcript.
230

231 **ATHB2miP can heterodimerize with itself and the full-length ATHB2 isoform**

232 We hypothesized that the ATHB2miP might inhibit the full-length ATHB2 in a manner similar
233 to that of the LITTLE ZIPPERs (Kim *et al.*, 2008; Wenkel *et al.*, 2007). To test this hypothesis,
234 we posed the following questions. Firstly, does the microProtein localize to the same cellular
235 compartment as the full-length protein? It was previously reported that ATHB4, another
236 member of the HD ZIPII family, contains a nuclear localization signal (NLS) in its homeodomain
237 that is highly conserved in ATHB2 (Gallemi *et al*, 2017). However, our own analysis using
238 LOCALIZER (Sperschneider *et al*, 2017), which has been trained specifically on plant proteins,
239 suggests that there is an NLS in the LZ domain that is conserved in other HD-ZIPIIs, including
240 ATHB4. Interestingly, LOCALISER also predicts a chloroplast transit peptide (cTP) at residues
241 1-46 of ATHB2 with a probability of 0.8. The cTP does not appear to be conserved, suggesting
242 that, if real, it is specific to ATHB2. To determine the subcellular localization of ATHB2 and
243 ATHB2miP, we transiently expressed eGFP fusions of the proteins in *Nicotiana benthamiana*
244 leaves (Fig. 2A; eGFP-ATHB2, eGFP-ATHB2miP). eGFP-ATHB2miP encompasses the
245 entirety of exon 3 and 4 of ATHB2. To further test whether the homeodomain of ATHB2 could
246 target the nucleus, as shown previously for ATHB4 (Gallemi *et al.*, 2017), we also fused a
247 variant of ATHB2 lacking the leucine zipper to eGFP (eGFP-ATHB2HD). We observed that all
248 variants of ATHB2 localized exclusively to the nucleus, unlike the negative control (eGFP)
249 which was dispersed along the cell membrane and in the nucleus (Fig. 2A). We also observed
250 that the distribution of eGFP-ATHB2 was not uniform in the nucleus, but instead localized to
251 distinct nuclear speckles or, in some case, to the nucleolus specifically. Nuclear speckles were
252 mostly absent in leaves expressing eGFP-ATHB2miP, but entirely so in leaves expressing
253 eGFP-ATHB2HD, suggesting that interactions mediated by the leucine zipper domain and in
254 part by the N-terminal domain are required for the speckles to form. We next asked whether
255 we could find evidence that ATHB2 and ATHB2miP can physically interact with each other. It
256 has been demonstrated that ATHB2 homodimerizes in order to bind to its target DNA

257 sequence and that this interaction is directed by the leucine zipper domain (Sessa *et al*, 1993).
258 To test whether ATHB2 and ATHB2miP can interact, we performed a yeast-2-hybrid (Y2H)
259 study. As in the localization assay, ATHB2 or ATHB2miP were fused to either the activation or
260 the DNA binding domain of GAL4. Constructs were transformed into yeast carrying the *HIS3*
261 gene fused to the target DNA sequence of *GAL4*. Growth of transformed cells on media lacking
262 histidine revealed that ATHB2 and ATHB2miP can interact and form heterodimeric complexes
263 as well as homodimers (Fig. 2B). To further confirm this interaction *in planta* we carried out a
264 FRET-FLIM assay. Here again, ATHB2 and ATHB2miP were fused to either eGFP or mCherry
265 and then transiently co-expressed in tobacco leaves. Using a confocal microscope,
266 fluorescence lifetime of eGFP-ATHB2 was measured. We observed that in cells that co-
267 expressed fluorescently tagged versions of ATHB2 or ATHB2miP there was a significant
268 decrease in fluorescence lifetime of eGFP-ATHB2 when compared to cells that expressed
269 eGFP-ATHB2 alone or with a control (mCherry; Fig. 2C), supporting that ATHB2 and
270 ATHB2miP can physically interact.
271

272 **ATHB2miP perturbs the ability of ATHB2 to regulate transcription**

273 To determine the functional consequences of the interaction between ATB2miP and ATHB2,
274 we decided to study the transcriptional changes elicited by ectopic expression of ATHB2miP.
275 For this, we grew transgenic plants overexpressing ATHB2miP, an *athb2* mutant, and wild-
276 type plants in white light conditions. Prior to RNA extraction, plants were exposed to 0, 45 and
277 90 minutes of additional far-red light. RNA-seq was then used to compare the transcriptomes
278 of respective plants under white light and shade conditions. We hypothesized that
279 overexpressing ATBH2miP would result in the inactivation of ATHB2, similar to the inactivation
280 of HD-ZIPIII transcription factors by ZPR microProteins (Kim *et al.*, 2008; Wenkel *et al.*, 2007).
281 Differential gene expression analysis, adjusting for genotype, environment and genotype by
282 environment interaction, identified approximately 5,900 differentially expressed genes (DEGs)
283 with FDR $\leqslant$0.05 and significant up/down-regulation (logFC>1; logFC<1). As it is known that
284 HD-ZIPIIs establish an autoregulatory feedback loop by binding to their own promoter and
285 repressing their own expression (Ciarbelli *et al.*, 2008), we examined the *ATHB2* gene in all
286 three genotypes. No reads mapping to the *ATHB2* locus were detected in the *t-athb2* mutant
287 (Figure S2), indicating a true null allele. The transgenic plants overexpressing ATHB2miP
288 showed an increase in the number of reads covering exons 1 and 2 compared to wild type
289 plants, suggesting an increase in the endogenous expression of *ATHB2* (Figure S2). These
290 findings indicate that ectopic expression of ATHB2miP sequesters the endogenous ATHB2
291 protein and inactivates the negative auto-repressor loop.

292 In total, we identified 95 DEGs whose expression depended on genotype, environment, and
293 genotype x environment (Fig. 3A). The GO analysis of these 95 genes revealed an
294 overrepresentation of genes involved in the shade avoidance response, growth, and
295 hormone/auxin signaling (Fig. S3A). These findings demonstrate that both *t-athb2* mutant
296 plants and the transgenic plants overexpressing ATBH2miP can detect shade and respond to
297 it at the molecular level. In order to compare the genotypes and environmental conditions, we
298 carried out an unbiased gene cluster analysis using mFuzz (Kumar & Futschik, 2007). Our
299 analysis revealed 17 individual clusters with distinct expression patterns (Fig. S3B). Cluster 1
300 (Fig. 3B) contains 200 genes that are upregulated only in *35S::miP*, but not in either of the
301 other two genotypes. We found an overrepresentation of genes involved in root development
302 in this cluster. Cluster 2 contains 382 genes with lower expression in both 35S::miP and *t-*
303 *athb2*. This cluster is also enriched in genes encoding hormonal regulators and regulators of
304 root development. Finally, cluster 14 contains 152 genes that are up-regulated in both
305 *35S::miP* and *t-athb2* plants (Fig. 3B). Genes in this cluster have known roles in the shade
306 avoidance response. We also found an overrepresentation of genes involved in iron uptake
307 and signalling, which are strongly upregulated in *35S::miP* and *t-athb2* plants. The latter
308 finding suggests that ATHB2 may play a role in mediating between nutritional status and
309 growth responses.
310 Having established that ectopic expression of ATHB2miP perturbs ATHB2 function and
311 alleviates the negative auto-repressive feedback loop we assessed DEGs in the three different
312 environmental conditions (0-, 45-, and 90-minutes additional FR-light). Focusing on the genes
313 upregulated in both *t-athb2* and *35S::miP* plants, we found 22 genes in white light (Fig. 3C),
314 157 genes in 45-minute shade (Fig. 3D), and 20 genes in 90-minute shade (Fig. 3E). As
315 previously observed, we identified upregulated genes related to auxin synthesis and signaling,
316 as well as genes involved in root development. Genes related to auxin-biology and hypocotyl
317 elongation (Table 1) include *YUCCA3* and *YUCC9*, both of which have known functions in
318 auxin biosynthesis in response to shade. (Goyal *et al*, 2016). Genes associated with root
319 development (Table 2) include several cell wall modifying enzymes, most notably expansins,
320 as well as transcription factors. Taken together, these findings support the function of ATHB2
321 as a suppressor of auxin signaling and root development. Analysis of genes down-regulated
322 in both *t-athb2* and *35S::miP* plants showed that genes encoding defence regulators were
323 enriched in all conditions (0, 45 and 90 min additional FR light). In addition, we found that
324 genes involved in the biosynthesis of tryptophan were down-regulated in the shaded
325 environments. Since auxin is mainly synthesised from tryptophan (Tao *et al*, 2008; Won *et al.*,
326 2011), the latter finding may explain why *athb2* mutant plants have defects in shade-induced
327 growth promotion (Ciarbelli *et al.*, 2008; Schena *et al.*, 1993).
328

329 **ATHB2 controls elongation growth and root development in response to shade**
330 In shade-avoidant plants, such as *A. thaliana*, shade caused by neighboring plants elicits a
331 set of growth responses including elongation of stem-like structures, hyponasty and leaf
332 extension, and acceleration of flowering (Martinez-Garcia & Rodriguez-Concepcion, 2023). It
333 has been demonstrated that ATHB2 plays a central role in the shade avoidance response,
334 with transgenic Arabidopsis plants where *ATHB2* is affected displaying varying degrees of
335 shade response defects, the most obvious being in the elongation of the hypocotyl (Steindler
336 *et al.*, 1999). Early studies found that when knocked down or overexpressed, *ATHB2* was
337 associated with subsequent shade responses where there was less or more hypocotyl
338 elongation compared to the WT, respectively. For this reason, and because we show that the
339 two proteins interact, we expected that transgenic plants where ATHB2miP (*35S::miP*) is
340 overexpressed would show a phenotype indicative of lost ATHB2 function, i.e. shade
341 insensitivity. To better understand the role of the region corresponding to ATHB2miP, we also
342 created mutant plants by CRISPR-Cas9 gene editing to compare with the overexpression and
343 KO line already described (Fig. 4A). The first, *athb2ΔLZ*, presents a loss-of-function allele of
344 *ATHB2* caused by a 25-bp deletion in the third exon, causing a frameshift and nonsense
345 mutation, and resulting in a protein product lacking the leucine zipper and part of helix III of
346 the homeodomain. Secondly, *athb2Δ,* is a complete loss-of-function of *ATHB2*, caused by a
347 frameshift mutation in the beginning of the first exon, resulting in a truncated protein that
348 shares only its first twelve amino acids with ATHB2. Based on RT-qPCR results where we
349 analyzed expression of *ATHB2* after 45 minutes in shade, we see that transcription is not
350 affected in any of the CRISPR lines we have generated, unlike in *t-athb2* where *ATHB2* is
351 clearly downregulated (Fig. 4B), as also shown in the RNA-seq analysis (Fig. S2). To measure
352 responsiveness to shade, we grew these mutant and transgenic plants alongside the Col-0
353 wild-type (WT) under three different shade regimes (Fig. 4C). We grew plants in deep shade
354 (13 µmol m-2 s-1 PAR in WL and WL+FR; R:FR = 0.13 in WL+FR) with reduced PAR and a
355 very low R/FR-ratio. In addition, we used canopy shade (45 µmol m-2 s-1 PAR in WL and
356 WL+FR; R:FR = 0.15 in WL+FR) and proximity shade (45 µmol m-2 s-1 PAR in WL and
357 WL+FR; R:FR = 0.06 in WL+FR) with higher PAR but different F/FR-ratios. In deep shade, we
358 observed that *athb2* mutants were hypersensitive, with elongated hypocotyls in white light
359 conditions and, as in the case of *35S::miP*, with very long hypocotyls in shade. Interestingly,
360 the CRISPR mutants *athb2Δ* and *athb2ΔLZ* exhibited responses similar to those of wild-type
361 plants. It is worth noting that ATHB2 is still expressed in these mutants (Fig. 4B), which
362 suggests that ATHBmiP and other transcript or protein isoforms may still be produced. Under
363 canopy shade, the *athb2* mutants exhibited a growth behaviour that was less sensitive to
364 shade, resulting in reduced hypocotyl elongation in shaded conditions. However, under

365 proximity shade, there were no significant differences in hypocotyl elongation between the wild
366 type and *athb2* mutant plants (Fig. 4C). The results indicate that ATHB2 has a dual function.
367 In deep shade, it acts as a growth repressor, whereas in canopy and proximity shade, it acts
368 as a growth promoter.
369 Class II HD-ZIPs have been suggested to have some redundant function in the shade
370 avoidance response, so it is possible that ATHB2miP also interacts with these other proteins
371 in the absence of full-length *ATHB2*. To determine whether the effect of *35S::miP* depends on
372 the presence of the endogenous copy of *ATHB2*, we complemented the *t-athb2* knock-out with
373 *35S::miP*. In deep shade, we observed that both *35S::miP* and *35S::miP t-athb2* plants
374 developed longer hypocotyls under white light, indicating that ATHB2miP may have inhibited
375 other transcription factors. Additionally, *35S::miP* plants had significantly longer hypocotyls
376 than *35S::miP t-athb2* plants in deep shade, suggesting that the absence of full-length ATHB2
377 may contribute to this difference.
378 The shade avoidance response is also associated with changes in root architecture. For
379 example, one study found that lateral root density was decreased in plants that were exposed
380 to longer periods of shade (van Gelderen *et al*, 2018). Mutations affecting ATHB2 have been
381 reported to affect root development, disrupting both root length and lateral root formation
382 (Schena *et al.*, 1993), and we also identified a number of genes related to root morphogenesis
383 in our RNA-seq analysis. Therefore, we questioned whether ATHB2miP could also have a role
384 in root development.
385 Following the shade regimes described earlier, plants were grown in either WL+FR or WL for
386 a period of 11 days and primary root length was measured. Our results indicate that there
387 were no significant differences in all conditions tested, except for the *t-athb2* mutant plants
388 which had slightly shorter roots in white light and even shorter roots in the proximity shade
389 regime (Suppl. Fig. S4). We also assessed the ability of plants to form lateral roots under light
390 and shade conditions. While no difference in the number of lateral roots respective to
391 geneotypes and environmenal conditions was observed in deep shade and canopy shade, we
392 detected significant differences in proximity shade conditions (Fig. 4D). In white light, *t-athb2*
393 mutants, CRISPR mutants (*athb2Δ* and *athb2ΔLZ*), and *35S::miP* transgenic plants all all had
394 significantly lower numbers of lateral roots, suggesting that ATHB2 promotes lateral root
395 emergence broadly. In response to shade, a decrease in the number of lateral roots was
396 observed in wild type plants. However, in *t-abhb2*, *athb2Δ* and *athb2ΔLZ* mutant plants, the
397 number of lateral roots did not change in shade. This indicates that ATHB2 is required to
398 repress lateral root emergence in shade. Ectopic expression of ATHB2miP severely affected
399 lateral root development in shade, suggesting that both ATHB2miP and ATHB2 act to repress
400 lateral root emergence in shade. The *35S::miP* transgenic plants exhibited a strong

suppression phenotype, which was reversed upon loss of *ATHB2* in *35S::miP* transgenic plants (*t-athb2 35S::miP*), indicating that a functional *ATHB2* is required and that ATHB2miP and ATHB2 work together in a repressive complex.

Taken together, the results indicate that ATHB2 regulates hypocotyl elongation in response to shade in a shade-dependent manner. It is hypothesized that ATHB2 homodimers exist in balance with heterodimers of ATHB2 and ATHB2miP. The role of ATHB2miP is to sequester ATHB2 and prevent excessive inhibition of hypocotyl growth in early shade. Furthermore, it was observed that ectopic expression of ATHB2miP resulted in heightened sensitivity to deep shade in comparison to the other LOF alleles. This observation may suggest that ATHB2miP has the potential to interfere with other HD-ZIPs involved in shade avoidance, although its primary target is ATHB2. This effect is also evident in the root, where ectopic expression of ATHB2miP affects lateral root emergence in an ATHB2-dependent manner.

The shade avoidance response is influenced by nutritional status regulated by ATHB2

ATHB2 regulates the expression of genes involved in shade avoidance and root development (Fig. 3) and mutations in *ATHB2* result in defects in these processes (Fig. 4). After observing an increase in the expression of genes involved in iron transport and metabolism (Supplementary dataset 2), we investigated whether iron content was altered in *athb2* mutants and whether iron affected shade-induced hypocotyl elongation and root development.

Iron homeostasis is under the regulation of a transcriptional cascade involving genes encoding iron transporters and a family of small peptides involved in iron deficiency signalling (Spielmann *et al*, 2023). The basic helix-loop-helix family of transcription factors, specifically bHLH38, bHLH39, bHLH100, and bHLH101, have been identified as intermediaries in iron deficiency signaling for the iron transporter IRT1. They achieve this by dimerizing with FIT1, FER-LIKE IRON DEFICIENCY INDUCED TRANSCRIPTION FACTOR1 (Colangelo & Guerinot, 2004).

Transcriptome profiling of wild type, *t-athb2* and *35S::miP* seedlings revealed genes that were upregulated in both *t-athb2* and *35S::miP* plants (Fig. 3B), indicating that these genes are repressed by ATHB2. The iron deficiency cascade, which includes *bHLH38*, *bHLH39*, *bHLH100*, *bHLH101*, *FIT1*, the iron transporter *IRT1*, and the *IRONMAN* genes (*IMA1*, *IMA2*, and *IMA3*), was among the ATHB2 repressed genes and found to be upregulated in *35S::miP* and *t-abth2* seedlings (Fig. 5A,B). In white light conditions, a smaller set of genes, including *bHLH38*, *bHLH101*, *IMA1*, and *IMA2*, are already highly expressed (Fig. 5A). The remaining genes are upregulated after 45 minutes of additional far-red light (Fig. 5B), suggesting negative regulation by ATHB2 in response to shading. This regulation is uncoupled in *t-athb2* and *35::miP* plants.

To determine whether the observed increases in transcription of genes encoding iron signaling and uptake components are related to iron levels, we measured the iron content of wild type, *35S::miP*, and *t-athb2* mutant seedlings. Iron is taken up by the roots and then distributed throughout the plant, and we therefore decided to measure iron levels in both roots and shoots. As anticipated, higher concentrations of iron were found in the roots than in the shoots (Fig. 5C). Both *t-athb2* and *35S::miP* transgenic plants exhibited significantly higher levels of iron in both roots and shoots. In comparison, *35S::miP* plants had the highest internal iron concentration, which can be explained by the ability of ATHB2miP to affect class II HD-ZIP transcription factors beyond ATHB2. These findings collectively support the role of ATHB2 as a negative regulator of iron uptake.

To test whether nutrient status affects root development and shade avoidance responses, we performed shade avoidance assays with seedlings grown on low iron (10μM) and high iron (100μM) media. Overall, we observed a significant hypocotyl shortening and reduced hypocotyl response when seedlings were grown on high iron media compared to low iron media. The responses to deep shade on low-iron medium were similar to those observed earlier (Fig. 4C), except that *t-athb2* mutants did not elongate hypocotyls under white light conditions. On high iron media, we observed a dampening of the shade avoidance response, but *t-athb2*, *athb2Δ* and *athb2ΔLZ* mutants showed longer hypocotyls than wild type under shade conditions. Transgenic plants ectopically expressing ATHB2miP (35S::miP and *t-athb2 35S::miP*) showed slightly elongated hypocotyls in white light and almost no elongation in shade. These results show that nutrient status has a strong influence on the shade avoidance response and that the function of ATHB2 is dependent on iron homeostasis. Under low iron conditions, ATHB2 acts as an activator of shade growth, whereas under high iron conditions it suppresses shade growth. With regard to root growth, we measured the length of the primary root on low and high iron media in the absence and presence of shade. Overall, we did not observe strong changes in root elongation between white light and shade grown seedlings, supporting what we had observed previously (Fig. S4). However, we observed that all *athb2* mutant plants, with the exception of *athb2Δ* mutants, developed significantly shorter roots when grown on low-iron medium. This finding may indicate that these mutants, in which iron uptake is strongly up-regulated, do not need to grow longer roots to increase their iron uptake. The latter hypothesis is supported by the finding that plants ectopically expressing ATHB2miP have the shortest roots but show the greatest hypocotyl elongation in shade. On high-iron medium, we did not observe major changes in root length, except for *athb2Δ* mutants, which had significantly shorter roots in both white light and shade conditions. In conclusion, these results show that nutrient status affects both root length and hypocotyl elongation in response

472 to shade. ATHB2 mediates between root growth and hypocotyl elongation and can either
473 activate or repress hypocotyl elongation in response to different iron concentrations.
474

475 Discussion

476 In this study, we assessed the impact of shading on transcript isoform production and their
477 potential to encode microProteins. It was found that almost half of the Arabidopsis genes that
478 had reads at the transcription start site also had alternative transcription start sites in the gene
479 body. Only a fraction of these genes had the potential to encode microProteins. It is
480 furthermore fair to assume that many more potential microProteins exist, but these are
481 produced in a tissue-specific and/or inducible fashion. We filtered the dataset to prioritize
482 alternative transcripts with robust expression levels, excluding those unsupported by multiple
483 reads, which again could be candidates that have a narrower pattern of expression. Our
484 approach focused on enriching microProtein candidates likely to have biological significance.
485 To enhance the likelihood of identifying microProteins involved in the shade avoidance
486 response, we considered alternative methods for studying transcript isoforms, including
487 parallel analysis of RNA 5′ends from low-input RNA (nanoPARE), and ribosome profiling
488 (Ingolia, 2014; Schon *et al*, 2018). Ribosome profiling, which employs deep sequencing to
489 monitor *in vivo* translation, can provide additional insights into protein formation. Such future
490 approaches can both confirm the purely sequencing-based identification of candidate
491 microProteins but also aid to identify novel candidates.
492 The list of microProtein candidates generated through the 5'PEAT-seq approach contained
493 few shade avoidance regulators. Notably, microProteins may serve distinct functions
494 compared to their full-length counterparts, albeit often targeting evolutionarily related proteins.
495 Adjusting filtering criteria could identify more transcript isoforms related to shade responses,
496 although they may not classify as microProteins. The ATHB2miP microProtein, chosen due to
497 its similarity to the known LITTLE ZIPPER microProteins (Kim *et al.*, 2008; Wenkel *et al.*,
498 2007), was expected to interact with ATHB2 and regulate its activity, akin to ZPRs and HD-
499 ZIPIIIs. Yeast-Two-Hybrid tests and pull-down assays confirmed physical interactions between
500 ATHB2 and ATHB2miP. Notably, ATHB2miP can also homodimerize, a feature attributed to the
501 leucine zipper domain's role as a protein-protein interaction motif in HD-ZIP proteins.
502 We assessed potential alternative *ATHB2miP* promoters using the GUS protein assay,
503 identifying one active promoter, *pATHB2miP::GUS*, which did not possess a typical ATG start
504 codon. This observation is consistent with prior findings of alternative start sites not
505 commencing with an ATG. Differences in GUS expression patterns between *ATHB2* and
506 *ATHB2miP* promoters were observed, with ATHB2miP primarily present in the shoot apical

507 meristem (SAM) and early leaf primordia. The absence of ATHB2miP in other tissues
508 suggested that it does not directly regulate ATHB2 in those areas.
509 Analysis of the subcellular localization patterns of ATHB2miP and ATHB2 protein-fusions to
510 GFP revealed that ATHB2 localizes mostly to larger subnuclear speckles, a phenomenon also
511 observed for phytochromes and other light signaling components and therefore often referred
512 to as photobodies (Kim *et al*, 2023; Kircher *et al*, 2002). Interestingly, ATHB2miP protein
513 showed a more diffuse nuclear localization implying discrete functions of both isoforms.
514 However, when co-expressed both isoforms appear in speckles, but these are more dispersed,
515 indicating a direct physical interaction where the full-length isoform is partially removed from
516 the speckles. The full-length ATBH2 isoform has an intrinsically disordered amino terminus
517 and could undergo phase separation and engage in more flexible protein-protein interactions
518 within the photobodies. As ATHB2miP lacks the disordered part, it may prevent the phase
519 separation activity of ATHB2.
520 To investigate the regulatory role of ATHB2miP, we compared *ATHB2miP* overexpression
521 plants to *athb2* mutants. Surprisingly, in deep shade conditions all *athb2* knock-down and
522 knock-out mutants as well as *ATHB2miP* overexpression plants showed elongated hypocotyls
523 in white light conditions compared to Col-0 wild type plants while the hypocotyl length in shade
524 grown seedlings showed no significant differences when compared to wild type. Except
525 35S::miP transgenic plants that showed excessive hypocotyl elongation in both white light and
526 far-red enriched conditions (Fig. 4C). This indicates that ATHB2 functions as a growth
527 repressor in deep shade and ATHB2miP has the capacity to inactivate ATHB2 which results
528 in hypocotyl elongation in shade. The finding that the growth differences can only be observed
529 in deep shade conditions indicates that ATHB2 activity is dependent on other components that
530 are only present in this type of shade. In canopy and proximity shade ATHB2 functions as a
531 promoter of elongation growth, hence the shorter hypocotyls in these types of shade.
532 ATHB2 has also been reported to be involved in controlling root growth and development
533 (Schena *et al.*, 1993) prompting us to investigate the role of ATHB2 and ATHB2miP in this
534 process. The analysis of root development in different types of shade did not reveal significant
535 differences in root length. However, there were notable differences in the ability of *athb2*
536 mutants to produce lateral roots. Additionally, all tested *athb2* mutants appeared to be
537 insensitive to shade, as the number of lateral roots did not change in response to proximity
538 shade, unlike in the wild type. In summary, these results suggest that ATHB2 acts as a
539 suppressor of lateral root growth in response to proximity shade. The overexpression of
540 ATHB2miP had a significant impact on lateral root development in both white light and shade.
541 Notably, in the *t-athb2* mutant background (*t-athb2 35S::miP*), the strong shade suppression
542 was not observed, indicating that ATHB2 is required for strong shade suppression. These
543 findings suggest that ATHB2miP has a role beyond ATHB2 inhibition. The discovery that

ATHB2 has diverging functions depending on the type of shade experienced by the plant raises questions about the upstream signalling processes that control ATHB2 activity and the type of protein complexes in which ATHB2 participates. Additionally, the localization of ATHB2 to photobodies hints at a role in phytochrome B signalling. In deep shade, phyA may have additional roles that cause ATHB2 to switch from being a growth activator to being a growth inhibitor. To better understand ATHB2's molecular functions in different light conditions, it is important to study the protein complexes it interacts with.

To elucidate the molecular events resulting from the absence of *ATHB2* or the inhibition of ATHB2 by overexpression of ATHB2miP, we conducted an RNA-sequencing analysis comparing Col-0 wild type with *t-athb2* and transgenic plants overexpressing ATHB2miP (*35S::miP*), all grown under both white light and deep shade conditions. In line with the exaggerated growth phenotype of the hypocotyl in deep shade, we found auxin- and shade-related genes to be expressed at elevated levels. The fact that the expression of these genes follows the same trend in both the *t-athb-2* and the *35S::miP* line also supports the negative feedback mechanism we hypothesised. In addition, we also found genes involved in root growth and development to be over-represented in white light grown seedlings, which corresponds to what we observe in our lateral root phenotyping results. Genes involved in the formation of lateral roots were found to be upregulated in both the *t-athb2* and the *35S::miP* seedlings, even in shade. Using gene clustering analysis, especially cluster 2 (Fig. 3B) comprises genes that are down-regulated in both the *t-athb2* and the *35S::miP* lines and several of these genes are involved in root growth and development. Surprisingly, genes that were upregulated in *t-athb2* and *ATHB2miP* overexpression plants encode iron transporters. The finding that iron concentrations were elevated in these plants implies that these roots might have stopped growing because of a better nutrient supply. Our findings show that iron availability influences the shade avoidance response and primary root length. Furthermore, it appears that the nutritional status influences the activity of ATHB2, suggesting that ATHB2 is integrating different environmental inputs (shade and nutrient availability) to formulate an appropriate response that can either activate or repress growth.

Shade not only promotes elongation growth, but it also affects the plant's hormone homeostasis, weakening its ability to defend against pathogens (de Wit *et al*, 2013). The recent discovery that regulation of iron uptake is influenced by internal IRONMAN peptides, which are sensitive to external flagellin perception (Cao *et al*, 2024), raises the question of whether ATHB2 may play a central role in mediating between nutritional immunity and shade-induced growth promotion or suppression.

In summary, our work uncovered transcript isoforms that are responsive to shading, including ATHB2miP, a microProtein derived from an alternative transcription start site in the *ATHB2* gene. This microProtein appears to intercede between root and shoot growth in response to

581 shade. In summary, our data support a model for ATHB2/ATHB2miP as integrator of nutrient
582 status and elongation growth of both the shoot and the root.
583

584 Materials and methods

585 Plant materials and growth conditions

586 All plants were grown in growth chambers on ½ Murashige and Skooge (MS) 1% agar under
587 long day (LD) conditions at 22°C during daytime and 20°C during night. For the shade
588 treatment for the 5'PEAT-seq and RNA-seq experiments, seedlings were transferred after 10
589 days from a white light only compartment (PAR of 13 µmol/m^2/s, R:FR = 7.7) to a compartment
590 with additional far-red lights (PAR of 13 µmol/m^2/s, R:FR = 0.2) for treatments of 0, 45 and 90
591 minutes. Whole seedlings frozen in liquid N_2 before grinding with a mortar and pestle and RNA
592 extraction. For the shade treatments related to measuring hypocotyl length and root growth,
593 half of the plants were transferred after two days from a white light only compartment (PAR of
594 45 µmol/m^2/s, R:FR = 5.5) to a compartment in the same chamber with additional far-red lights
595 (PAR of 45 µmol/m^2/s, R:FR = 0.15). The other half remained in the white light only
596 compartment. After five and eleven days, seedlings were photographed and hypocotyls and
597 roots measured using IMAGEJ.
598 For the growth and phenotyping of seedlings on iron-enriched and iron-depleted media, two
599 different growth media were prepared. Both media contained the same vitamins, macro and
600 micro elements of standard MS salt (M0222, Duchefa) and in the same concentrations, with
601 the exception of FeNAEDTA, used in a concentration of 10uM in the iron-depleted medium
602 and 100uM in the iron-enriched medium. Seedlings were grown on 1% agar iron-depleted and
603 iron-enriched media under the same growth conditions of the other growth experiments. Half
604 of the seedlings on plates were transferred after two days from a white light only compartment
605 (PAR of 13 µmol/m2 /s, R:FR = 7.7) to a compartment in the same chamber with additional
606 far-red lights (PAR of 13 µmol/m2 /s, R:FR = 0.2). The other half remained in the white light
607 only compartment. After five and eleven days, seedlings were photographed and hypocotyls
608 and roots measured using IMAGEJ.
609

610 5'-PEAT library preparation and analysis

611 Arabidopsis Col-wild type seedlings were grown for 10 days on 1/2 MS agar plates as
612 described before. Whole seedlings were harvested directly into liquid nitrogen after 0, 45 or
613 90 minutes of shade treatment. Total RNA was extracted with Spectrum Plant Total RNA Kit
614 (Sigma). PEAT library preparation was done exactly as described earlier (Ni *et al.*, 2010), but
615 all steps beginning with the circularization of the PCR product were omitted and the fragments
616 were inserted into NEBNext Ultra™ II DNA Library Prep K it for Illumina and quantified with

617 NEBNext Library Quant Kit for Illumina. The finished libraries had a fragment size of 500 bp.
618 BGI tech solutions (Hongkong) performed sequencing on an Illumina HiSeq (150bp paired
619 end, 40 M read pairs per sample).
620 Quality control was done with FastQC (v0.11.5), available online at:
621 http://www.bioinformatics.babraham.ac.uk/projects/fastqc) and generic Illumina adapters
622 were removed with Cutadapt (Martin, 2011). Additionally, Cutadapt removed PEAT-adapters
623 and marked read direction for later identification of the RNA start. HISAT2 (v2.0.5) (Kim et al.,
624 2015) aligned the reads with more than 90% success to the Arabidopsis thaliana TAIR9
625 genome assembly (Lamesch *et al*, 2012). Samtools (v0.1.19) (Li *et al*, 2009) and bedtools
626 (v2.17.0) (Quinlan & Hall, 2010) filtered for properly paired reads with high aligment score (-q
627 10) and performed file conversions. Around 30 million read pairs per sample were further
628 analysed. Using previously marked read directions, RNA starts were identified and recorded
629 in bedgraph format. Only genes with a transcription start site (TSS) in the 5'UTR or upstream
630 region and at least one alternative TSS in coding sequence or intron region were further
631 analysed. The difference between the TSS and the alternative start site was set to 10%. To
632 test whether the regulation of the 5'UTR-TSS and the alternative TSSs are different, the
633 number of TSS-reads in the 5'UTR-TSS and one alternative TSS were compared in two
634 samples with Fisher's exact test and Benjamini-Hochberg corrected false discovery rate in R
635 software (v3.2.2). The protein size was set between 7 and 20 kDa. Pfam-A domains (v28,
636 ftp://ftp.ebi.ac.uk/pub/databases/Pfam/) (Finn *et al*, 2016) were searched in protein sequences
637 with hmmer3 (http://hmmer.org/) and cutoffs of e-value <= 0.1 and c-value <= 0.05. Interaction
638 domains were identified using iPfam (v1.0, http://www.ipfam.org/) (Finn *et al*, 2014).
639

640 **Generation of transgenic and CRISPR-Cas9 edited plants**

641 To generate the overexpression of ATHB2miP (*35S::ATHB2miP*) in Arabidopsis oligos were
642 designed to amplify exons 3 and 4 of *ATHB2* and the resulting sequence then cloned into the
643 vector pJAN33 by gateway cloning. Plants were transformed by floral dipping and
644 transformants selected for by BASTA treatment.
645 For CRISPR mutagenesis, sgRNAS were cloned into the CRISPR/Cas9 vector pKI1.1R. The
646 cloning was performed as described (Tsutsui and Higashiyama, 2016). *Athb2-cr2* was made
647 by Cas9 gene editing targeted using the following sgRNA combination: 5'-
648 TACAGTCTCAAGCTCTACA-3' and 5'-GTTTCAGAACAGACGAGCA-3'. *Athb2-cr3* was
649 made by Cas9 gene editing targeted using the single sgRNA: 5'-ACGATCTGGGTCTAAGCTT-
650 3'. Plants were transformed by floral dipping and then RFP-positive seeds selected for in the
651 T1 generation. Genomic DNA of 4-week old plants was isolated using the Edwards method
652 and the plants were genotyped by PCR. The PCR products were gel purified using the E.Z.N.A
653 ® Cycle-Pure Kit (Omega Bio-tek) and sequenced at Macrogen (Amsterdam).

654 Expression in transgenic and mutant plants was measured by RT-qPCR. First, RNA was
655 extracted from 4-week old plants using the Spectrum Plant Total RNA Kit (Sigma). Purified
656 RNA (1 µg) was used for reverse transcription using the iScript™ cDNA Synthesis Kit (BIO-
657 RAD). RT-qPCR was performed using SYBR green (ThermoScientific) on a Biorad CFX384,
658 using oligos against *ATHB2*: Region A FWD 5'-GATTGCAAAATCTCTCTCTCTCTC-3' with
659 REV 5'-GACCCAGATCGTCTTTCTCG-3'; Region B FWD 5'-
660 TGTTCGAGAAAGACGATCTGG-3' with REV 5'-GTCGATTCCTCGGATGAAAG-3'; Region C
661 FWD 5'-CAGTGACGATGAAGATGGTGA-3' with REV 5'-GAAACCAAACTTCCACTTGTCT-
662 3'; Region D FWD 5'-GCTGAAGCAAACGGAGGTAG-3' with REV 5'-
663 GGACCTAGGACGAAGAGCGT-3'. Gene expression levels were standardized to the
664 housekeeping gene GAPDH (FWD 5'- AAAGTGTTGCCATCCCTCAA- 3' with REV 5'-
665 TCGGTAGACACAACATCATCCT-3') using the Pfaffl equation.
666

667 **Yeast-2-Hybrid**

668 For the Y2H assay, the coding sequences of *ATHB2* and *ATHB2miP* (exons 3 and 4 of *ATHB2,*
669 as above) were recombined into vectors pGBKT7-GW and pGADT7-GW by gateway cloning.
670 The constructs were transformed to yeast strains PJ694α and YM4271A using lithium acetate
671 and heat shock for 15 min (42°C). Yeast colonies that grew after two days at 30°C on SD
672 dropout medium plates lacking either tryptophan (-W) or leucine (-L) were selected for, mixed
673 with each other and spread on SD dropout medium lacking tryptophan and leucine (-LW).
674 Colonies were finally plated on dropout -LW medium that additionally lacked histidine (–LWH),
675 with additional 10 mM 3-aminotriazole to test for protein interactions.
676

677 **GUS localization assay**

678 Oligos were designed to amplify the 1 kb sequence upstream of *ATHB2,* and the sequences
679 upstream of the alternative start codon TTG in exon 3 and upstream of a control start codon
680 ATG in intron 2, up to and including the putative start codon ATG of *ATHB2.* Sequences were
681 cloned into pBGWFS7 by gateway cloning and the resulting vectors transformed into plants
682 by floral dipping. Arabidopsis thaliana (ecotype Col-0) seeds were surface-sterilized and sown
683 on Murashige and Skoog (MS) agar plates. The plates were incubated at 4°C for 2 days for
684 stratification and subsequently transferred to a growth chamber under long-day conditions.
685 Seven day old seedlings were immersed in GUS staining solution (XGLUC; 14-mM NaH2PO4,
686 36-mM Na2HPO4, 2-mM K4[Fe(CN)6], 2-M K3[Fe(CN)6], 10% Triton X-100). Seedlings were
687 vacuum infiltrated for 5 minutes, incubated overnight at 37 °C in the dark, and rinsed with PBS
688 to remove excess staining solution and dehydrated through an ethanol series (30%, 50%,
689 70%, 90%, and 100%) to remove chlorophyll and enhance contrast.. Pictures of seedlings

690 were taken with a canon EOS 450D camera on a Nikon Eclipse 80i Fluorescence microscope.
691 White balance and scalebar were added via ImageJ.
692

693 **Subcellular localization and FRET-FLIM analysis**

694 As above, the coding sequences of *ATHB2, ATHB2miP* and *ATHB2HD* (from exon 1 to the
695 beginning of exon 4) were cloned into the vectors pK7WGF2 that contains an N-terminal eGFP
696 driven by a CaMV35S promoter and a modified vector of pEarlyGate104 containing an N-
697 terminal mCherry driven by a CaMV35S promoter (provided by Sabine Müller/Dorothee
698 Stöckle, ZMBP Tübingen). Tobacco leaves were transiently transformed by agrobacterium
699 infiltration (three individual plants per infiltration) and nuclei of the epidermal cells of leaves
700 were imaged 2-3 days post-infiltration using a Leica Stellaris 8 confocal laser-scanning
701 microscope. All vectors were co-transformed with p19 and for the FRET-FLIM assay, eGFP-
702 ATHB2 was additionally co-transformed with either mCherry, mCherry-ATHB2, or mCherry-
703 ATHB2miP. Fluorescence lifetime measurements were collected using the TauInteraction tool
704 in the LAS X software; samples were excited with a 470 nm pulsed laser (10 MHz) and
705 emission from 500 nm to 560 nm was recorded.
706

707 **Differential Gene Expression Analysis and Clustering**

708 Total RNA was extracted from Arabidopsis thaliana Col-0 wild-type (WT), t-athb2, and
709 35S::ATHB2miP genotypes under three different light conditions as described: White Light
710 (WL), 45 minutes of Far-Red (FR) enrichment, and 90 minutes of FR enrichment. Three
711 biological replicates per genotype-condition combination were used. RNA quality and
712 concentration were assessed using a NanoDrop spectrophotometer and Agilent Bioanalyzer.
713 RNA was sequenced on an Illumina HiSeq platform.
714 Raw sequencing reads were quality-checked using FastQC to ensure data integrity. Trimming
715 of reads to remove adapter sequences and low-quality bases was performed using
716 Trimmomatic. After preprocessing, high-quality reads were retained for downstream analysis.
717 Read alignment to the Arabidopsis reference genome (TAIR10) was conducted using the
718 STAR aligner within the R environment. This generated BAM files for each sample, which were
719 subsequently used for quantification of gene expression. Differential gene expression analysis
720 was carried out using the edgeR package in R. Raw counts were normalized using the TMM
721 (Trimmed Mean of M-values) method. Genes with low expression were filtered out, and
722 statistical analysis was performed to identify differentially expressed genes (DEGs) between
723 genotypes and conditions. P-values were adjusted for multiple testing using the false
724 discovery rate (FDR) method. DEGs were subjected to gene ontology (GO) enrichment
725 analysis to gain insights into biological processes and pathways associated with the observed
726 expression changes. The topGO package in R was used for this purpose.

727 The raw counts obtained from the RNA-seq analysis were used as input for the MFuzz
728 clustering. These count data represented the expression levels of genes across the three
729 genotypes and three light conditions.
730 For the clustering, the count data were first normalized (TMM method) to account for
731 differences in sequencing depth and library size. MFuzz, an R package specifically designed
732 for model-based clustering of high-dimensional gene expression data, was utilized for
733 clustering analysis. The resulting 17 gene clusters were visually inspected to find the ones
734 with the wished expression patterns across the samples. Genes belongind to clusters of
735 interest were extracted and subjected to gene ontology (GO) enrichment analysis.
736 The reads were visualised on IGV (Integrative Genome Viewer) from the BAM mapping files
737 generated.
738
739 **Preparation of samples and iron concentration measurements**
740 WT, t-athb2 and 35S::ATHB2miP seedlings (about 200 per genotype) were grown on standard
741 MS medium for 12 days. They were carefully removed from the medium to avoid damage and
742 the aerial part of each seedling was excised from the roots. In all steps, no iron nor glass tools
743 were used. The roots and shoots were rinsed in a 2 mM $CaCl_2$ solution to remove iron
744 accumulated on the surface. The material was placed in an Eppendorf tube and dried in a
745 stove at 70°C for 3 days. Each of the dry samples was weighed (dry root weights ranged from
746 2 to 6 mg, seedlings from 15 to 25 mg). Samples were digested in 2.5 mL of 70% HNO3 + 1
747 mL of H2O2. Digestion conditions: maximum temperature = 240°C; maximum pressure = 200
748 bar; cycle duration: 1-2 hours. Following digestion, the samples were diluted in ultrapure Milli-
749 Q water to achieve a [HNO3] = 3.5%. The liquid samples were then analyzed using an ICP-
750 QQQ-MS 8800 Agilent instrument. NIST certified reference materials were employed to verify
751 the accuracy of the data.
752

753 ## Figure legends

754 **Figure 1. An alternative transcription start site in the third exon of ATHB2 gives rise to**
755 **a shade sensitive microProtein. (A)** 5'PEAT-seq reads map to 12,398 unique TSSs, 43%
756 representing novel TSSs and 1.5% being candidates for alternative microProteins. **(B)** 5'PEAT-
757 seq reads mapping to *ATHB2*. **(C)** Iso-seq reads mapping to *ATHB2*. **(D)** Design of constructs
758 with GUS fused to either the promoter of *ATHB2* (*pATHB2::GUS*), the sequence upstream of
759 an ATG in the second intron of *ATHB2* (*pATHB2control::GUS*; negative control), or the
760 sequence upstream of a near-cognate start codon TTG in the third exon of *ATHB2*
761 (*pATHB2miP::GUS*). **(E)** GUS staining in *pATHB2::GUS* accumulates in the hypocotyl, SAM
762 and cotyledon vasculature. **(F)** GUS staining in *pATHB2control::GUS* does not produce any

763 GUS signal. **(G)** GUS staining in *pATHB2miP::GUS* plants accumulates in the SAM and
764 hypocotyl. Scale bars = 200 μM.
765
766 **Figure 2. ATHB2 and ATHB2miP localise to the nucleus and interact in vitro and *in*
767 *planta*. (A)** Graphic representation of eGFP fused to ATHB2 variants: ATHB2, ATHB2miP, and
768 ATHB2HD; panels show subcellular localisation of ATHB2 variants when fused to eGFP in
769 tobacco leaves, with nuclear localisation shown in square and enlarged in lower panel. Arrows
770 point to nuclear speckles that form when eGFP-ATHB2miP is expressed. Scale bars: upper
771 panel = 50 μM; lower panel = 5 μM. **(B)** Yeast-2-hybrid with growth on histidine drop-out media
772 when ATHB2 or ATHB2miP are fused to the activation domain and binding domain of GAL4.
773 **(C)** FRET-FLIM assay. Co-expression of eGFP-ATHB2 with mCherry-ATHB2 or mCherry-
774 ATHB2miP in tobacco leaves shows co-localization in the nucleus and causes significant
775 reduction in fluorescent lifetime of eGFP-ATHB2. Scale bar = 5 μM. Significance levels: * =
776 $p<0.05$, ** = $p<0.01$, *** = $p<0.001$, ns = not significant.
777
778 **Figure 3. Gene expression profiling of the *t-athb2* and *35S::miP lines*. (A)** Heatmap
779 showing expression changes in 95 genes whose expression depends on both the genotype
780 and the shade treatment. Lowly-expressed genes are in blue, highly-expressed genes are in
781 yellow. **(B)** 3 of the 17 clusters generated with MFuzz. Expression changes of the genes
782 included in each cluster are on the y-axis. The 3 genotypes and shade conditions are on the
783 x-axis. **(C, D, E)** Venn diagrams showing the number of genes that are upregulated and
784 downregulated in each of the three genotypes included in the RNA-Seq and the overlap
785 between them in WL, and 45 and 90 mins of WL+FR.
786
787 **Figure 4. ATHB2 and ATHB2mip mediate hypocotyl elongation and root growth. (A)**
788 Graphic representation of the gene structure of *ATHB2*. Location of T-DNA insertion
789 (SALK_106790) is shown and below that the predicted protein structure of WT, transgenic and
790 mutant isoforms of *ATHB2*. Blue and red triangles indicate Cas9 targeted sites. **(B)** RT-qPCR
791 of *ATHB2* in mutant and transgenic plants with and without WL+FR treatment. Coloured
792 regions correspond to those in part A. **(C)** Plots of hypocotyl length in plants after 5 days
793 growth in three different shade regimes: deep shade, canopy shade, and proximity shade. **(D)**
794 Plot of the number of lateral roots in plants after 11 days of growth in WL or WL+FR (proximity
795 shade). Letters in all signify results of pairwise comparisons by two-way ANOVA.
796
797 **Figure 5. ATHB2 and ATHB2miP have a role in iron transport regulation and
798 homeostasis. (A)** Heatmap showing the expression levels in white light of genes upregulated
799 in both the *t-athb2* mutant and the *35S::miP* lines that are involved in iron transport, iron

homeostasis and iron deficiency response. Lowly expressed genes are coloured blue, highly expressed genes are coloured red. Additionally, the expression level for all three replicates per genotype is displayed. **(B)** Heatmap showing the gene expression levels in shade conditions. **(C)** Concentration of Fe (mg/g) measured in the roots (grey) and the shoots of seedlings of the three genotypes (green). **(D)** Plots of hypocotyl length (red) and root length (blue) of plants grown on low or high iron supplemented ½ MS media in WL and WL+FR. Hypocotyls were measured after five days in either condition whilst roots were measured after nine days.

Supplementary figures

Figure S1. Identification of an alternative ORF in the coding sequence of *ATHB2*. (A) 5'RACE carried out on 10 day old seedlings that were exposed to 0, 45 and 90 mins of WL+FR light. Red asterisk signifies the small transcript arising from *ATHB2* that when purified and sequenced corresponds to the underlined portion of *ATHB2* in part B. **(B)** Genomic sequence of *ATHB2*. The exonic sequence with protein sequence below is shown in bold. Double-sided blue arrows denote sites corresponding to 5'PEAT-seq read clusters. Blue highlighted sequence denotes the leucine zipper domain. Red circles indicate two alternative non-canonical start codons identified by TIS predictor. Underlined sequence corresponds to the transcript purified from 5'RACE in part A.

Figure S2. RNA-seq reads corresponding to ATHB2 at 0, 45 and 90 mins of WL+FR. RNA-seq reads peak in the WT at 45 mins of WL+FR and then decrease (see y axis for maximum reads) at 90 mins. No reads are detected in t-athb2 mutants whereas 35S::miP mutant plants are highly overexpressing exon 3 and 4 of ATHB2.

Figure S3. (A) GO enrichment analysis performed on the 95 DEGs whose expression is shown in Figure 3 yielded several significant GO terms (FDR < 0.05). These GO terms are related to shade avoidance, hormone/auxin responses, and growth. **(B)** The 17 clusters generated by MFuzz. The expression changes of the genes included in each cluster are on the y-axis. The three genotypes and shade conditions are on the x-axis.

Figure S4. ATH2B and ATHB2miP slightly effect root elongation. Main root length of 11-day old plants grown in either WL or WL+FR in three different shade regimes: deep shade, canopy shade and proximity shade.

Acknowledgements

We thank Jaume Martinez-Garcia for comments on the manuscript.

Additional information

Funding

We acknowledge funding through NovoCrops Centre (Novo Nordisk Foundation project number 2019OC53580 to S. W.), the Independent Research Fund Denmark (0136- 00015B and 0135-00014B to S. W.) and the Novo Nordisk Foundation (NNF18OC0034226 and NNF20OC0061440 to S. W.).

Author contribution

Conceptualization: Stephan Wenkel.

Data curation: Daniel Straub, Maurizio J. Chiurazzi.

Formal analysis: Ashleigh Edwards, Maurizio Junior Chiurazzi, Anko Blaakmeer, Ylenia Vittozzi, Ashish Sharma, Sanne Matton, Valdeko Kruusvee, Daniel Straub, Giovanna Sessa, Monica Carabelli, Giorgio Morelli, Stephan Wenkel

Funding acquisition: Stephan Wenkel.

Investigation: Ashleigh Edwards, Maurizio Junior Chiurazzi, Anko Blaakmeer, Ylenia Vittozzi, Ashish Sharma, Sanne Matton, Valdeko Kruusvee, Daniel Straub, Giovanna Sessa, Monica Carabelli

Methodology: Daniel Straub, Maurizio Junior Chiurazzi, Valdeko Kruusvee

Supervision: Stephan Wenkel.

Validation: Daniel Straub, Valdeko Kruusvee, Giovanna Sessa, Monica Carabelli, Giorgio Morelli

Writing – original draft: Ashleigh Edwards, Maurizio Junior Chiurazzi, Anko Blaakmeer

Writing – review & editing: Giovanna Sessa, Monica Carabelli, Giorgio Morelli, Stephan Wenkel.

Author ORCIDs

Daniel Straub: 0000-0002-2553-0660

Anko Blaakmeer: 0000-0002-1834-2573

Stephan Wenkel: 0000-0001-5764-9423

Additional files

Supplementary dataset 1. Alternative transcripts encoding potential microProteins discovered by 5'PEAT seq.

Supplementary dataset 2. Differentially expressed genes in wild type, t-athb2 and 35S::miP seedlings exposed to 45 and 90 minutes of shade.

Supplementary dataset 3. Gene clusters identified by RNA-seq.

Data availability

All sequencing data have been deposited in NCBI's Gene Expression Omnibus under GEO Series accession no. GSE241976, comprising RNA-seq (GSE241974) and 5'PEAT-seq (GSE241975).

References

Beadle GW, Tatum EL (1941) Genetic Control of Biochemical Reactions in Neurospora. *Proc Natl Acad Sci U S A* 27: 499-506

Berget SM, Moore C, Sharp PA (1977) Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci U S A* 74: 3171-3175

Bhati KK, Dolde U, Wenkel S (2021) MicroProteins: Expanding functions and novel modes of regulation. *Mol Plant* 14: 705-707

Botterweg-Paredes E, Blaakmeer A, Hong SY, Sun B, Mineri L, Kruusvee V, Xie Y, Straub D, Ménard D, Pesquet E, Wenkel S (2020) Light affects tissue patterning of the hypocotyl in the shade-avoidance response. *PLoS Genet* 16: e1008678

Bou-Torrent J, Salla-Martret M, Brandt R, Musielak T, Palauqui JC, Martinez-Garcia JF, Wenkel S (2012) ATHB4 and HAT3, two class II HD-ZIP transcription factors, control leaf development in Arabidopsis. *Plant Signal Behav* 7: 1382-1387

Brandt R, Salla-Martret M, Bou-Torrent J, Musielak T, Stahl M, Lanz C, Ott F, Schmid M, Greb T, Schwarz M *et al* (2012) Genome-wide binding-site analysis of REVOLUTA reveals a link between leaf patterning and light-mediated growth responses. *Plant J* 72: 31-42

Brandt R, Xie Y, Musielak T, Graeff M, Stierhof YD, Huang H, Liu CM, Wenkel S (2013) Control of stem cell homeostasis via interlocking microRNA and microProtein feedback loops. *Mech Dev* 130: 25-33

Cao M, Platre MP, Tsai HH, Zhang L, Nobori T, Armengot L, Chen Y, He W, Brent L, Coll NS *et al* (2024) Spatial IMA1 regulation restricts root iron acquisition on MAMP perception. *Nature* 625: 750-759

Carabelli M, Possenti M, Sessa G, Ruzza V, Morelli G, Ruberti I (2018) Arabidopsis HD-Zip II proteins regulate the exit from proliferation during leaf development in canopy shade. *J Exp Bot* 69: 5419-5431

Carabelli M, Sessa G, Baima S, Morelli G, Ruberti I (1993) The Arabidopsis Athb-2 and -4 genes are strongly induced by far-red-rich light. *Plant J* 4: 469-479

Carabelli M, Turchi L, Ruzza V, Morelli G, Ruberti I (2013) Homeodomain-Leucine Zipper II family of transcription factors to the limelight: central regulators of plant development. *Plant Signal Behav* 8

Chong C, Müller M, Pak H, Harnett D, Huber F, Grun D, Leleu M, Auger A, Arnaud M, Stevenson BJ *et al* (2020) Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes. *Nat Commun* 11: 1293

Ciarbelli AR, Ciolfi A, Salvucci S, Ruzza V, Possenti M, Carabelli M, Fruscalzo A, Sessa G, Morelli G, Ruberti I (2008) The Arabidopsis homeodomain-leucine zipper II gene family: diversity and redundancy. *Plant Mol Biol* 68: 465-478

Colangelo EP, Guerinot ML (2004) The essential basic helix-loop-helix protein FIT1 is required for the iron deficiency response. *Plant Cell* 16: 3400-3412

915 de Wit M, Spoel SH, Sanchez-Perez GF, Gommers CMM, Pieterse CMJ, Voesenek L, Pierik R (2013) Perception
916 of low red:far-red ratio compromises both salicylic acid- and jasmonic acid-dependent pathogen defences in
917 Arabidopsis. *Plant J* 75: 90-103

918 Eguen T, Straub D, Graeff M, Wenkel S (2015) MicroProteins: small size-big impact. *Trends Plant Sci* 20: 477-
919 482

920 Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-
921 Vegas A *et al* (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*
922 44: D279-285

923 Finn RD, Miller BL, Clements J, Bateman A (2014) iPfam: a database of protein family and domain interactions
924 found in the Protein Data Bank. *Nucleic Acids Res* 42: D364-373

925 Floyd SK, Ryan JG, Conway SJ, Brenner E, Burris KP, Burris JN, Chen T, Edger PP, Graham SW, Leebens-Mack
926 JH *et al* (2014) Origin of a novel regulatory module by duplication and degeneration of an ancient plant
927 transcription factor. *Mol Phylogenet Evol* 81: 159-173

928 Gallemi M, Molina-Contreras MJ, Paulisic S, Salla-Martret M, Sorin C, Godoy M, Franco-Zorrilla JM, Solano R,
929 Martinez-Garcia JF (2017) A non-DNA-binding activity for the ATHB4 transcription factor in the control of
930 vegetation proximity. *New Phytol* 216: 798-813

931 Gleason AC, Ghadge G, Chen J, Sonobe Y, Roos RP (2022) Machine learning predicts translation initiation sites
932 in neurologic diseases with nucleotide repeat expansions. *PLoS One* 17: e0256411

933 Goyal A, Karayekov E, Galvão VC, Ren H, Casal JJ, Fankhauser C (2016) Shade Promotes Phototropism
934 through Phytochrome B-Controlled Auxin Production. *Curr Biol* 26: 3280-3287

935 Halliday KJ, Koornneef M, Whitelam GC (1994) Phytochrome B and at Least One Other Phytochrome Mediate
936 the Accelerated Flowering Response of Arabidopsis thaliana L. to Low Red/Far-Red Ratio. *Plant Physiol* 104:
937 1311-1315

938 Hanada K, Higuchi-Takeuchi M, Okamoto M, Yoshizumi T, Shimizu M, Nakaminami K, Nishi R, Ohashi C, Iida K,
939 Tanaka M *et al* (2013) Small open reading frames associated with morphogenesis are hidden in plant genomes.
940 *Proc Natl Acad Sci U S A* 110: 2395-2400

941 Ingolia NT (2014) Ribosome profiling: new views of translation, from single codons to genome scale. *Nat Rev*
942 *Genet* 15: 205-213

943 Kim C, Kwon Y, Jeong J, Kang M, Lee GS, Moon JH, Lee HJ, Park YI, Choi G (2023) Phytochrome B
944 photobodies are comprised of phytochrome B and its primary and secondary interacting proteins. *Nat Commun*
945 14: 1708

946 Kim Y-S, Kim S-G, Lee M, Lee I, Park H-Y, Seo PJ, Jung J-H, Kwon E-J, Suh SW, Paek K-H, Park C-M (2008)
947 HD-ZIP III Activity Is Modulated by Competitive Inhibitors via a Feedback Loop in Arabidopsis Shoot Apical
948 Meristem Development. *Plant Cell* 20: 920-933

949 Kircher S, Gil P, Kozma-Bognár L, Fejes E, Speth V, Husselstein-Muller T, Bauer D, Adám E, Schäfer E, Nagy F
950 (2002) Nucleocytoplasmic partitioning of the plant photoreceptors phytochrome A, B, C, D, and E is regulated
951 differentially by light and exhibits a diurnal rhythm. *Plant Cell* 14: 1541-1555

952 Kondo T, Plaza S, Zanet J, Benrabah E, Valenti P, Hashimoto Y, Kobayashi S, Payre F, Kageyama Y (2010)
953 Small peptides switch the transcriptional activity of Shavenbaby during Drosophila embryogenesis. *Science* 329:
954 336-339

955 Kumar L, Futschik E (2007) Mfuzz: a software package for soft clustering of microarray data. *Bioinformation* 2: 5-
956 7

957 Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-
958 Hernandez M *et al* (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new
959 tools. *Nucleic Acids Res* 40: D1202-1210

960 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The
961 Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078-2079

962 Lorrain S, Allen T, Duek PD, Whitelam GC, Fankhauser C (2008) Phytochrome-mediated inhibition of shade
963 avoidance involves degradation of growth-promoting bHLH transcription factors. *Plant J* 53: 312-323

964 Luo Z, Freitag M, Sachs MS (1995) Translational regulation in response to changes in amino acid availability in
965 Neurospora crassa. *Mol Cell Biol* 15: 5235-5245

966 Mallory AC, Reinhart BJ, Jones-Rhoades MW, Tang G, Zamore PD, Barton MK, Bartel DP (2004) MicroRNA
967 control of PHABULOSA in leaf development: importance of pairing to the microRNA 5[prime] region. *EMBO J* 23:
968 3356-3364

969 Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* 17: 3

970 Martinez-Garcia JF, Rodriguez-Concepcion M (2023) Molecular mechanisms of shade tolerance in plants. *New*
971 *Phytol* 239: 1190-1202

972 Merelo P, Ram H, Pia Caggiano M, Ohno C, Ott F, Straub D, Graeff M, Cho SK, Yang SW, Wenkel S, Heisler MG
973 (2016) Regulation of MIR165/166 by class II and class III homeodomain leucine zipper proteins establishes leaf
974 polarity. *Proc Natl Acad Sci U S A* 113: 11973-11978

975 Mudge JM, Ruiz-Orera J, Prensner JR, Brunet MA, Calvet F, Jungreis I, Gonzalez JM, Magrane M, Martinez TF,
976 Schulz JF *et al* (2022) Standardized annotation of translated open reading frames. *Nat Biotechnol* 40: 994-999

977 Müller-Moulé P, Nozue K, Pytlak ML, Palmer CM, Covington MF, Wallace AD, Harmer SL, Maloof JN (2016)
978 YUCCA auxin biosynthetic genes are required for Arabidopsis shade avoidance. *PeerJ* 4: e2574

979 Ni T, Corcoran DL, Rach EA, Song S, Spana EP, Gao Y, Ohler U, Zhu J (2010) A paired-end sequencing strategy
980 to map the complex landscape of transcription initiation. *Nat Methods* 7: 521-527

981 Ohgishi M, Oka A, Morelli G, Ruberti I, Aoyama T (2001) Negative autoregulation of the Arabidopsis homeobox
982 gene ATHB-2. *Plant J* 25: 389-398

983 Orbach MJ, Sachs MS, Yanofsky C (1990) The Neurospora crassa arg-2 locus. Structure and expression of the
984 gene encoding the small subunit of arginine-specific carbamoyl phosphate synthetase. *J Biol Chem* 265: 10981-
985 10987

986 Pedmale UV, Huang SC, Zander M, Cole BJ, Hetzel J, Ljung K, Reis PAB, Sridevi P, Nito K, Nery JR *et al* (2016)
987 Cryptochromes Interact Directly with PIFs to Control Plant Growth in Limiting Blue Light. *Cell* 164: 233-245

988 Quail PH (2002) Phytochrome photosensory signalling networks. *Nat Rev Mol Cell Biol* 3: 85-93

989 Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*
990 26: 841-842

991 Ruiz-Orera J, Villanueva-Cañas JL, Albà MM (2020) Evolution of new proteins from translated sORFs in long
992 non-coding RNAs. *Exp Cell Res* 391: 111940

993 Schena M, Lloyd AM, Davis RW (1993) The HAT4 gene of Arabidopsis encodes a developmental regulator.
994 *Genes Dev* 7: 367-379

995 Schon MA, Kellner MJ, Plotnikova A, Hofmann F, Nodine MD (2018) NanoPARE: parallel analysis of RNA 5' ends
996 from low-input RNA. *Genome Res* 28: 1931-1942

997 Sessa G, Morelli G, Ruberti I (1993) The Athb-1 and -2 HD-Zip domains homodimerize forming complexes of
998 different DNA binding specificities. *Embo j* 12: 3507-3517

999 Sorin C, Salla-Martret M, Bou-Torrent J, Roig-Villanova I, Martinez-Garcia JF (2009) ATHB4, a regulator of shade
1000 avoidance, modulates hormone response in Arabidopsis seedlings. *Plant J* 59: 266-277

1001 Sperschneider J, Catanzariti AM, DeBoer K, Petre B, Gardiner DM, Singh KB, Dodds PN, Taylor JM (2017)
1002 LOCALIZER: subcellular localization prediction of both plant and effector proteins in the plant cell. *Sci Rep* 7:
1003 44598

1004 Spielmann J, Fanara S, Cotelle V, Vert G (2023) Multilayered regulation of iron homeostasis in Arabidopsis. *Front*
1005 *Plant Sci* 14: 1250588

1006 Srivastava AK, Lu Y, Zinta G, Lang Z, Zhu JK (2018) UTR-Dependent Control of Gene Expression in Plants.
1007 *Trends Plant Sci* 23: 248-259

1008 Steindler C, Matteucci A, Sessa G, Weimar T, Ohgishi M, Aoyama T, Morelli G, Ruberti I (1999) Shade avoidance
1009 responses are mediated by the ATHB-2 HD-zip protein, a negative regulator of gene expression. *Development*
1010 126: 4235-4245

1011 Svoboda KK, Nishimura I, Sugrue SP, Ninomiya Y, Olsen BR (1988) Embryonic chicken cornea and cartilage
1012 synthesize type IX collagen molecules with different amino-terminal domains. *Proc Natl Acad Sci U S A* 85: 7496-
1013 7500

1014 Tao Y, Ferrer JL, Ljung K, Pojer F, Hong F, Long JA, Li L, Moreno JE, Bowman ME, Ivans LJ *et al* (2008) Rapid
1015 synthesis of auxin via a new tryptophan-dependent pathway is required for shade avoidance in plants. *Cell* 133:
1016 164-176

1017 Turchi L, Baima S, Morelli G, Ruberti I (2015) Interplay of HD-Zip II and III transcription factors in auxin-regulated
1018 plant development. *J Exp Bot* 66: 5043-5053

1019 van Gelderen K, Kang C, Pierik R (2018) Light Signaling, Root Development, and Plasticity. *Plant Physiol* 176:
1020 1049-1060

1021 Wenkel S, Emery J, Hou B-H, Evans MMS, Barton MK (2007) A Feedback Regulatory Module Formed by LITTLE
1022 ZIPPER and HD-ZIPIII Genes. *Plant Cell* 19: 3379-3390

1023 Won C, Shen X, Mashiguchi K, Zheng Z, Dai X, Cheng Y, Kasahara H, Kamiya Y, Chory J, Zhao Y (2011)
1024 Conversion of tryptophan to indole-3-acetic acid by TRYPTOPHAN AMINOTRANSFERASES OF ARABIDOPSIS
1025 and YUCCAs in Arabidopsis. *Proc Natl Acad Sci U S A* 108: 18518-18523

1026 Zhang S, Reljić B, Liang C, Kerouanton B, Francisco JC, Peh JH, Mary C, Jagannathan NS, Olexiouk V, Tang C
1027 *et al* (2020) Mitochondrial peptide BRAWNIN is essential for vertebrate respiratory complex III assembly. *Nat*
1028 *Commun* 11: 1312
1029

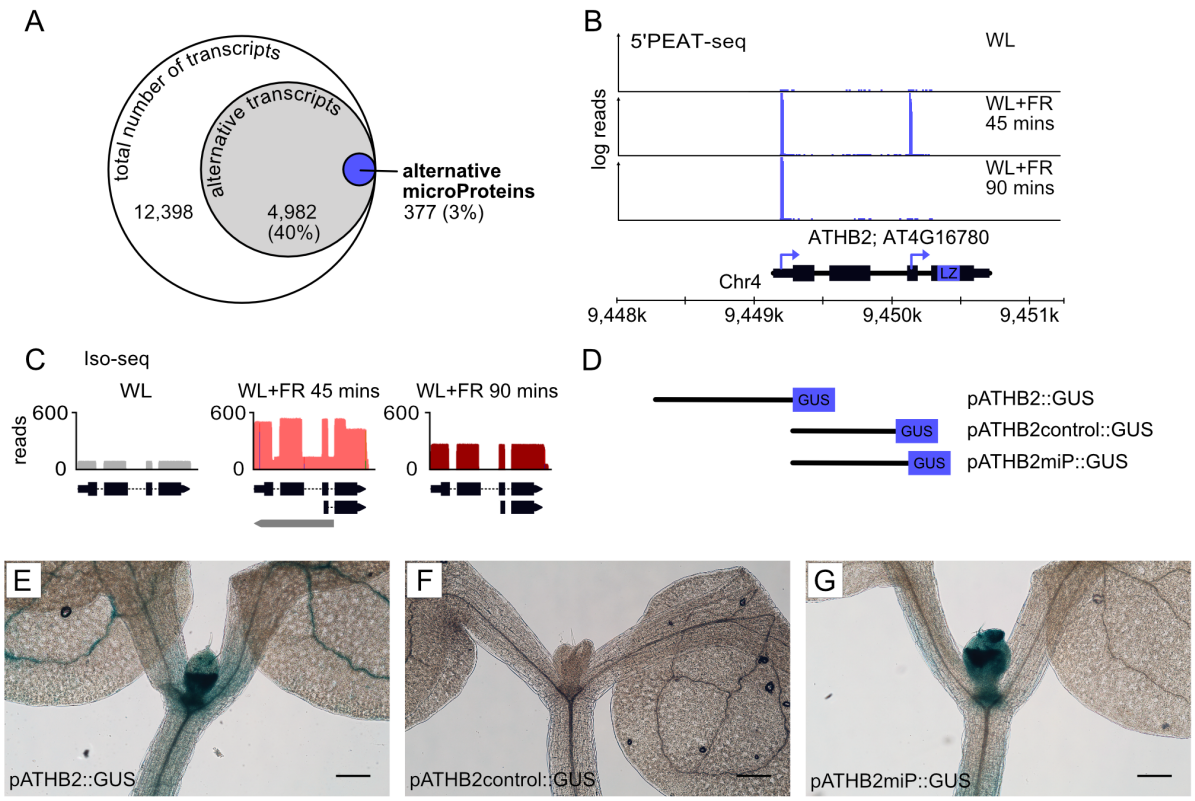

Figure 1. An alternative transcription start site in the third exon of *ATHB2* gives rise to a shade sensitive microProtein. (A) 5'PEAT-seq reads map to 12,398 unique TSSs, 43% representing novel TSSs and 1.5% being candidates for alternative microProteins. **(B)** 5'PEAT-seq reads mapping to *ATHB2*. **(C)** Iso-seq reads mapping to *ATHB2*. **(D)** Design of constructs with GUS fused to either the promoter of *ATHB2* (*pATHB2::GUS*), the sequence upstream of an ATG in the second intron of *ATHB2* (*pATHB2control::GUS*; negative control), or the sequence upstream of a near-cognate start codon TTG in the third exon of *ATHB2* (*pATHB2miP::GUS*). **(E)** GUS staining in *pATHB2::GUS* accumulates in the hypocotyl, SAM and cotyledon vasculature. **(F)** GUS staining in *pATHB2control::GUS* does not produce any GUS signal. **(G)** GUS staining in *pATHB2miP::GUS* plants accumulates in the SAM and hypocotyl. Scale bars = 200 µM.

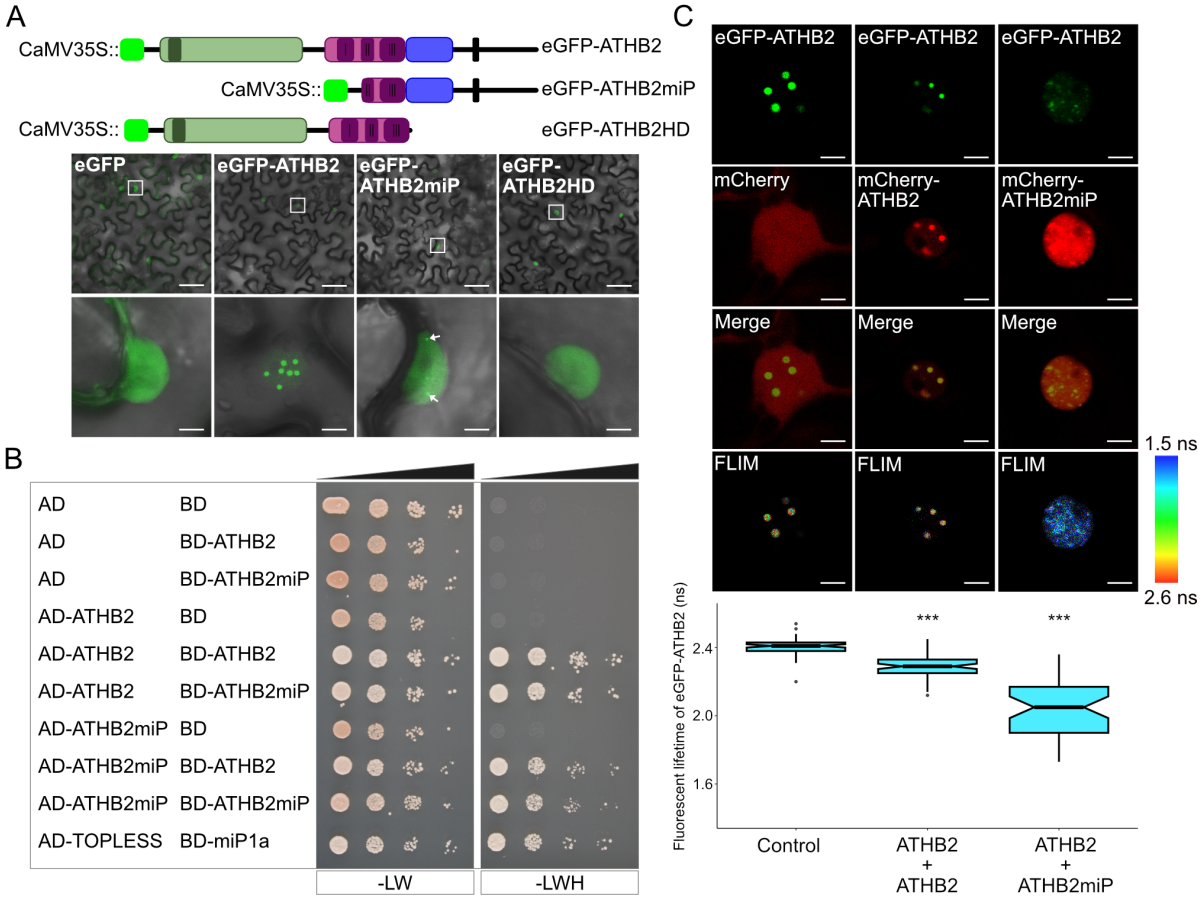

Figure 2. ATHB2 and ATHB2miP localise to the nucleus and interact *in vitro* and *in planta*. (A) Graphic representation of eGFP fused to ATHB2 variants: ATHB2, ATHB2miP, and ATHB2HD; panels show subcellular localisation of ATHB2 variants when fused to eGFP in tobacco leaves, with nuclear localisation shown in square and enlarged in lower panel. Arrows point to nuclear speckles that form when eGFP-ATHB2miP is expressed. Scale bars: upper panel = 50 µM; lower panel = 5 µM. **(B)** Yeast-2-hybrid with growth on histidine drop-out media when ATHB2 or ATHB2miP are fused to the activation domain and binding domain of GAL4. **(C)** FRET-FLIM assay. Co-expression of eGFP-ATHB2 with mCherry-ATHB2 or mCherry-ATHB2miP in tobacco leaves shows co-localization in the nucleus and causes significant reduction in fluorescent lifetime of eGFP-ATHB2. Scale bar = 5 µM. Significance levels: * = $p<0.05$, ** = $p<0.01$, *** = $p<0.001$, ns = not significant.

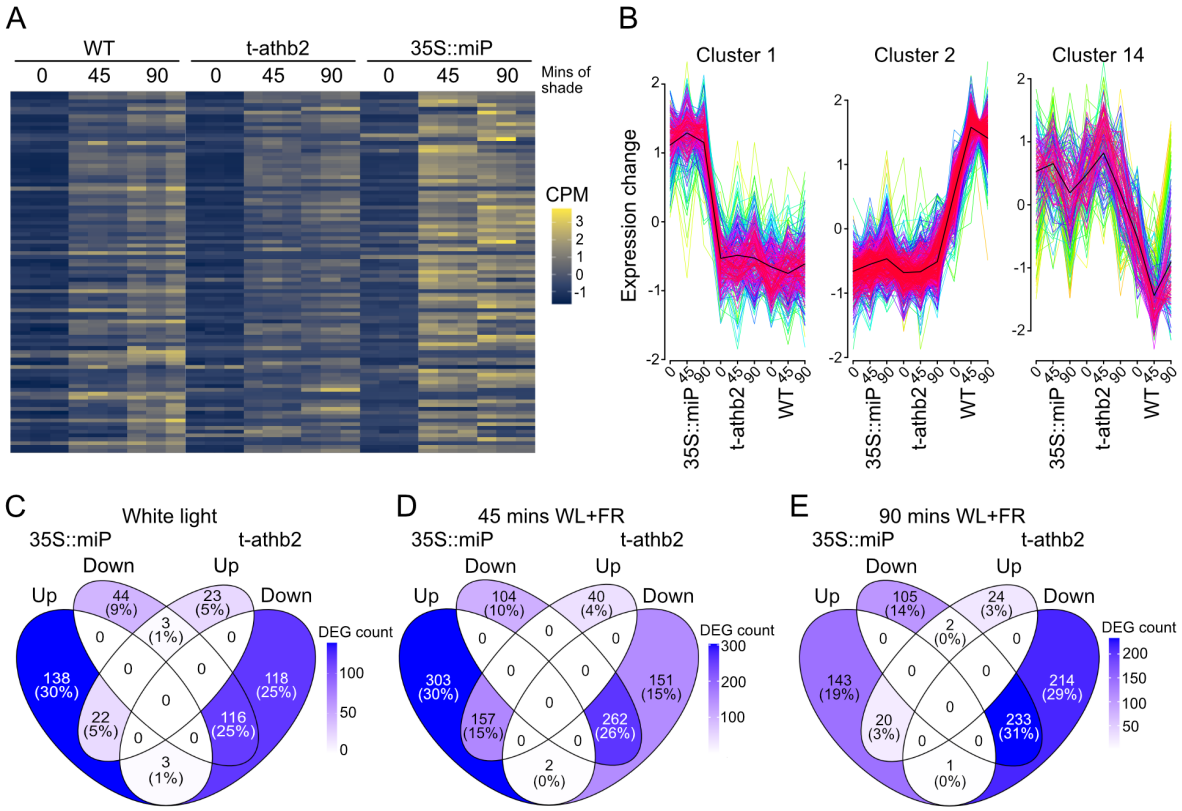

Figure 3. Gene expression profiling of the *t-athb2* and *35S::miP* lines. (A) Heatmap showing expression changes in 95 genes whose expression depends on both the genotype and the shade treatment. Lowly-expressed genes are in blue, highly-expressed genes are in yellow. **(B)** 3 of the 17 clusters generated with MFuzz. Expression changes of the genes included in each cluster are on the y-axis. The 3 genotypes and shade conditions are on the x-axis. (**C, D, E**) Venn diagrams showing the number of genes that are upregulated and downregulated in each of the three genotypes included in the RNA-Seq and the overlap between them in WL, and 45 and 90 mins of WL+FR.

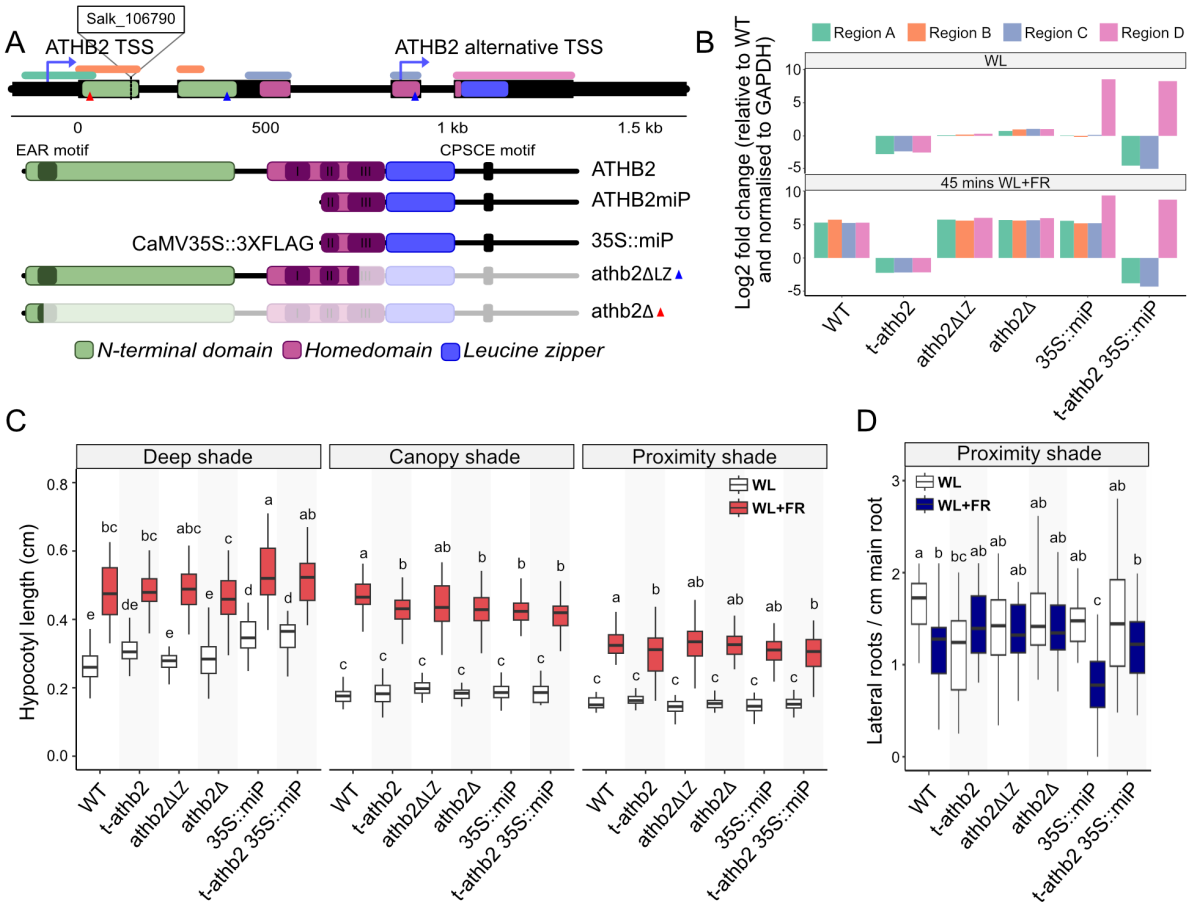

Figure 4. ATHB2 and ATHB2mip mediate hypocotyl elongation and root growth. (A) Graphic representation of the gene structure of *ATHB2*. Location of T-DNA insertion (SALK_106790) is shown and below that the predicted protein structure of WT, transgenic and mutant isoforms of *ATHB2*. Blue and red triangles indicate Cas9 targeted sites. **(B)** RT-qPCR of *ATHB2* in mutant and transgenic plants with and without WL+FR treatment. Coloured regions correspond to those in part A. **(C)** Plots of hypocotyl length in plants after 5 days growth in three different shade regimes: deep shade, canopy shade, and proximity shade. **(D)** Plot of the number of lateral roots in plants after 11 days of growth in WL or WL+FR (proximity shade). Letters in all signify results of pairwise comparisons by two-way ANOVA.

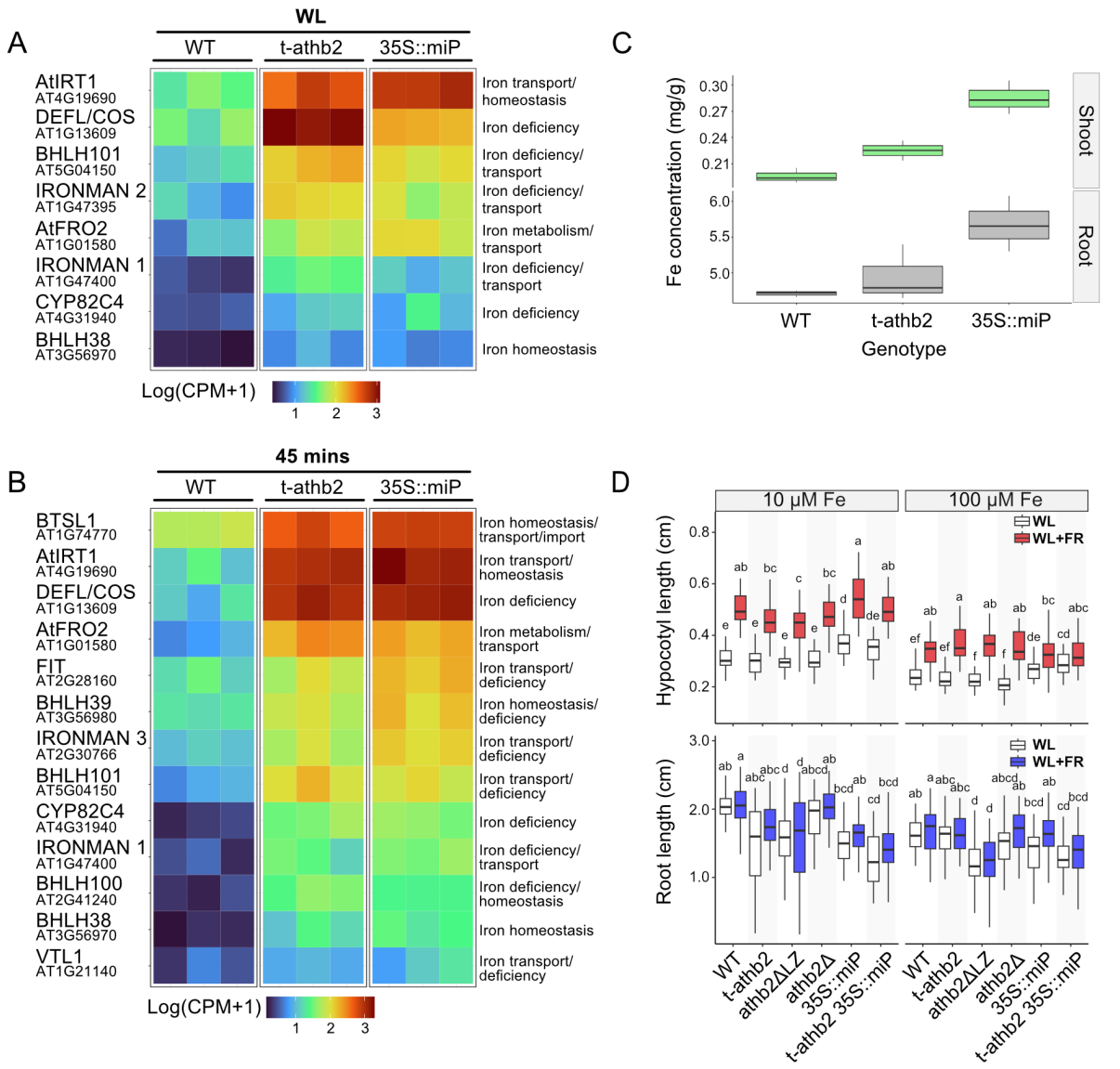

Figure 5. ATHB2 and ATHB2miP have a role in iron transport regulation and homeostasis. (A) Heatmap showing the expression levels in white light of genes upregulated in both the *t-athb2* mutant and the *35S::miP* lines that are involved in iron transport, iron homeostasis and iron deficiency response. Lowly expressed genes are coloured blue, highly expressed genes are coloured red. Additionally, the expression level for all three replicates per genotype is displayed. **(B)** Heatmap showing the gene expression levels in shade conditions **(C)** Concentration of Fe (mg/g) measured in the roots (grey) and the shoots of seedlings of the three genotypes (green). **(D)** Plots of hypocotyl length (red) and root length (blue) of plants grown on low or high iron supplemented ½ MS media in WL and WL+FR. Hypocotyls were measured after five days in either condition whilst roots were measured after nine days.

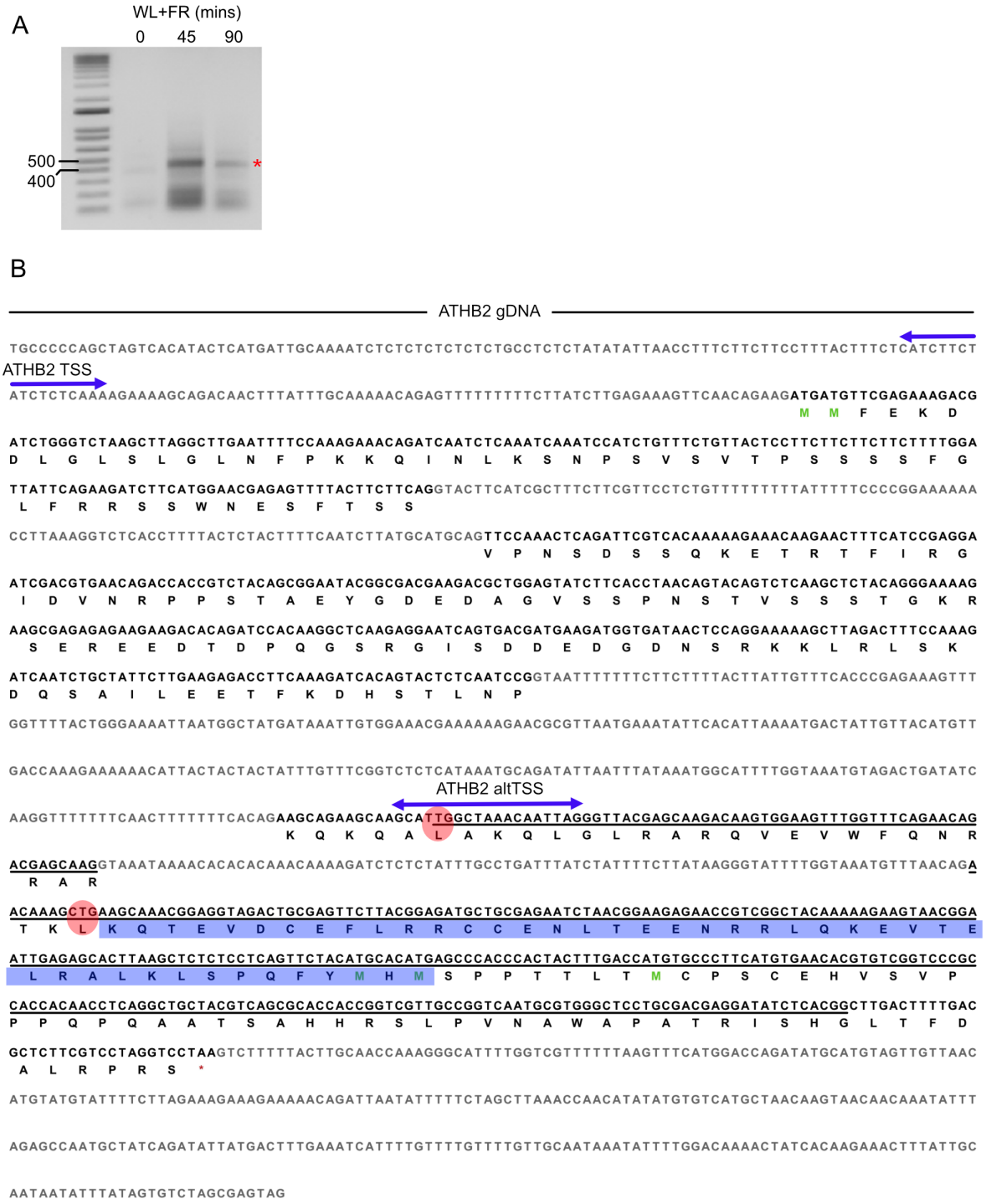

Figure S1. Identification of an alternative ORF in the coding sequence of ATHB2. (A) 5'RACE carried out on 10 day old seedlings that were exposed to 0, 45 and 90 mins of WL+FR light. Red asterisk signifies the small transcript arising from *ATHB2* that when purified and sequenced corresponds to the underlined portion of *ATHB2* in part B. **(B)** Genomic sequence of *ATHB2*. The exonic sequence with protein sequence below is shown in bold. Double-sided blue arrows denote sites corresponding to 5'PEAT-seq read clusters. Blue highlighted sequence denotes the leucine zipper domain. Red circles indicate two alternative non-canonical start codons identified by TIS predictor. Underlined sequence corresponds to the transcript purified from 5'RACE in part A.

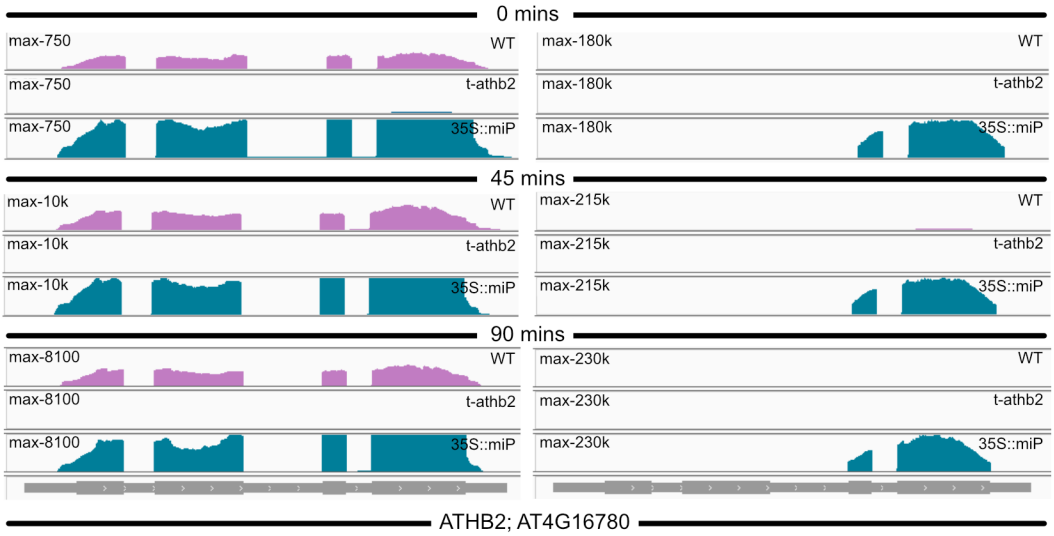

Figure S2. RNA-seq reads corresponding to ATHB2 at 0, 45 and 90 mins of WL+FR. RNA-seq reads peak in the WT at 45 mins of WL+FR and then decrease (see y axis for maximum reads) at 90 mins. No reads are detected in t-athb2 mutants whereas 35S::miP mutant plants are highly overexpressing exon 3 and 4 of *ATHB2*.

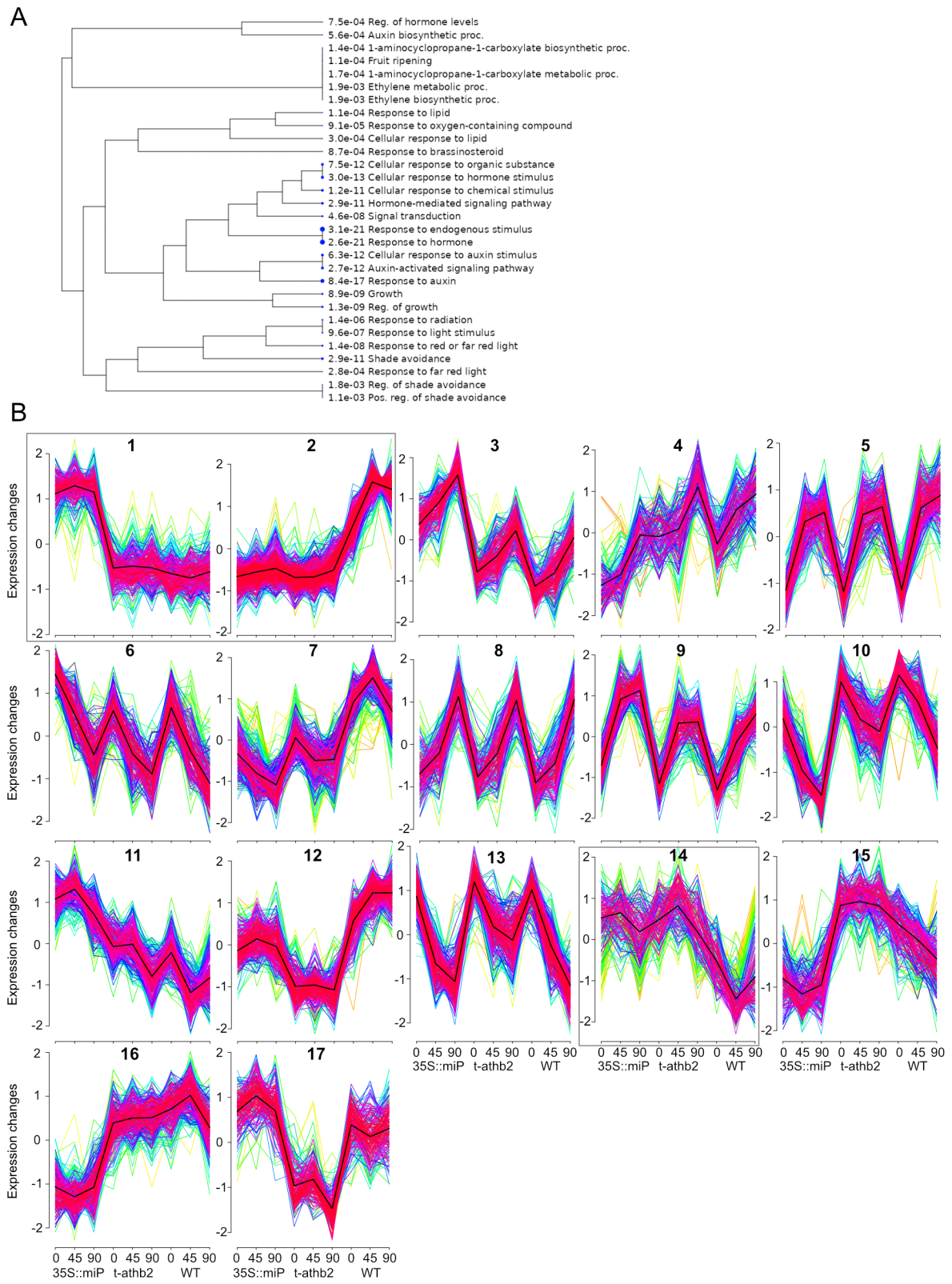

Figure S3. (A) GO enrichment analysis performed on the 95 DEGs whose expression is shown in Figure 3 yielded several significant GO terms (FDR < 0.05). These GO terms are related to shade avoidance, hormone/auxin responses, and growth. **(B)** The 17 clusters generated by MFuzz. The expression changes of the genes included in each cluster are on the y-axis. The three genotypes and shade conditions are on the x-axis.

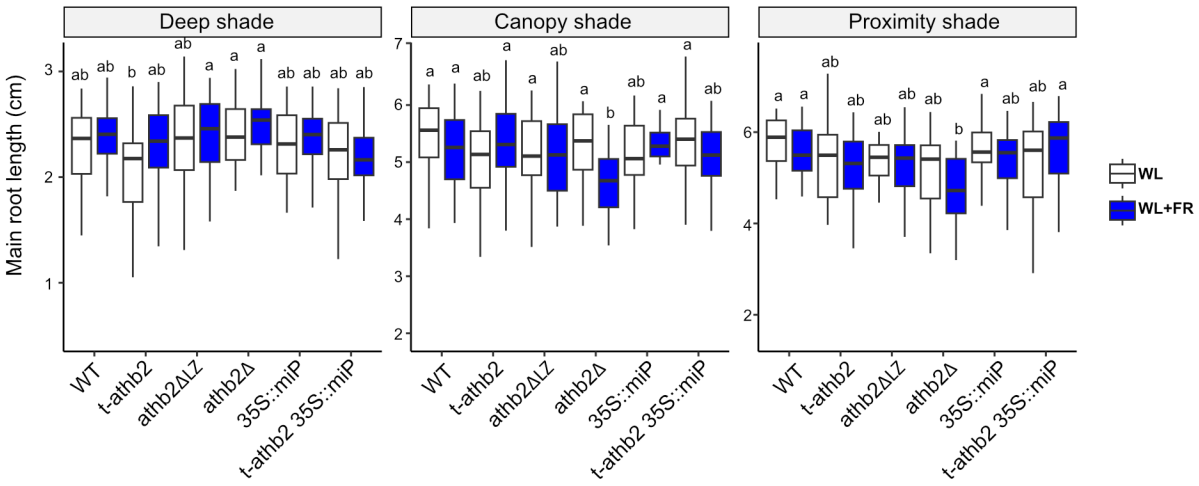

Figure S4. ATH2B and ATHB2miP slightly effect root elongation. Main root length of 11-day old plants grown in either WL or WL+FR in three different shade regimes: deep shade, canopy shade and proximity shade.

genes	Proteins	logFC 35S:ATHB2miP	logFC t-athb2
Auxin – Auxin biosynthesis - Elongation (White light)			
AT1G04180	YUCCA 9	1,139682	1,64699
AT2G46970	Phytochrome interacting factor 3-like 1	1,305277	2,439173
AT3G45960	expansin-like A3	2,080293	2,036176
Auxin – Auxin biosynthesis - Elongation (45 minutes FR)			
AT1G04610	YUCCA 3	1,144326	1,599915
AT1G15580	IAA5	1,938751	1,410173
AT1G29490	SAUR-like auxin-responsive protein family	1,008418	1,038856
AT1G79840	HD-ZIP IV family of homeobox-leucine zipper protein	1,078921	1,186589
AT2G22810	1-aminocyclopropane-1-carboxylate synthase 4	1,205152	2,452735
AT3G15540	indole-3-acetic acid inducible 19	1,037735	1,543168

genes	Proteins	logFC 35S:ATHB2miP	logFC t-athb2
Root system development (45 minutes FR)			
AT1G12040	Leucine-rich repeat extensin-like protein 1	1,284717	1,615868
AT1G12560	expansin A7	1,055020	1,253633
AT1G62980	expansin A 18	1,218306	1,585069
AT1G74500	Transcription factor PRE3	1,441103	1,542728
AT3G10710	root hair specific 12	1,272890	1,539051
AT4G01630	expansin A 17	1,883914	1,279127
AT4G02270	root hair specific 13	1,120669	1,399511
AT4G15290	Cellulose synthase-like protein B5	2,416704	1,639997
AT4G29180	root hair specific 16	1,533472	1,800483
AT4G33880	Transcription factor RSL2	1,849955	1,112459
AT5G58010	Transcription factor LRL3	1,502507	1,246907
AT5G60660	Probable aquaporin PIP2-4	1,910733	1,290471
AT5G67400	root hair specific 19	1,294708	1,624888
Root system development (90 minutes FR)			
AT1G05650	Pectin lyase-like superfamily protein	1,924806	1,033376
AT1G14960	Polyketide cyclase/dehydrase and lipid transport superfamily protein	1,027060	1,184112
AT1G53680	glutathione S-transferase TAU 28	1,438246	1,052538
AT4G01630	expansin A 17	1,536239	1,167905
AT4G29180	root hair specific 16	1,269614	1,078967
AT4G33120	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	1,665754	1,080399
AT5G49270	COBRA-like extracellular glycosyl-phosphatidyl inositol-anchored protein family	1,785649	1,208562