Plant Biology

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

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

University of Copenhagen, Department of Plant and Environmental Sciences, Thorvaldsensvej 40, 1871 Frederiksberg, Denmark • Department of Plant Physiology, Umeå Plant Science Centre, Umeå University, SE-901 87, Umeå, Sweden • current address: Carlsberg Research Laboratory, J.C. Jacobsens Gade 4, Copenhagen V DK-1799, Denmark • current address: CSIR, Central Institute of Medicinal and Aromatic Plants, Lucknow, India • current address: Department of Plant Science, Wagening University, The Netherlands • current address: Quantitative Biology Center (QBiC), University of Tübingen, Auf der Morgenstelle, Tübingen, Germany • Institute of Molecular Biology and Pathology, National Research Council, P.le A. Moro 5, 00185 Rome, Italy • Research Centre for Genomics and Bioinformatics, Council for Agricultural Research and Economics (CREA), Via Ardeatina 546, 00178 Rome, Italy

Reviewed Preprint

Published from the original preprint after peer review and assessment by eLife.

About eLife's process

Reviewed preprint version 1
April 26, 2024 (this version)

Posted to preprint server
February 15, 2024

Sent for peer review
February 13, 2024

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.

eLife assessment

Through a genome-wide screen for functional alternative transcription start sites (TSS) in Arabidopsis, the authors provide evidence for widespread transcription of potential microproteins from previously annotated protein-coding genes. Functional analysis of AtHB2-miP, derived from the C-terminal region of transcription factor AtHB2 and predicted to form non-productive dimers with ATHB2, suggested that this microprotein could affect AtHB2 functions in shade responses, root growth, and iron homeostasis. The work is **valuable** as a case study of how new microproteins could act to modulate gene regulation in response to environmental change, but the focus on a single gene, the lack of precision in AtHB2-miP measurement with missing controls, and the relatively minor phenotypic effects in the specific case investigated, leave it unclear how **important** microprotein production is as a general regulatory strategy.

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

Results

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

To mine the transcriptome of *Arabidopsis thaliana* for novel microProteins regulating shade avoidance, we used 5' PEAT sequencing, a method that enriches capped RNA transcripts and results in reads that correspond to transcription start sites (Ni *et al*, 2010). Plants were grown for ten days in a white light regime (WL; R:FR = 7.7) and then exposed to additional far-red light, to simulate shade (WL+FR; R:FR = 0.13; deep shade). After shade treatment for 45- and 90-minutes samples, along with an untreated control (0 minutes shade) were collected. Overall, we identified 12,398 transcription start sites (TSSs). The majority of these corresponded to putative TSSs from TAIR10, but 40% (4,982) corresponded to sites that had not previously been annotated (**Fig. 1A**). Most of the reads mapped to the 5'UTR of already annotated genes, but a subset was found to map to regions within gene bodies. As we were particularly interested in identifying novel transcripts that might encode microProteins (i.e. overlapping gene products), we focused on the subset of genes encoding such alternative microProteins. We therefore established a set of criteria in order to identify sites in our dataset that might correspond to an alternative microProtein mRNA transcript. Firstly, we limited our search to open reading frames (ORFs) that would encode a protein within the size range of a typical microProtein of 5-15 kDa but would differ significantly in size to the overlapping full-length gene product. Secondly, we selected for ORFs that contained a single protein-protein interaction domain allowing the alternative proteins to interfere with the full-length isoforms by forming heterodimers. Finally, we considered the biological relevance of each alternative transcript: the number of reads must exceed a certain threshold to ensure robust expression of the transcript isoform. By applying these criteria, we identified 377 potential microProtein candidates, corresponding to 3% of the total number of TSSs identified in our dataset (**Fig.1A**). One of the alternative microProtein candidates we identified overlaps with the end of *ATHB2* (**Fig. 1B**), a gene encoding a class II HD-ZIP transcription factor that plays a central role in the shade avoidance response (Ciarbelli *et al*., 2008; Schena *et al*, 1993). Like the full-length transcript, the alternative transcript appears to be shade-sensitive, with 5'PEAT-seq showing that expression peaks at 45 minutes. The *ATHB2* alternative transcript was also identified by 5'RACE (**Fig. S1A**) and by sequencing the PCR product, we found that the alternative transcript is spliced

at the same site as the full-length transcript and starts around position 653 of the full-length transcript (**Fig. S1B**). In Arabidopsis, the average length of a 5'UTR is 155 bp (Srivastava *et al*, 2018), therefore, we searched for alternative start codons no more than 300 bp downstream of the alternative TSS. We identified two near-cognate start codons with strong Kozak sequence similarity (≥ 0.7 and < 0.8) using the online tool TISpredictor (Gleason *et al*, 2022). The two start codons are TTG and CTG, both in-frame with the full-length ATHB2 protein sequence and producing ORFs that are 126 and 102 amino acids long, respectively, and consisting of the leucine zipper domain, and, in the case of the former, parts of the homeodomain (**Fig. S1B**). It is unclear whether the alternative transcript we have identified is the result of an independent transcription event or due to post-transcriptional processing of the full-length *ATHB2* transcript involving downstream recapping (Ni *et al*., 2010). To confirm that ATHB2miP is the result of an independent transcription event, we sequenced full-length cDNAs using PacBio isoseq. This analysis revealed isoforms that start in the third exon of ATHB2, supporting an independent transcription event (**Fig. 1C**). Furthermore, we also observed antisense transcripts covering the introns. These changes occurred for both isoseq and the 5'PEAT-seq only at the 45-minute timepoint, indicating major chromatin reorganization in response to short-term shading. At 90 minutes, the situation stabilized, and ATHB2 was expressed at higher levels (**Fig. 1C**). To determine whether there are sequences in the gene body of *ATHB2* that can promote transcription and translation, we fused a β-glucuronidase *uidA* (GUS) reporter gene to different sequences from the *ATHB2* gene body (**Fig. 1D**). First, we fused GUS to the sequence 1 kb upstream of the *ATHB2* full-length start codon ATG (*pATHB2::GUS*). As a control, we fused GUS to the sequence immediately upstream of an ATG located in the middle of intron two (*pATHB2control::GUS*). Finally, we fused GUS to the sequence immediately upstream of the first alternative, in-frame codon that we had identified, TTG (*pATHB2miP::GUS*). Whilst *pATHB2control::GUS* did not display any GUS staining upon treatment with the substrate 5-bromo-4-chloro-3-indolyl ß-d-glucuronide (X-Gluc; **Fig. 1F**), both *pATHB2::GUS* and *pATHB2miP::GUS* showed staining in the shoot apical meristem (SAM) and the hypocotyl, though to a lesser extent in the latter in *pATHB2miP::GUS* (**Fig. 1E, G**). This suggests that there are sequences within the second intron and third exon, that are excluded from *pATHB2control::GUS*, that can promote both transcription and translation. This might also lend support to our hypothesis that the alternative transcript arises through an independent transcription event to the full-length *ATHB2* transcript.

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

We hypothesized that the ATHB2miP might inhibit the full-length ATHB2 in a manner similar to that of the LITTLE ZIPPERs (Kim *et al*., 2008; Wenkel *et al*., 2007). To test this hypothesis, we posed the following questions. Firstly, does the microProtein localize to the same cellular compartment as the full-length protein? It was previously reported that ATHB4, another member of the HD ZIPII family, contains a nuclear localization signal (NLS) in its homeodomain that is highly conserved in ATHB2 (Gallemi *et al*, 2017). However, our own analysis using LOCALIZER (Sperschneider *et al*, 2017), which has been trained specifically on plant proteins, suggests that there is an NLS in the LZ domain that is conserved in other HD-ZIPIIs, including ATHB4. Interestingly, LOCALISER also predicts a chloroplast transit peptide (cTP) at residues 1-46 of ATHB2 with a probability of 0.8. The cTP does not appear to be conserved, suggesting that, if real, it is specific to ATHB2. To determine the subcellular localization of ATHB2 and ATHB2miP, we transiently expressed eGFP fusions of the proteins in *Nicotiana benthamiana* leaves (**Fig. 2A**; eGFP-ATHB2, eGFP-ATHB2miP). eGFP-ATHB2miP encompasses the entirety of exon 3 and 4 of ATHB2. To further test whether the homeodomain of ATHB2 could target the nucleus, as shown previously for ATHB4 (Gallemi *et al*., 2017), we also fused a variant of ATHB2 lacking the leucine zipper to eGFP (eGFP-ATHB2HD). We observed that all variants of ATHB2 localized exclusively to the nucleus, unlike the negative control (eGFP) which was dispersed along the cell membrane and in the nucleus (**Fig. 2A**). We also observed that the distribution of eGFP-ATHB2 was not uniform

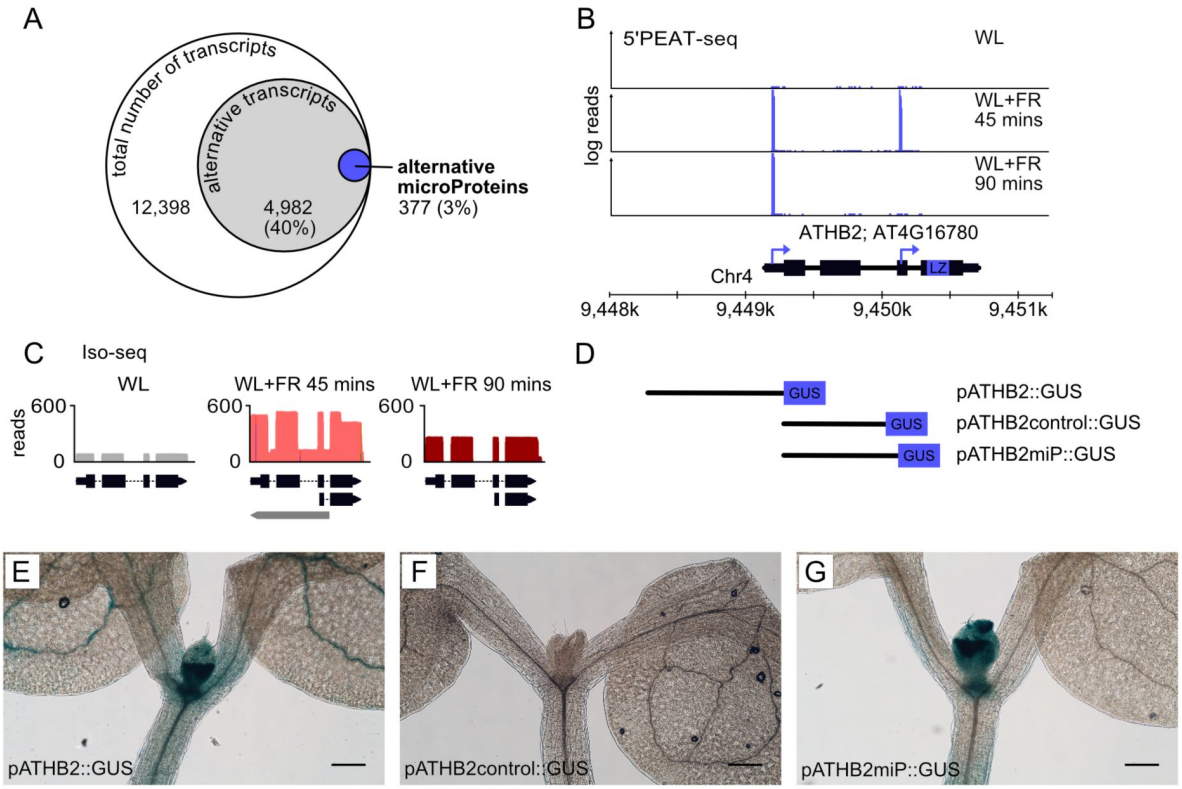

Figure 1.

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.

in the nucleus, but instead localized to distinct nuclear speckles or, in some case, to the nucleolus specifically. Nuclear speckles were mostly absent in leaves expressing eGFP-ATHB2miP, but entirely so in leaves expressing eGFP-ATHB2HD, suggesting that interactions mediated by the leucine zipper domain and in part by the N-terminal domain are required for the speckles to form. We next asked whether we could find evidence that ATHB2 and ATHB2miP can physically interact with each other. It has been demonstrated that ATHB2 homodimerizes in order to bind to its target DNA sequence and that this interaction is directed by the leucine zipper domain (Sessa *et al*, 1993). To test whether ATHB2 and ATHB2miP can interact, we performed a yeast-2-hybrid (Y2H) study. As in the localization assay, ATHB2 or ATHB2miP were fused to either the activation or the DNA binding domain of GAL4. Constructs were transformed into yeast carrying the *HIS3* gene fused to the target DNA sequence of *GAL4*. Growth of transformed cells on media lacking histidine revealed that ATHB2 and ATHB2miP can interact and form heterodimeric complexes as well as homodimers (**Fig. 2B**). To further confirm this interaction *in planta* we carried out a FRET-FLIM assay. Here again, ATHB2 and ATHB2miP were fused to either eGFP or mCherry and then transiently co-expressed in tobacco leaves. Using a confocal microscope, fluorescence lifetime of eGFP-ATHB2 was measured. We observed that in cells that co-expressed fluorescently tagged versions of ATHB2 or ATHB2miP there was a significant decrease in fluorescence lifetime of eGFP-ATHB2 when compared to cells that expressed eGFP-ATHB2 alone or with a control (mCherry; **Fig. 2C**), supporting that ATHB2 and ATHB2miP can physically interact.

ATHB2miP perturbs the ability of ATHB2 to regulate transcription

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

In total, we identified 95 DEGs whose expression depended on genotype, environment, and genotype x environment (**Fig. 3A**). The GO analysis of these 95 genes revealed an overrepresentation of genes involved in the shade avoidance response, growth, and hormone/auxin signaling (**Fig. S3A**). These findings demonstrate that both *t-athb2* mutant plants and the transgenic plants overexpressing ATBH2miP can detect shade and respond to it at the molecular level. In order to compare the genotypes and environmental conditions, we carried out an unbiased gene cluster analysis using mFuzz (Kumar & Futschik, 2007). Our analysis revealed 17 individual clusters with distinct expression patterns (**Fig. S3B**). Cluster 1 (**Fig. 3B**) contains 200 genes that are upregulated only in *35S::miP*, but not in either of the other two genotypes. We found an overrepresentation of genes involved in root development in this cluster. Cluster 2 contains 382 genes with lower expression in both 35S::miP and *t-athb2*. This cluster is also enriched in genes encoding hormonal regulators and regulators of root development. Finally, cluster 14 contains 152 genes that are up-regulated in both *35S::miP* and *t-athb2* plants (**Fig. 3B**).

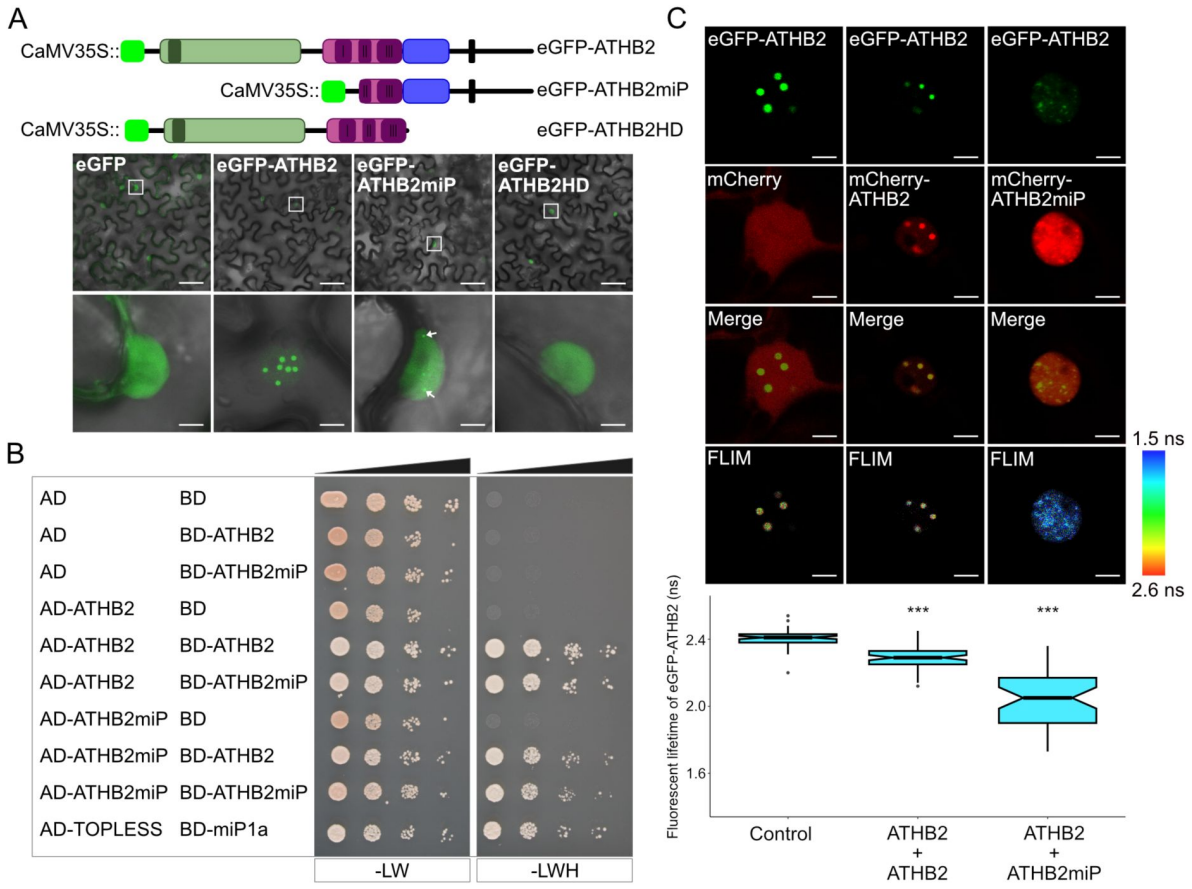

Figure 2.

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.

Genes in this cluster have known roles in the shade avoidance response. We also found an overrepresentation of genes involved in iron uptake and signalling, which are strongly upregulated in *35S::miP* and *t-athb2* plants. The latter finding suggests that ATHB2 may play a role in mediating between nutritional status and growth responses.

Having established that ectopic expression of ATHB2miP perturbs ATHB2 function and alleviates the negative auto-repressive feedback loop we assessed DEGs in the three different environmental conditions (0-, 45-, and 90-minutes additional FR-light). Focusing on the genes upregulated in both *t-athb2* and *35S::miP* plants, we found 22 genes in white light (**Fig. 3C**), 157 genes in 45-minute shade (**Fig. 3D**), and 20 genes in 90-minute shade (**Fig. 3E**). As previously observed, we identified upregulated genes related to auxin synthesis and signaling, as well as genes involved in root development. Genes related to auxin-biology and hypocotyl elongation (Table 1) include *YUCCA3* and *YUCC9*, both of which have known functions in auxin biosynthesis in response to shade. (Goyal *et al*, 2016). Genes associated with root development (Table 2) include several cell wall modifying enzymes, most notably expansins, as well as transcription factors. Taken together, these findings support the function of ATHB2 as a suppressor of auxin signaling and root development. Analysis of genes down-regulated in both *t-athb2* and *35S::miP* plants showed that genes encoding defence regulators were enriched in all conditions (0, 45 and 90 min additional FR light). In addition, we found that genes involved in the biosynthesis of tryptophan were down-regulated in the shaded environments. Since auxin is mainly synthesised from tryptophan (Tao *et al*, 2008; Won *et al*., 2011), the latter finding may explain why *athb2* mutant plants have defects in shade-induced growth promotion (Ciarbelli *et al*., 2008; Schena *et al*., 1993).

ATHB2 controls elongation growth and root development in response to shade

In shade-avoidant plants, such as *A. thaliana*, shade caused by neighboring plants elicits a set of growth responses including elongation of stem-like structures, hyponasty and leaf extension, and acceleration of flowering (Martinez-Garcia & Rodriguez-Concepcion, 2023). It has been demonstrated that ATHB2 plays a central role in the shade avoidance response, with transgenic Arabidopsis plants where *ATHB2* is affected displaying varying degrees of shade response defects, the most obvious being in the elongation of the hypocotyl (Steindler *et al*., 1999). Early studies found that when knocked down or overexpressed, *ATHB2* was associated with subsequent shade responses where there was less or more hypocotyl elongation compared to the WT, respectively. For this reason, and because we show that the two proteins interact, we expected that transgenic plants where ATHB2miP (*35S::miP*) is overexpressed would show a phenotype indicative of lost ATHB2 function, i.e. shade insensitivity. To better understand the role of the region corresponding to ATHB2miP, we also created mutant plants by CRISPR-Cas9 gene editing to compare with the overexpression and KO line already described (**Fig. 4A**). The first, *athb21LZ*, presents a loss-of-function allele of *ATHB2* caused by a 25-bp deletion in the third exon, causing a frameshift and nonsense mutation, and resulting in a protein product lacking the leucine zipper and part of helix III of the homeodomain. Secondly, *athb21*, is a complete loss-of-function of *ATHB2*, caused by a frameshift mutation in the beginning of the first exon, resulting in a truncated protein that shares only its first twelve amino acids with ATHB2. Based on RT-qPCR results where we analyzed expression of *ATHB2* after 45 minutes in shade, we see that transcription is not affected in any of the CRISPR lines we have generated, unlike in *t-athb2* where *ATHB2* is clearly downregulated (**Fig. 4B**), as also shown in the RNA-seq analysis (**Fig. S2**). To measure responsiveness to shade, we grew these mutant and transgenic plants alongside the Col-0 wild-type (WT) under three different shade regimes (**Fig. 4C**). We grew plants in deep shade (13 µmol m-2 s-1 PAR in WL and WL+FR; R:FR = 0.13 in WL+FR) with reduced PAR and a very low R/FR-ratio. In addition, we used canopy shade (45 µmol m-2 s-1 PAR in WL and WL+FR; R:FR = 0.15 in WL+FR) and proximity shade (45 µmol m-2 s-1 PAR in WL and WL+FR; R:FR = 0.06 in WL+FR) with higher PAR but different F/FR-ratios. In deep shade, we observed that *athb2* mutants were hypersensitive, with elongated hypocotyls in white light conditions and, as in the case of *35S::miP*, with very long hypocotyls in

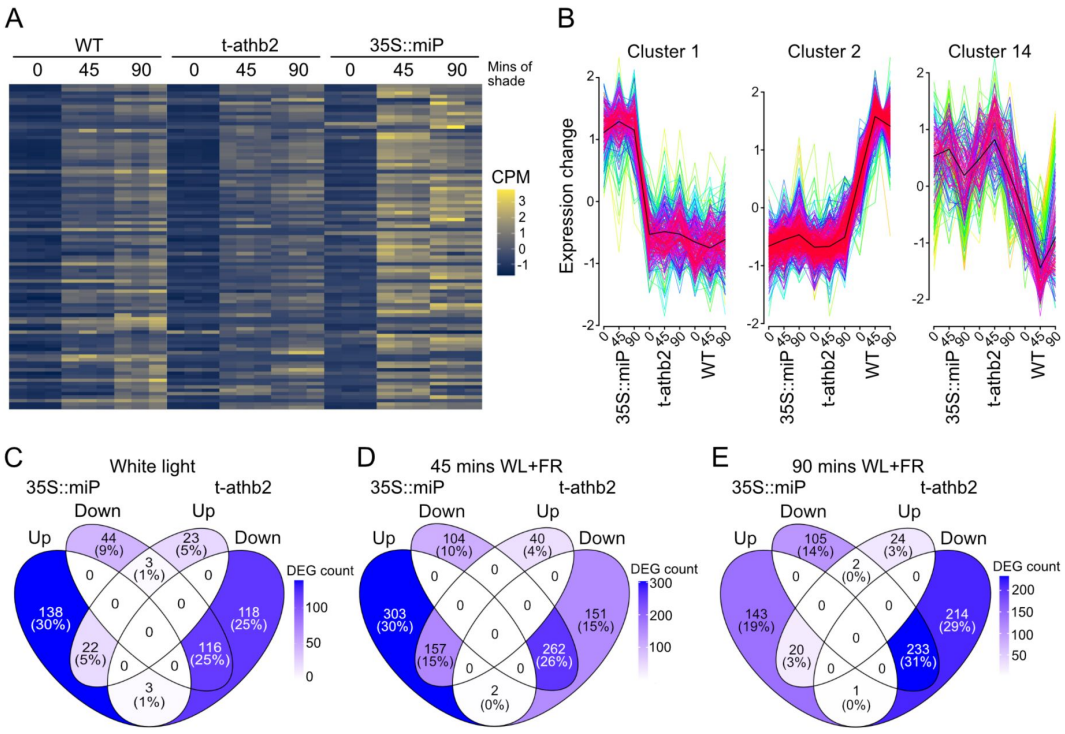

Figure 3.

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.

shade. Interestingly, the CRISPR mutants *athb21* and *athb21LZ* exhibited responses similar to those of wild-type plants. It is worth noting that ATHB2 is still expressed in these mutants (**Fig. 4B**), which suggests that ATHBmiP and other transcript or protein isoforms may still be produced. Under canopy shade, the *athb2* mutants exhibited a growth behaviour that was less sensitive to shade, resulting in reduced hypocotyl elongation in shaded conditions. However, under proximity shade, there were no significant differences in hypocotyl elongation between the wild type and *athb2* mutant plants (**Fig. 4C**). The results indicate that ATHB2 has a dual function. In deep shade, it acts as a growth repressor, whereas in canopy and proximity shade, it acts as a growth promoter.

Class II HD-ZIPs have been suggested to have some redundant function in the shade avoidance response, so it is possible that ATHB2miP also interacts with these other proteins in the absence of full-length *ATHB2*. To determine whether the effect of *35S::miP* depends on the presence of the endogenous copy of *ATHB2*, we complemented the *t-athb2* knock-out with *35S::miP*. In deep shade, we observed that both *35S::miP* and *35S::miP t-athb2* plants developed longer hypocotyls under white light, indicating that ATHB2miP may have inhibited other transcription factors. Additionally, *35S::miP* plants had significantly longer hypocotyls than *35S::miP t-athb2* plants in deep shade, suggesting that the absence of full-length ATHB2 may contribute to this difference.

The shade avoidance response is also associated with changes in root architecture. For example, one study found that lateral root density was decreased in plants that were exposed to longer periods of shade (van Gelderen *et al*, 2018). Mutations affecting ATHB2 have been reported to affect root development, disrupting both root length and lateral root formation (Schena *et al*., 1993), and we also identified a number of genes related to root morphogenesis in our RNA-seq analysis. Therefore, we questioned whether ATHB2miP could also have a role in root development.

Following the shade regimes described earlier, plants were grown in either WL+FR or WL for a period of 11 days and primary root length was measured. Our results indicate that there were no significant differences in all conditions tested, except for the *t-athb2* mutant plants which had slightly shorter roots in white light and even shorter roots in the proximity shade regime (**Suppl. Fig. S4**). We also assessed the ability of plants to form lateral roots under light and shade conditions. While no difference in the number of lateral roots respective to geneotypes and environmenal conditions was observed in deep shade and canopy shade, we detected significant differences in proximity shade conditions (**Fig. 4D**). In white light, *t-athb2* mutants, CRISPR mutants (*athb21* and *athb21LZ*), and *35S::miP* transgenic plants all all had significantly lower numbers of lateral roots, suggesting that ATHB2 promotes lateral root emergence broadly. In response to shade, a decrease in the number of lateral roots was observed in wild type plants. However, in *t-abhb2*, *athb21* and *athb21LZ* mutant plants, the number of lateral roots did not change in shade. This indicates that ATHB2 is required to repress lateral root emergence in shade. Ectopic expression of ATHB2miP severely affected lateral root development in shade, suggesting that both ATHB2miP and ATHB2 act to repress 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.

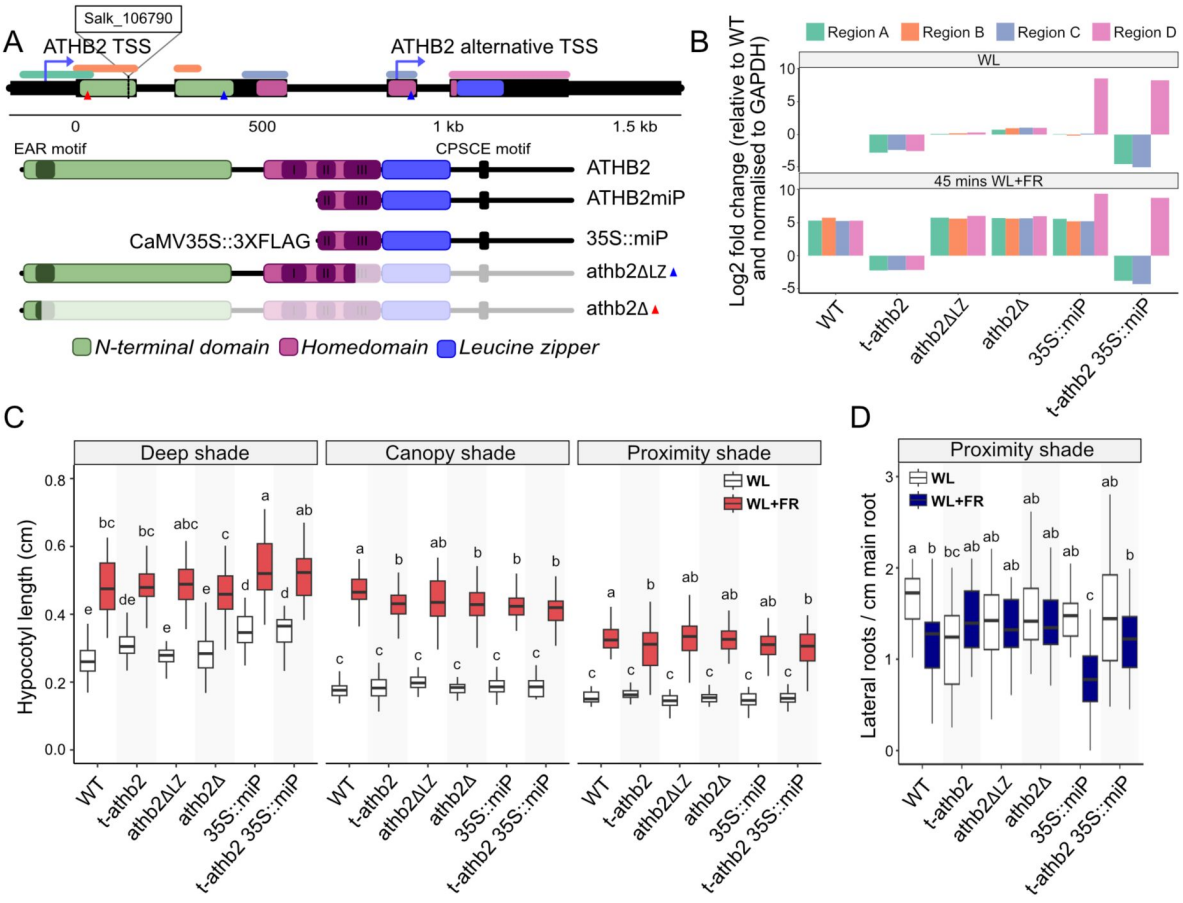

Figure 4.

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.

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*, *athb21* and *athb21LZ* 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**).

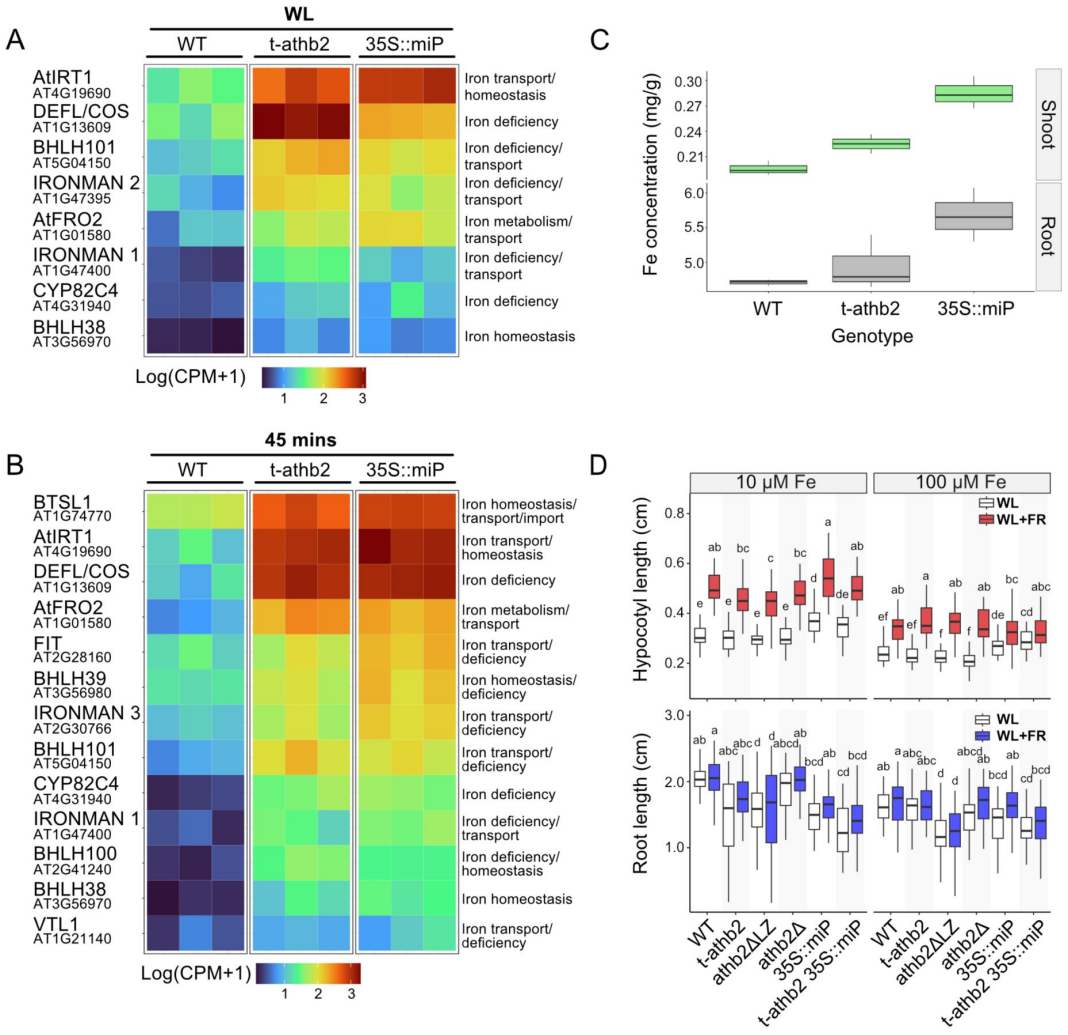

Figure 5.

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.

However, we observed that all *athb2* mutant plants, with the exception of *athb21* 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 *athb21* 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 to shade. ATHB2 mediates between root growth and hypocotyl elongation and can either activate or repress hypocotyl elongation in response to different iron concentrations.

Discussion

In this study, we assessed the impact of shading on transcript isoform production and their potential to encode microProteins. It was found that almost half of the Arabidopsis genes that had reads at the transcription start site also had alternative transcription start sites in the gene body. Only a fraction of these genes had the potential to encode microProteins. It is furthermore fair to assume that many more potential microProteins exist, but these are produced in a tissue-specific and/or inducible fashion. We filtered the dataset to prioritize alternative transcripts with robust expression levels, excluding those unsupported by multiple reads, which again could be candidates that have a narrower pattern of expression. Our approach focused on enriching microProtein candidates likely to have biological significance. To enhance the likelihood of identifying microProteins involved in the shade avoidance response, we considered alternative methods for studying transcript isoforms, including parallel analysis of RNA 5'ends from low-input RNA (nanoPARE), and ribosome profiling (Ingolia, 2014; Schon *et al*, 2018). Ribosome profiling, which employs deep sequencing to monitor *in vivo* translation, can provide additional insights into protein formation. Such future approaches can both confirm the purely sequencing-based identification of candidate microProteins but also aid to identify novel candidates.

The list of microProtein candidates generated through the 5'PEAT-seq approach contained few shade avoidance regulators. Notably, microProteins may serve distinct functions compared to their full-length counterparts, albeit often targeting evolutionarily related proteins. Adjusting filtering criteria could identify more transcript isoforms related to shade responses, although they may not classify as microProteins. The ATHB2miP microProtein, chosen due to its similarity to the known LITTLE ZIPPER microProteins (Kim *et al*., 2008; Wenkel *et al*., 2007), was expected to interact with ATHB2 and regulate its activity, akin to ZPRs and HD-ZIPIIIs. Yeast-Two-Hybrid tests and pull-down assays confirmed physical interactions between ATHB2 and ATHB2miP. Notably, ATHB2miP can also homodimerize, a feature attributed to the leucine zipper domain's role as a protein-protein interaction motif in HD-ZIP proteins.

We assessed potential alternative *ATHB2miP* promoters using the GUS protein assay, identifying one active promoter, *pATHB2miP::GUS*, which did not possess a typical ATG start codon. This observation is consistent with prior findings of alternative start sites not commencing with an ATG. Differences in GUS expression patterns between *ATHB2* and *ATHB2miP* promoters were observed, with ATHB2miP primarily present in the shoot apical meristem (SAM) and early leaf primordia. The absence of ATHB2miP in other tissues suggested that it does not directly regulate ATHB2 in those areas.

Analysis of the subcellular localization patterns of ATHB2miP and ATHB2 protein-fusions to GFP revealed that ATHB2 localizes mostly to larger subnuclear speckles, a phenomenon also observed for phytochromes and other light signaling components and therefore often referred to as photobodies (Kim *et al*, 2023; Kircher *et al*, 2002). Interestingly, ATHB2miP protein showed a

more diffuse nuclear localization implying discrete functions of both isoforms. However, when co-expressed both isoforms appear in speckles, but these are more dispersed, indicating a direct physical interaction where the full-length isoform is partially removed from the speckles. The full-length ATBH2 isoform has an intrinsically disordered amino terminus and could undergo phase separation and engage in more flexible protein-protein interactions within the photobodies. As ATHB2miP lacks the disordered part, it may prevent the phase separation activity of ATHB2.

To investigate the regulatory role of ATHB2miP, we compared *ATHB2miP* overexpression plants to *athb2* mutants. Surprisingly, in deep shade conditions all *athb2* knock-down and knock-out mutants as well as *ATHB2miP* overexpression plants showed elongated hypocotyls in white light conditions compared to Col-0 wild type plants while the hypocotyl length in shade grown seedlings showed no significant differences when compared to wild type. Except 35S::miP transgenic plants that showed excessive hypocotyl elongation in both white light and far-red enriched conditions (**Fig. 4C**). This indicates that ATHB2 functions as a growth repressor in deep shade and ATHB2miP has the capacity to inactivate ATHB2 which results in hypocotyl elongation in shade. The finding that the growth differences can only be observed in deep shade conditions indicates that ATHB2 activity is dependent on other components that are only present in this type of shade. In canopy and proximity shade ATHB2 functions as a promoter of elongation growth, hence the shorter hypocotyls in these types of shade.

ATHB2 has also been reported to be involved in controlling root growth and development (Schena *et al*., 1993) prompting us to investigate the role of ATHB2 and ATHB2miP in this process. The analysis of root development in different types of shade did not reveal significant differences in root length. However, there were notable differences in the ability of *athb2* mutants to produce lateral roots. Additionally, all tested *athb2* mutants appeared to be insensitive to shade, as the number of lateral roots did not change in response to proximity shade, unlike in the wild type. In summary, these results suggest that ATHB2 acts as a suppressor of lateral root growth in response to proximity shade. The overexpression of ATHB2miP had a significant impact on lateral root development in both white light and shade. Notably, in the *t-athb2* mutant background (*t-athb2 35S::miP*), the strong shade suppression was not observed, indicating that ATHB2 is required for strong shade suppression. These 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 shade. In summary, our data support a model for ATHB2/ATHB2miP as integrator of nutrient status and elongation growth of both the shoot and the root.

Materials and methods

Plant materials and growth conditions

All plants were grown in growth chambers on ½ Murashige and Skooge (MS) 1% agar under long day (LD) conditions at 22°C during daytime and 20°C during night. For the shade treatment for the 5'PEAT-seq and RNA-seq experiments, seedlings were transferred after 10 days from a white light only compartment (PAR of 13 µmol/m^2/s, R:FR = 7.7) to a compartment with additional far-red lights (PAR of 13 µmol/m^2/s, R:FR = 0.2) for treatments of 0, 45 and 90 minutes. Whole seedlings frozen in liquid N_2 before grinding with a mortar and pestle and RNA extraction. For the shade treatments related to measuring hypocotyl length and root growth, half of the plants were transferred after two days from a white light only compartment (PAR of 45 µmol/m^2/s, R:FR = 5.5) to a compartment in the same chamber with additional far-red lights (PAR of 45 µmol/m^2/s, R:FR = 0.15). The other half remained in the white light only compartment. After five and eleven days, seedlings were photographed and hypocotyls and roots measured using IMAGEJ.

For the growth and phenotyping of seedlings on iron-enriched and iron-depleted media, two different growth media were prepared. Both media contained the same vitamins, macro and micro elements of standard MS salt (M0222, Duchefa) and in the same concentrations, with the exception of FeNAEDTA, used in a concentration of 10uM in the iron-depleted medium and 100uM in the iron-enriched medium. Seedlings were grown on 1% agar iron-depleted and iron-enriched media under the same growth conditions of the other growth experiments. Half of the seedlings on plates were transferred after two days from a white light only compartment (PAR of 13 µmol/m2 /s, R:FR = 7.7) to a compartment in the same chamber with additional far-red lights (PAR of 13 µmol/m2 /s, R:FR = 0.2). The other half remained in the white light only compartment. After five and eleven days, seedlings were photographed and hypocotyls and roots measured using IMAGEJ.

5'-PEAT library preparation and analysis

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

Kit for Illumina. The finished libraries had a fragment size of 500 bp. BGI tech solutions (Hongkong) performed sequencing on an Illumina HiSeq (150bp paired end, 40 M read pairs per sample).

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

Generation of transgenic and CRISPR-Cas9 edited plants

To generate the overexpression of ATHB2miP (*35S::ATHB2miP*) in Arabidopsis oligos were designed to amplify exons 3 and 4 of *ATHB2* and the resulting sequence then cloned into the vector pJAN33 by gateway cloning. Plants were transformed by floral dipping and transformants selected for by BASTA treatment.

For CRISPR mutagenesis, sgRNAS were cloned into the CRISPR/Cas9 vector pKI1.1R. The cloning was performed as described (Tsutsui and Higashiyama, 2016). *Athb2-cr2* was made by Cas9 gene editing targeted using the following sgRNA combination: 5'-TACAGTCTCAAGCTCTACA-3' and 5'-GTTTCAGAACAGACGAGCA-3'. *Athb2-cr3* was made by Cas9 gene editing targeted using the single sgRNA: 5'-ACGATCTGGGTCTAAGCTT-3'. Plants were transformed by floral dipping and then RFP-positive seeds selected for in the T1 generation. Genomic DNA of 4-week old plants was isolated using the Edwards method and the plants were genotyped by PCR. The PCR products were gel purified using the E.Z.N.A® Cycle-Pure Kit (Omega Bio-tek) and sequenced at Macrogen (Amsterdam).

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

Yeast-2-Hybrid

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

GUS localization assay

Oligos were designed to amplify the 1 kb sequence upstream of *ATHB2*, and the sequences upstream of the alternative start codon TTG in exon 3 and upstream of a control start codon ATG in intron 2, up to and including the putative start codon ATG of *ATHB2*. Sequences were cloned into pBGWFS7 by gateway cloning and the resulting vectors transformed into plants by floral dipping. Arabidopsis thaliana (ecotype Col-0) seeds were surface-sterilized and sown on Murashige and Skoog (MS) agar plates. The plates were incubated at 4°C for 2 days for stratification and subsequently transferred to a growth chamber under long-day conditions. Seven day old seedlings were immersed in GUS staining solution (XGLUC; 14-mM NaH2PO4, 36-mM Na2HPO4, 2-mM K4[Fe(CN)6], 2-M K3[Fe(CN)6], 10% Triton X-100). Seedlings were vacuum infiltrated for 5 minutes, incubated overnight at 37 °C in the dark, and rinsed with PBS to remove excess staining solution and dehydrated through an ethanol series (30%, 50%, 70%, 90%, and 100%) to remove chlorophyll and enhance contrast.. Pictures of seedlings were taken with a canon EOS 450D camera on a Nikon Eclipse 80i Fluorescence microscope. White balance and scalebar were added via ImageJ.

Subcellular localization and FRET-FLIM analysis

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

Differential Gene Expression Analysis and Clustering

Total RNA was extracted from Arabidopsis thaliana Col-0 wild-type (WT), t-athb2, and 35S::ATHB2miP genotypes under three different light conditions as described: White Light (WL), 45 minutes of Far-Red (FR) enrichment, and 90 minutes of FR enrichment. Three biological replicates per genotype-condition combination were used. RNA quality and concentration were assessed using a NanoDrop spectrophotometer and Agilent Bioanalyzer. RNA was sequenced on an Illumina HiSeq platform.

Raw sequencing reads were quality-checked using FastQC to ensure data integrity. Trimming of reads to remove adapter sequences and low-quality bases was performed using Trimmomatic. After preprocessing, high-quality reads were retained for downstream analysis. Read alignment to the Arabidopsis reference genome (TAIR10) was conducted using the STAR aligner within the R environment. This generated BAM files for each sample, which were subsequently used for

quantification of gene expression. Differential gene expression analysis was carried out using the edgeR package in R. Raw counts were normalized using the TMM (Trimmed Mean of M-values) method. Genes with low expression were filtered out, and statistical analysis was performed to identify differentially expressed genes (DEGs) between genotypes and conditions. P-values were adjusted for multiple testing using the false discovery rate (FDR) method. DEGs were subjected to gene ontology (GO) enrichment analysis to gain insights into biological processes and pathways associated with the observed expression changes. The topGO package in R was used for this purpose.

The raw counts obtained from the RNA-seq analysis were used as input for the MFuzz clustering. These count data represented the expression levels of genes across the three genotypes and three light conditions.

For the clustering, the count data were first normalized (TMM method) to account for differences in sequencing depth and library size. MFuzz, an R package specifically designed for model-based clustering of high-dimensional gene expression data, was utilized for clustering analysis. The resulting 17 gene clusters were visually inspected to find the ones with the wished expression patterns across the samples. Genes belongind to clusters of interest were extracted and subjected to gene ontology (GO) enrichment analysis.

The reads were visualised on IGV (Integrative Genome Viewer) from the BAM mapping files generated.

Preparation of samples and iron concentration measurements

WT, t-athb2 and 35S::ATHB2miP seedlings (about 200 per genotype) were grown on standard MS medium for 12 days. They were carefully removed from the medium to avoid damage and the aerial part of each seedling was excised from the roots. In all steps, no iron nor glass tools were used. The roots and shoots were rinsed in a 2 mM $CaCl_2$ solution to remove iron accumulated on the surface. The material was placed in an Eppendorf tube and dried in a stove at 70°C for 3 days. Each of the dry samples was weighed (dry root weights ranged from 2 to 6 mg, seedlings from 15 to 25 mg). Samples were digested in 2.5 mL of 70% HNO3 + 1 mL of H2O2. Digestion conditions: maximum temperature = 240°C; maximum pressure = 200 bar; cycle duration: 1-2 hours. Following digestion, the samples were diluted in ultrapure Milli-Q water to achieve a [HNO3] = 3.5%. The liquid samples were then analyzed using an ICP-QQQ-MS 8800 Agilent instrument. NIST certified reference materials were employed to verify the accuracy of the data.

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.

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).

Note

This reviewed preprint has been updated to use the correct text; previously, due to a productione error, content from an unrelated submission was presented here.

Supplementary figures

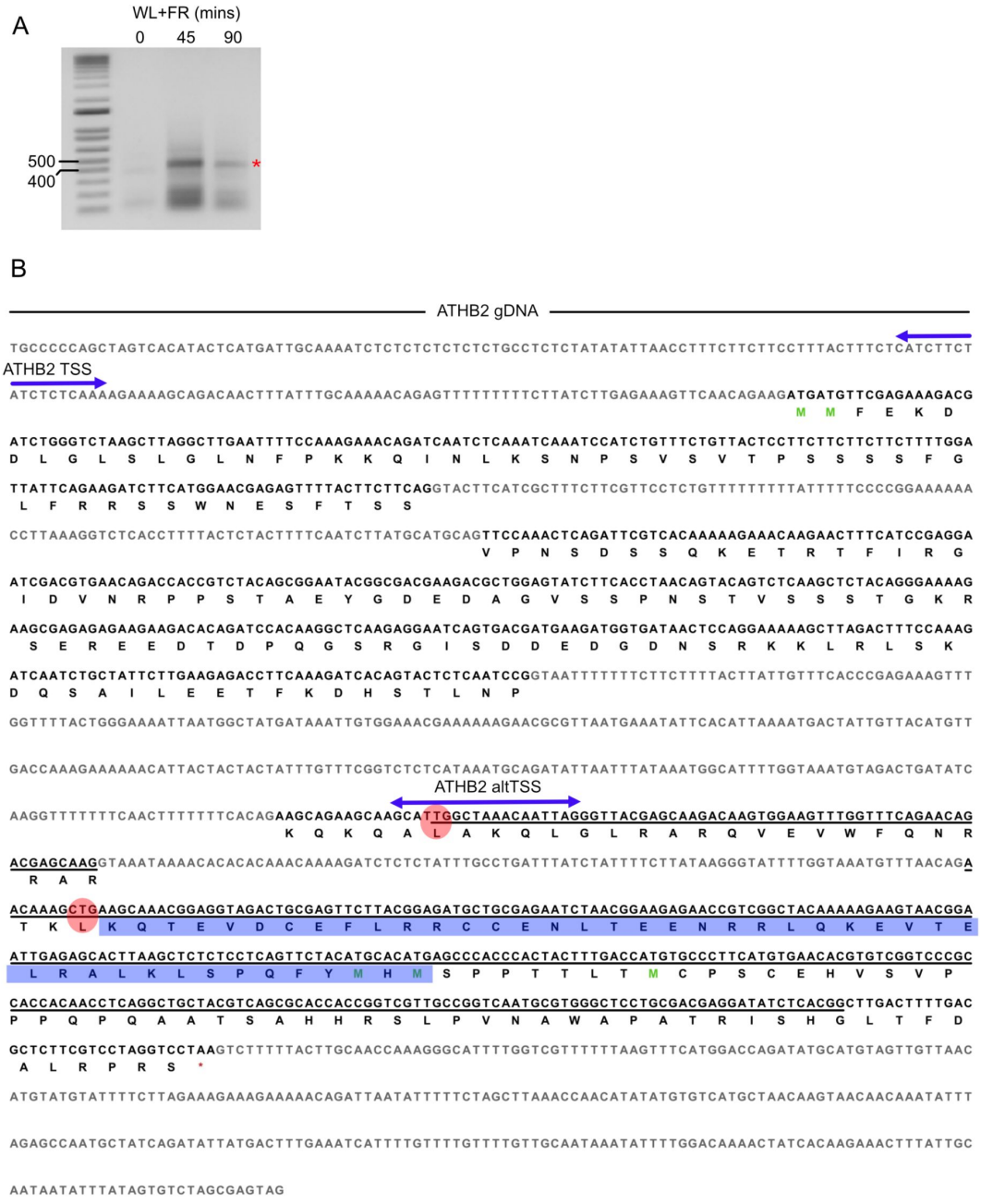

Figure S1.

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.

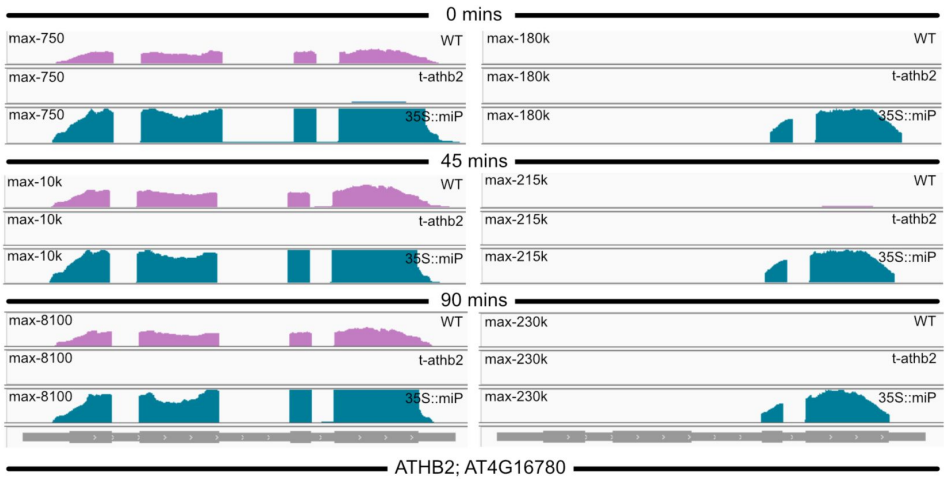

Figure S2.

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.

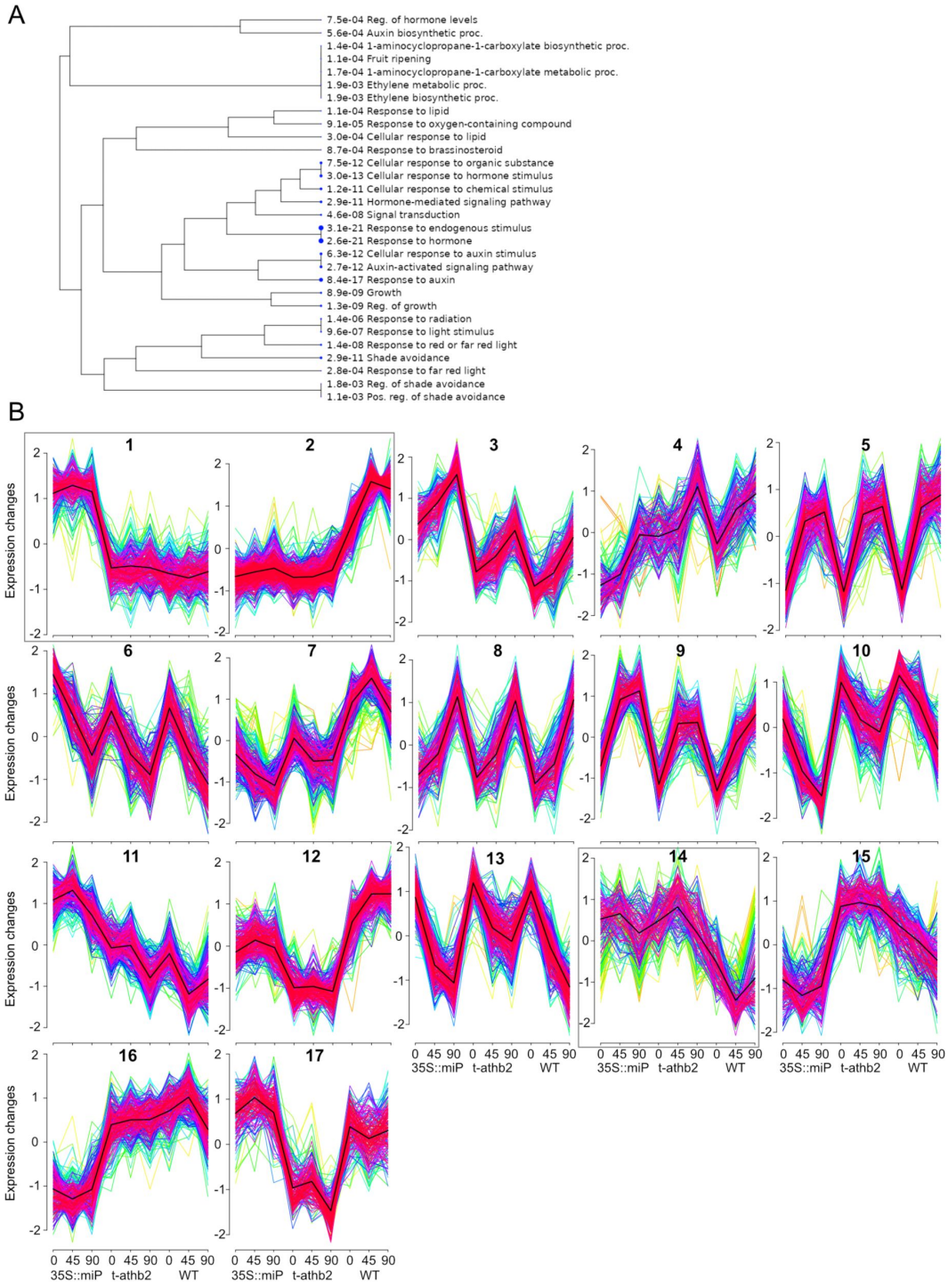

Figure S3.

(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.

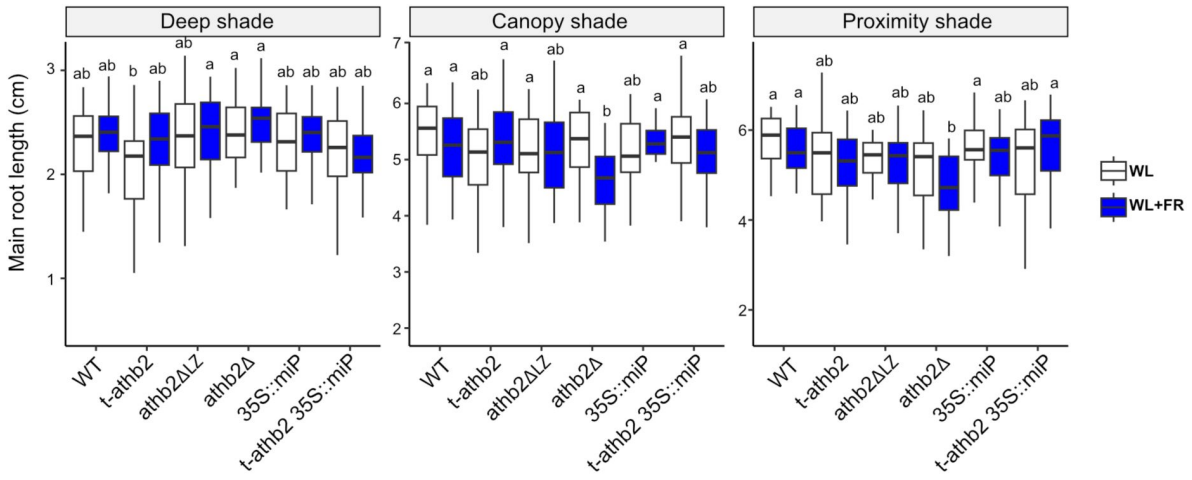

Figure S4.

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.

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

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 *et al.* (2020) **Light affects tissue patterning of the hypocotyl in the shade-avoidance response** *PLoS Genet* **16**

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 *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 *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*

Chong C *et al.* (2020) **Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes** *Nat Commun* **11**

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

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

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

Finn RD *et al.* (2016) **The Pfam protein families database: towards a more sustainable future** *Nucleic Acids Res* **44**:D279–285

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

Floyd SK *et al.* (2014) **Origin of a novel regulatory module by duplication and degeneration of an ancient plant transcription factor** *Mol Phylogenet Evol* **81**:159–173

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

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

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

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

Hanada K *et al.* (2013) **Small open reading frames associated with morphogenesis are hidden in plant genomes** *Proc Natl Acad Sci U S A* **110**:2395–2400

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

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

Kim Y-S *et al.* (2008) **HD-ZIP III Activity Is Modulated by Competitive Inhibitors via a Feedback Loop in Arabidopsis Shoot Apical Meristem Development** *Plant Cell* **20**:920–933

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

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

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

Lamesch P *et al.* (2012) **The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools** *Nucleic Acids Res* **40**:D1202–1210

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

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

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

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

Martin M (2011) **Cutadapt removes adapter sequences from high-throughput sequencing reads**

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

Merelo P *et al.* (2016) **Regulation of MIR165/166 by class II and class III homeodomain leucine zipper proteins establishes leaf polarity** *Proc Natl Acad Sci U S A* **113**:11973–11978

Mudge JM *et al.* (2022) **Standardized annotation of translated open reading frames** *Nat Biotechnol* **40**:994–999

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

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

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

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

Pedmale UV *et al.* (2016) **Cryptochromes Interact Directly with PIFs to Control Plant Growth in Limiting Blue Light** *Cell* **164**:233–245

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

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

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

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

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

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

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

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

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

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

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

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

Tao Y *et al.* (2008) **Rapid synthesis of auxin via a new tryptophan-dependent pathway is required for shade avoidance in plants** *Cell* **133**:164–176

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

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

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

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

Zhang S *et al.* (2020) **Mitochondrial peptide BRAWNIN is essential for vertebrate respiratory complex III assembly** *Nat Commun* **11**

Article and author information

Ashleigh Edwards
University of Copenhagen, Department of Plant and Environmental Sciences, Thorvaldsensvej 40, 1871 Frederiksberg, Denmark, Department of Plant Physiology, Umeå Plant Science Centre, Umeå University, SE-901 87, Umeå, Sweden

Maurizio Junior Chiurazzi
University of Copenhagen, Department of Plant and Environmental Sciences, Thorvaldsensvej 40, 1871 Frederiksberg, Denmark, Department of Plant Physiology, Umeå Plant Science Centre, Umeå University, SE-901 87, Umeå, Sweden

Anko Blaakmeer
University of Copenhagen, Department of Plant and Environmental Sciences, Thorvaldsensvej 40, 1871 Frederiksberg, Denmark, current address: Carlsberg Research Laboratory, J.C. Jacobsens Gade 4, Copenhagen V DK-1799, Denmark
ORCID iD: 0000-0002-1834-2573

Ylenia Vittozzi
University of Copenhagen, Department of Plant and Environmental Sciences, Thorvaldsensvej 40, 1871 Frederiksberg, Denmark, Department of Plant Physiology, Umeå Plant Science Centre, Umeå University, SE-901 87, Umeå, Sweden

Ashish Sharma
University of Copenhagen, Department of Plant and Environmental Sciences, Thorvaldsensvej 40, 1871 Frederiksberg, Denmark, current address: CSIR, Central Institute of Medicinal and Aromatic Plants, Lucknow, India

Sanne Matton
University of Copenhagen, Department of Plant and Environmental Sciences, Thorvaldsensvej 40, 1871 Frederiksberg, Denmark, current address: Department of Plant Science, Wagening University, The Netherlands

Valdeko Kruusvee
University of Copenhagen, Department of Plant and Environmental Sciences, Thorvaldsensvej 40, 1871 Frederiksberg, Denmark, Department of Plant Physiology, Umeå Plant Science Centre, Umeå University, SE-901 87, Umeå, Sweden

Daniel Straub
University of Copenhagen, Department of Plant and Environmental Sciences, Thorvaldsensvej 40, 1871 Frederiksberg, Denmark, current address: Quantitative Biology Center (QBiC), University of Tübingen, Auf der Morgenstelle, Tübingen, Germany
ORCID iD: 0000-0002-2553-0660

Giovanna Sessa
Institute of Molecular Biology and Pathology, National Research Council, P.le A. Moro 5, 00185 Rome, Italy

Monica Carabelli
Institute of Molecular Biology and Pathology, National Research Council, P.le A. Moro 5, 00185 Rome, Italy

Giorgio Morelli
Research Centre for Genomics and Bioinformatics, Council for Agricultural Research and Economics (CREA), Via Ardeatina 546, 00178 Rome, Italy

Stephan Wenkel
University of Copenhagen, Department of Plant and Environmental Sciences, Thorvaldsensvej 40, 1871 Frederiksberg, Denmark, Department of Plant Physiology, Umeå Plant Science Centre, Umeå University, SE-901 87, Umeå, Sweden
For correspondence: stephan.wenkel@umu.se
ORCID iD: 0000-0001-5764-9423

Copyright

Editors

Reviewing Editor
Dominique Bergmann
Stanford University, Stanford, United States of America

Senior Editor
Jürgen Kleine-Vehn
University of Freiburg, Freiburg, Germany

Reviewer #1 (Public Review):

Summary:

The authors report evidence for a microprotein of AtHB2-miP. The authors came across HB2 in a screen for alternative transcription start sites in Arabidopsis in response to white light or a white light followed by a far red light representative of shade. Out of 337 potential microproteins, authors selected AtHB2. At the beginning of the manuscript, it is investigated that an alternative transcription start site of HB2 gene can be used in response to far red light. The resulting shorter protein form seems to interact with HB2 protein forms, altering the localization of HB2 in transient expression assays. The functionality of HB2-miP overexpression has been addressed in transgenic Arabidopsis lines using a 35S promoter. The responses and phenotypes were compared with either WT or various types of athb2 mutant lines with disrupted HB2 gene. Such mutants and the 35S promoter-driven AtHB2-miP line showed various types of phenotypes versus each other that can be classified as mild or none, e.g. small effects on root growth, iron homeostasis gene expression, and iron contents.

Strengths:

The authors performed an interesting screen for alternative transcription start sites which resulted in 337 candidates (Figure 1A). Principally, it can be interesting to find that plants may use alternative start sites for HB2 in response to shading light. The authors provide evidence that alternative transcription start sites of HB2 can be present and used in response to FR. The possibility that potentially resulting small protein may have effects under FR light, causing alteration of root growth and physiology, is an interesting idea.

Weaknesses:

In the present manuscript, there are several signs of incomplete analysis.

(1) The transient expression experiments are not conducted with much detail to demonstrate that indeed HB2 miP is produced and can interact with regular protein. The localization of HB2 was found to be linked with condensates, but perhaps not in the presence of HB2 miP. Clearly, the lack of quantitative and qualitative analysis hampers a clear assessment of this point.

(2) The authors, unfortunately, did not provide the data of the screen to demonstrate which concrete candidates may have miPs and whether there is enrichment of certain functions. There is no supplemental table accompanying Figure 1A.

(3) One of the major unclear points that is also not addressed in the discussion is that the function of miR is studied in overexpression plants (35S promoter::miP). The effects are only compared to wild type and various lines of HB2 knockouts or knockdowns, partly with fairly uncharacterized phenotypes. It can now not be clearly determined whether the miP effects are due to a regular function of miP or due to overexpression of it. A needed control would be a 35S::AtHB2 line, or better at least two different lines (only a single miP overexpression line investigated). Since it has not been assessed by deletion mutant analysis to determine which protein parts of miP are involved in the protein regulation, it cannot be ruled out that the observed miP effects are not naturally occurring but the result of ectopic expression of a protein. Clearly, the effect of miP would be ideally studied in an environment where the levels can be controlled and the resulting phenotypes and protein levels quantified.

(4) It is not shown that the microprotein is generated in Arabidopsis in response to shade, e.g. through Western or fluorescence protein detection. The main idea that authors want to claim, namely that miP binds with regular protein and thereby controls its localization or activity has not been addressed in Arabidopsis. There are no localization experiments of HB2 protein data in the presence of miP in Arabidopsis.

(5) The plants with altered HB2 forms seem to grow well and the recorded phenotypes are rather minor. Photos are not shown. At some point, the authors discuss that there could be redundancy or that HB miP might interact with other HB proteins. However, such protein interactions have not been experimentally investigated.

https://doi.org/10.7554/eLife.96725.1.sa2

Reviewer #2 (Public Review):

The first portion of the manuscript centered on identifying and confirming the ATHB2 microprotein (ATHB2miP), which constitutes the core message of this study. Overall, I find no issue with the selection criteria employed for identifying alternative microprotein mRNA transcripts. However, I do have some queries that I hope the authors can address for clarity.

(1) Upon reviewing the supplemental dataset where the authors listed the 377 unique novel miPs, along with those specifically in WL or shade treatments, I sought to comprehend the

rationale behind focusing on ATHB2. Have the authors examined the shade response of all 377 potential microprotein candidates? Readers may be intrigued to learn how many of these candidates exhibit induction or repression under shade conditions, and whether such changes correlate positively or negatively with alterations in the full-length TSSs in response to shade. Essentially, I aim to discern the prevalence of microprotein production during shade responses and any shared characteristics among these microprotein transcripts. This inquiry also aims to uncover the existence of a common mechanism regulating microprotein transcription.

(2) To confirm that ATHB2miP stems from an independent transcription event, the authors sequenced full-length cDNAs using PacBio isoseq. However, I find the information regarding isoseq missing from the manuscript. My assumption is that the full-length cDNAs were reverse transcribed from mRNAs isolated from whole seedlings, where mature mRNAs in the cytoplasm predominate, making it challenging to evaluate whether a specific mRNA undergoes post-transcriptional processing. One approach to confirming ATHB2miP as a product of independent transcription involves examining nascent mRNA produced in the nucleus. The authors may need to isolate nascent mRNAs associated with RNA Polymerase II in the nucleus from seedlings treated with shade for 45 min, and then perform reverse transcription and PacBio isoseq.

(3) The authors noted the identification of two potential start codons, TTG and CTG, in the alternative TSS of ATHB2 using TISpredictor. Yet, it's imperative to identify the actual translation initiation site and the full-length sequence of ATHB2miP. I suggest the authors fuse an epitope tag (e.g., 3xFLAG) to the C-terminus of ATHB2 (utilizing the genomic sequence of ATHB2) and generate transgenic lines to be treated with shade to induce ATHB2miP-3xFLAG production. Affinity purification (anti-FLAG beads) and mass spectrometry can then identify the actual start site of ATHB2miP. This step is crucial, as the current ATHB2miP used may not be the exact sequence, and any observed phenotype could be artifacts arising from these lines.

(4) My confusion arose when analyzing the results in Figures 1E - G. The authors didn't specify whether these plants were subjected to shade treatment. What are the sequences within the second intron and third exon excluded from pATHB2control::GUS that promote transcription and translation? Have the authors examined the sequence features? This information is pivotal and related to the above question #1 because it may tell us whether the sequence feature is shared by other miP candidates.

The latter part of the manuscript focused on the functional characterization of ATHB2miP. The approaches adopted by the authors resemble those used in studying antimorphic (dominant negative) alleles. However, I have several concerns regarding the approaches and conclusions.

(5) Firstly, as mentioned in question #3, the authors did not map the actual translation initiation site of ATHB2miP. Therefore, all constructs involving ATHB2miP, such as eGFP-ATHB2miP, BD-ATHB2miP, and mCherry-ATHB2miP in Figure 2, and 35S::miP in Figures 3-5, may contain extra amino acids in the N-terminus, given that epitope tags were all added to the N terminus. These additional amino acids could potentially impact the behavior of ATHB2miP and lead to artifacts. Identifying the translation initiation site in ATHB2miP would facilitate the development of tools to disrupt ATHB2miP expression without affecting full-length ATHB2 expression. For instance, if the "CTG" before the leucine zipper domain is confirmed as the translation initiation site, mutating it to another Leu codon (e.g., TTA) could generate transgenic lines using the genomic sequence of ATHB2, including this mutation, to evaluate the impact of losing ATHB2miP on shade responses.

(6) Another concern pertains to the 35S::miP line utilized in Figures 3-5. The authors only presented results from one 35S::miP line, raising the possibility of T-DNA insertion disrupting

an endogenous gene in the transgenic plant genome. It is essential to clarify how many individual T1 plants were generated and how many of them showed the same phenotype as the line used in the manuscript. Additionally, the use of the constitutive CaMV35S promoter could generate artifacts akin to neomorphic mutations. For example, the authors identified Cluster 1 genes that were only induced in 35S::miP, but not in t-athb2 or WT plants (Figure 3B); moreover, they found an overrepresentation of genes involved in root development in this cluster. This observation correlated well with the root phenotype of 35S::miP under the proximity shade (Figure 4D), in which the short-root phenotype was only observed in lines expressing 35S::miP. These data could be artifacts due to the constitutive expression of ATHB2miP in roots but didn't necessarily reflect the natural function of ATHB2miP.

(7) Furthermore, I seek clarification regarding the rationale behind employing different shade conditions, including deep shade, canopy shade, and proximity shade, and the significance of treating plants with these conditions. The results were challenging to interpret, and I have reservations about some statements made. The authors claimed that ATHB2 acts as a growth repressor in deep shade but a growth promoter in the canopy and proximity shade (Lines 366-368). However, it appears that regardless of the shade conditions, most mutant and transgenic lines were not significantly different from WT (Figure 4C). Additionally, the definition of proximity shade in this manuscript (R:FR = 0.06) differs from that in Roig-Villanova & Martinez-Garcia (Front. Plant Sci., 2016; R:FR, 0.5-0.3). Clarity on this disparity would be appreciated.

(8) In Figure 5, no statistical analyses were presented in Figure 5C. It remains unclear whether the differences observed are statistically significant. Moreover, the values appear quite similar among all three genotypes. Even if statistically significant, do these minor differences in Fe concentrations significantly impact plant physiology? Additionally, some statements related to Figure 5 do not align with the data presented. For instance, claims about longer hypocotyls in t-athb2, athb2Δ, and atbh2ΔLZ mutants compared to wild type under shade conditions on high iron media (lines 453-455) were not supported by the data in Figure 5D. Similarly, statements about the differences between mutants (lines 458-460) were not substantiated by the data.

https://doi.org/10.7554/eLife.96725.1.sa1

Reviewer #3 (Public Review):

Summary and Strengths:

In this interesting manuscript, the authors identify a large number of alternative transcription start sites (TSS) and focus their functional analysis on an alternative TSS that is expected to produce a micro-protein (miP) encoding the C-terminus of ATHB2 (ATHB2miP). ATHB2miP is expected to comprise the leucine zipper part of ATHB2 and hence interact with the full-length protein through this dimerization motif. Such interactions are shown using yeast two-hybrid and FRET-FLIM assays. ATHB2 is a well-known shade-induced gene that has been implicated in shade-regulated growth responses. The authors then test the potential role for ATHB2miP genetically by comparing several athb2 loss-of-function (LOF) alleles: one does not express either full-length ATHB2 or the short ATHB2miP (t-ATHB2), two CRISPR alleles give rise to frameshift mutations in the full-length transcript but still express a potentially functional short ATHB2miP (athb2deltaLZ and athb2delta). The authors also use plants that over and ectopically express ATHB2miP (35S:miP). Overall, the results are consistent with the hypothesis that ATHB2miP inhibits the function of ATHB2, which constitutes a novel negative feedback loop. Potentially ATHB2miP may also inhibit the activity of other related HD ZIP proteins (based on 35S:miP). The effects of these genetic alterations on shade-regulated hypocotyl growth are relatively modest. Effects on root growth are also investigated and in one intriguing case, the negative feedback model does not appear to explain the data (Figure

4D, effect on lateral roots, because for this phenotype 35S:miP is very different from the lof alleles). The authors also identify a potentially interesting link between shade-regulated hypocotyl growth and iron uptake. A number of text changes and corrections to the figures would be important for clarity. They primarily concern three issues: names of the alleles, names of the studied shade conditions, and statements about significant differences between genotypes. Also, it would be interesting to know whether the effects of ATHB2 on iron uptake are due to local effects of ATHB2. Is ATHB2 expressed in roots?

Weaknesses:

(1) The naming of the different shade conditions is difficult to follow and not consistent with the way most authors in the field call such conditions. Deep shade is ok (low PAR and low R/FR, WL, PAR 13microE, R/FR 0.13). This condition is clearly defined for experiments in Figure 4. However, data in Figure 1 also use Deep shade (line 174) but PAR is not defined there. I suggest that all light conditions are clearly defined in the figure legends and in the M&M (not the case in this ms). Regarding Canopy shade (WL, PAR 45microE, R/FR 0.15) and proximity shade (WL, PAR 45microE, R/FR 0.06), see lines 355-357, this nomenclature is unclear. First proximity shade has a higher R/FR ratio than canopy shade. Second for canopy shade (compared to the WL control) PAR should decrease which is not what is done here. What is called proximity shade and canopy shade are 2 WL conditions with different R/FR ratios, which are compared to WL controls with the same PAR. It would make more sense to call them proximity shade and indicate the different R/FR ratios. Finally, extensive literature from many plant species and numerous labs has shown that hypocotyl elongation increases with R/FR decreasing. In the data shown in Figure 4, it is the opposite. Hypocotyls in Canopy shade (WL, PAR 45microE, R/FR 0.15) are longer than those in proximity shade (WL, PAR 45microE, R/FR 0.06), while with these R/FR ratios the opposite is expected. Could this be a mistake in the text? Please check.

(2) In several instances (in particular regarding data from Figures 4 and 5), the authors write that 2 genotypes are significantly different while the statistical analysis of the data does not support such statements. For example lines 392-395, the authors write that in WL the t-DNA mutant, both CRISPR mutants and 35S:miP lines all had significantly lower number of lateral roots than the WT. This is true for the t-DNA mutant (group bc, while the WT is in group a), however, all other genotypes are in group ab, hence not significantly different from the WT. Please carefully check all such statements about significant differences.

(3) The naming of the CRISPR mutants is problematic. In particular athb2delta, such a name suggests that the gene is deleted (also suggested by Figure 4A), which is not the case in this CRISPR allele leading to a frameshift early in the coding sequence. This is particularly problematic because in this allele ATHB2miP is still expressed, while based on such a name one would expect that in this mutant both the full length and the miP are lost. Both CRISPR alleles lead to a frameshift and this should be clarified in Figure 4A and in the text.

(4) Overall hypocotyl growth phenotypes of athb2 lof mutants and 35S:miP are similar and consistent with a model according to which ATHB2miP inhibits the full-length protein. However, this is not the case for the root phenotype described in 4D. It would be interesting to discuss this.

(5) The authors propose a role for ATHB2 in the root, in particular linked to iron uptake. Is this due to a local effect of ATHB2 in the roots? Is ATHB2 expressed in roots? It would be very informative if the authors would show such data, e.g. using the reporter lines used in Figure 1. Are both the FL and the miP expressed in roots?

(6) From the description regarding 5'PEAT.seq data presented in Figure 1 (see lines 174-177) it is not clear in which light conditions the seedlings were grown. It appears that samples were collected in 3 conditions. WL and after 45 and 90 minutes of low R/FR treatment. However,

then the data is discussed collectively. Does the 12398 TSS correspond to what was found in all three conditions together? Are the authors showing shade-regulation of TSS? This is clearly the case for ATHB2miP. This needs to be clarified.

(7) The way gene expression of low F/FR effects is done might conflate circadian effects and low R/FR effects because the samples from different light conditions are not collected at the same ZT. This is how I understood the text. If I'm wrong please clarify the text. If I am right, this potential problem should be mentioned in the text.

(8) Could the authors envisage a way to genetically test the role of ATHB2miP by using an allele that makes the full length but not the miP? Currently, the authors use lof alleles that either make none of the transcripts (t-DNA) or potentially only the miP (CRISPR alleles). Overall, these alleles do not appear to differ in their phenotypes, suggesting that most of the effect of ATHB2miP is through ATHB2 FL. Having an allele only producing the FL would be nice (but technically I'm not sure how one could do that).

https://doi.org/10.7554/eLife.96725.1.sa0