TTG1 proteins regulate circadian activity as well as epidermal cell fate and pigmentation

Chiara A. Airoldi, Timothy J. Hearn, Samuel F. Brockington, Alex A. R. Webb* and Beverley J. Glover*

Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge, CB2 3EA, UK

*For correspondence: bjg26@cam.ac.uk, 44 1223 333938; aarw2@cam.ac.uk, 44 1223 333948

Abstract

The Arabidopsis genome contains three members of the *TTG1 (TRANSPARENT TESTA GLABRA 1)* WDR subgroup of the WDR family, with very different reported roles. *TTG1* is a regulator of epidermal cell differentiation, and of the production of pigments, while *LWD1* (*LIGHT-REGULATED WD1*) and *LWD2* (*LIGHT-REGULATED WD2*) are regulators of the circadian clock. We discovered a new central role for TTG1 WDR proteins as regulators of the circadian system, demonstrated by a lack of detectable circadian rhythms in a triple *lwd1lwd2ttg1* mutant. We have demonstrated that there has been subfunctionalisation by protein changes within the angiosperms, with some TTG1 WDR proteins developing a stronger role in circadian clock regulation while losing the protein characteristics essential for pigment production and epidermal cell specification, and others weakening their ability to drive circadian clock regulation. Our work demonstrates that even where proteins are very conserved, small changes can drive big functional differences.

Introduction

A central paradigm in evolutionary developmental biology is that functional novelty arises through changes to the regulation and expression, both spatially and temporally, of otherwise well

conserved proteins[1,2,3]. Evolution of coding sequences is considered unlikely to occur, as mutations to essential proteins are most likely to be deleterious[3]. This is particularly true of transcription factors and other proteins involved in transcriptional regulation, as the loss of a protein necessary for the downstream regulation of multiple target genes is unlikely to be selectively advantageous. We have been investigating the WD-repeat (WDR) protein family to understand how new functions can arise in highly conserved families of transcriptional regulators. The WDR protein family plays an important role in the transcriptional regulation of many processes in plants. WDR proteins are a family in the β propeller protein group characterized by the presence of a 40 residue core region delineated by a glycine-histidine (GH) dipeptide and a tryptophan-aspartate (WD) dipeptide[4]. This motif is repeated in tandem between four and sixteen times in each protein. WDR proteins facilitate protein-protein interactions but have no direct DNA binding activity[5].

In Arabidopsis, extensive research has demonstrated that the WDR protein TTG1 is central to all aspects of epidermal cell fate through its role in forming the MBW complex with MYB and bHLH transcription factors. TTG1 positively regulates trichome formation, anthocyanin production, seed coat pigmentation and seed coat mucilage production, and negatively regulates root hair formation[5,6,7,8,9]. In the MBW complex the WDR protein functions as a scaffold, on which the DNA-binding MYB and bHLH proteins interact to generate the transcriptional complex, and in some cases WDR and MYB proteins compete to bind the bHLH protein, with bHLH-WDR dimers activating a different cascade of downstream genes to bHLH-MYB dimers[10,11].

More recently two further genes belonging to this small WDR subfamily have been characterized in Arabidopsis, and named *LIGHT REGULATED WD1* (*LWD1)* and *LWD2*[12,13]. These proteins function in transcriptional regulation of the central circadian clock component *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*). The *lwd1lwd2* double mutant has a short circadian period and the plants are early flowering. Regulation is achieved in part by binding to TEOSINTE BRANCHED 1-CYCLOIDEA-PCF20 (TCP20) and TCP22[14], which promote expression of

CCA1. The *LWD1* and *LWD2* genes are fully redundant, and a mutant phenotype is only revealed when both genes are mutated. Their sequence similarity to *TTG1* could suggest that these very different functions in the plant are the result of changes to the regulation of these WDR genes, rather than functional changes to the proteins they encode.

We show here that local changes in WDR function in Arabidopsis family members can be attributed to mutation of key residues followed by positive selection. Combinatorial mutant analysis reveals the central role of the TTG1 WDR protein family, including TTG1 itself, in circadian clock regulation. These results shed new light on protein functional evolution through small changes and point to a much more significant role than previously suspected for this particular protein family in circadian regulation.

Results

Protein function diverges in the TTG1-like WDR protein clade

LWD1 and LWD2 are the only two WDR proteins in Arabidopsis that are closely related to TTG1, although our phylogenetic analysis suggests that TTG1 and the LWD proteins fall into two clades resulting from a gene duplication event at the base of the seed plant lineage (Fig. 1a, Supplementary Fig. 1). Despite their sequence similarity to TTG1 (Extended Data Fig. 1), LWD1 and LWD2 have been described to have a very different function to TTG1, acting as scaffolds for the transcriptional regulators functioning in the circadian oscillator[14]. We were interested to determine whether this different function of proteins with such high similarity depends on differential regulation. To test this hypothesis, we first investigated whether these two proteins have similar function to TTG1 through a transgenic rescue test by ectopic expression in the *ttg1-1* mutant. We found that LWD1 and LWD2 did not rescue pigment absence in the leaves and in the seeds of the Arabidopsis *ttg1-1* mutant, nor were these genes capable of rescuing the root hair and trichome phenotypes (Fig. 1b, 1c; Extended Data Fig. 2, 3). We assessed the transcript levels of

LWD1 and *LWD2* in these transgenic lines and confirmed that the transgene is expressed (Extended Data Fig. 4).

Combinatorial mutant analysis provides new insights into the function and evolution of the TTG1 family

The apparent lack of TTG1-like function demonstrates that LWD1/2 and the closely related TTG1 have different protein activities that could have arisen as a result of evolution in LWD1/2 or TTG1 of a new functional role specific to the circadian clock or epidermal cell characteristics (neofunctionalization) or could be the effect of subfunctionalization and division of roles in the TTG1 family. To address this question, it was necessary to investigate whether TTG1 can have LWD1/2-like function in circadian clock regulation. The *ttg1-21* mutant has no effect on circadian period (Fig. 2a, 3a); however this does not demonstrate that TTG1 is without function in the circadian oscillator because that role might be masked by the presence of functional LWD1 and LWD2 proteins in *ttg1-21* mutant plants. To address the level of functional redundancy between *TTG1, LWD1* and *LWD2* genes and the possible involvement of TTG1 in circadian clock regulation, we crossed the *ttg1-21* Arabidopsis Columbia mutant with the double *lwd1lwd2* mutant [20]. We performed leaf movement and *CCA1:LUC* reporter gene assays and measured a severe impairment of the circadian clock in the triple mutant line (Fig. 2a, b). The *lwd1lwd2ttg1* triple mutant was arrhythmic in constant light with all replicates having an RAE > 0.5 with FFT-NLLS analysis for *CCA1:LUC* rhythms (Fig. 2b) and 60 of 83 triple mutant plants were arrhythmic for leaf movement (RAE > 0.5; Fig. 2a). Consistent with these results, the *lwd1lwd2ttg1* triple mutant was also very delayed in flower induction (Extended Data Fig. 5a, b). These data demonstrate that the TTG1 protein is necessary for circadian clock function in the absence of LWD1 and LWD2, and when all are absent there is a loss of circadian rhythms comparable to that for loss of the major components of the central oscillator (e.g. *prr5,7,9* triple mutants[15]). We conclude that TTG1 is capable of regulating circadian activity, therefore circadian function is not newly acquired and unique to the LWD1/LWD2 clade.

We reasoned that the modifications in LWD1 and LWD2 proteins that have restricted their function to the circadian clock might have improved their ability to participate in this process. Therefore, TTG1 would not be expected to be as effective in affecting circadian rhythms as are LWD1 and LWD2. We used single and double mutant combinations to test this hypothesis. We measured circadian function using *CCA1:LUC* expression in all single and double mutant combinations *ttg1, lwd1, lwd2, lwd1ttg1, lwd2ttg1* and *lwd1lwd2* (Fig. 3). If TTG1 is functionally equivalent to LWD1 and LWD2 in the circadian clock, the effect we observe in the double mutant *lwd1lwd2* would be present also in *lwd1ttg1* or *lwd2ttg1* mutants. In the single mutants *ttg1* (period: 23.4 ± 0.3 h, n = 11) and *lwd2* (24.1 ± 0.2 h, n = 11) we do not observe a strong effect on the period of the central oscillator compared to Col-0. *lwd1* was significantly reduced in period compared to Col-0 (Col-0 25.0 ± 0.8 h, *lwd1* 23.0 ± 0.6 h; t-value = 7.12, df=20, p< 0.00001) (Fig. 3, Extended Data Fig. 6a) and the *lwd1lwd2* double mutant had an extremely short circadian period (17.0 ± 0.2 h, T=22.72, df=14, p < 0.00001), confirming previous studies[12]. In contrast, double mutants between *ttg1* and *lwd1* or *lwd2* had very little effect on circadian period (Fig. 3, *lwd1ttg1* period 24.86 ± 0.34 h, T=0.16, df=21, p=0.44; *lwd2ttg1* period 24.34 ± 0.095 h, T =1.5, df=20, p=0.08). In line with the effects on circadian period, only the *lwd1lwd2* double mutant affected flowering time (Extended Data Fig. 5c). These data show that TTG1 is not as competent to affect circadian activity as are *LWD1* and *LWD2*. The small effect of *lwd1* was not detected in a *lwd1ttg1* double mutant (Fig. 3 and Extended Data Fig. 6a). This might provide evidence that TTG1 competes with LWD2 to regulate similar promoter sites, and that in the absence of TTG1, LWD2 can complement the effects of *lwd1* loss of function. Alternatively, it is possible that changes in the expression of *LWD2* in *ttg1* loss of function lines might affect circadian period. However, we found no differences in *LWD2* expression in Col-0 versus the *lwd1ttg1* mutant (Extended Data Fig. 6b), although we note that unexplored post-transcriptional differences might explain this result.

In addition to the circadian clock phenotype the triple mutants had several consistent morphological defects. Leaf shape was severely altered with increased serration and consistently

shorter leaves. Furthermore, the triple mutants had early signs of senescence localized at the tip of the leaves (Extended Data Fig. 7a, 7b). These defects were not present in any of the double mutant combinations we analysed (Extended Data Fig. 7a). Trichomes on the leaf margins were present in the *ttg1* mutant plants but their number was significantly reduced in the triple mutant and there was a small decrease in the number of trichomes present in the double mutants *lwd1ttg1* and *lwd2 ttg1* (Extended Data Fig.7c, p values in Extended Data Fig. 7d).

TTG1-like WDR protein functional capability is conserved across land plants

To investigate the evolution of the TTG1 WDR proteins we reconstructed the relationships between proteins inferred from sequences derived from a combination of genomes and transcriptomes from extant land plants (Supplementary Fig. 1). Our phylogenetic analyses identified that a gene duplication event had occurred, inferred to have arisen following the divergence of the Monilophytes. The gene duplication gave rise to two major clades, each containing representative proteins derived from gymnosperms and angiosperms, one including the *Arabidopsis thaliana TTG1* (and termed the *TTG1* clade), and the other including the *Arabidopsis thaliana LWD1* and *LWD2* genes (and termed the *LWD1/2* clade). On the basis of *loci* from fully sequenced genomes, all gymnosperms and angiosperms are inferred to have at least one gene copy from the *TTG1* clade, and at least one gene copy from the *LDW1/2* clade. Additional deep level duplications are inferred to have occurred within the 'Bryophytes' *sensu lato*, with three copies of TTG1 WDR proteins present in *Marchantia polymorpha*. Of these, MpWDR3 and its respective homologs are on a very long branch in comparison to the other two *Marchantia* sequences (Supplementary Fig. 1).

Focusing on early diverging land plants we isolated the three *Marchantia polymorpha TTG1 WDR* genes *MpWDR1, MpWDR2* and *MpWDR3*, and ectopically expressed them (using the CaMV 35S promoter) in the Arabidopsis *ttg1-1* mutant to address the extent of functional conservation between these three proteins (Extended Data Fig. 1). We used ectopic expression of *AtTTG1* (from the same promoter) as a positive control for the rescue of the *ttg1-1* phenotype. Mutant plants

expressing *MpWDR1* and plants expressing *MpWDR2* are capable of producing pro-anthocyanidin in the seed coat (Fig. 4). We also observed production of anthocyanins in leaves, normally absent in the *ttg1-1* mutant (Extended Data Fig. 2). Furthermore, 35S:*MpWDR1* and 35S:*MpWDR2* can also rescue the absence of trichomes on the leaf blade seen in the *ttg1-1* mutant (Fig. 4), and the increased number of trichoblasts (root hairs) found on the root (Extended Data Fig. 3). However, 35S:*MpWDR3* expression in the *ttg1-1* mutant did not rescue any of these defects to the extent of being approximately equal to the wild type (Fig. 4, Extended Data Fig. 2, 3). The *MpWDR3* gene sits on a very long branch compared to the other two *Marchantia TTG1* genes, as seen in our phylogenetic analysis (Supplementary Fig. 1). Long branches can skew apparent relationships in phylogenetic trees, but they also suggest faster mutation rates and the potential for gain or loss of function relative to *MpWDR1/2*. This faster evolution might explain the loss of the protein characteristics necessary to rescue the *ttg1-1* mutant phenotypes. Our data show that the protein characteristics essential for pigment regulation, but also those which allow the Rosid-specific evolution of trichome and atrichoblast regulation, are present in proteins in *Marchantia* that diverged 450 MYA from the angiosperm lineage.

The capability of TTG1 proteins to regulate the circadian clock is a remnant of an ancient function

To acquire information about the evolution of the circadian clock function in the TTG1 WDR proteins we used ectopic expression of TTG1 WDR proteins in the triple mutant *lwd1lwd2ttg1*. First, we confirmed that overexpression of *LWD1* in the arrhythmic triple mutant restored circadian rhythms of *CCA1:LUC* (26.3 ± 0.6 h, Fig. 5). This shows that, despite a high and constant expression pattern driven by the 35S promoter, the LWD1 protein is sufficient to restore wild type circadian cycling. This can be explained by the fact that TTG1-WDR proteins act as a scaffold to transcription factors, therefore a high and constant expression is not damaging to circadian clock regulation because the expression profiles of the transcription factors that are part of the complex remain unaltered. This level of protein competence is not present in TTG1: the ectopic expression of *TTG1* in *lwd1lwd2ttg1* mutants resulted in a short circadian period similar to

196 the *lwd1wd2* mutant (20.7 ± 0.8 h) (Fig. 5). We repeated the experiment with several independent
197 lines (expression of transgene confirmed in Extended Data Fig. 4 and Extended Data Fig. 8). We
198 always obtained comparable results, demonstrating that the short circadian period of the transgenic
199 lines is not due to the level of expression of *TTG1*. We conclude that the main reason that TTG1
200 has a reduced ability to rescue circadian clock phenotype defects is that the TTG1 protein is
201 different to the LWD1 protein. In short, the differential abilities of the proteins are attributable to
202 their protein properties, not their expression patterns.
203
204 To assess the likelihood that ancestral TTG1 WDR proteins could regulate the circadian oscillator
205 we explored the function of modern *Marchantia* sequences using transgenic rescue of the
206 *lwd1lwd2ttg1* triple mutant (Tr) with *MpWDR1*, *MpWDR2* and *MpWDR3* driven by the 35S
207 promoter. Unlike TTG1, both MpWDR1 and MpWDR2 could restore near wild type period in
208 *lwd1lwd2ttg1* and were as effective as transgenic rescue with LWD1 (*Tr35S:MpWDR1* 26.5 ± 0.6
209 h, *Tr35S:MpWDR2*: 25.7 ± 0.4, Fig. 6). Furthermore, we also observed rescue of the late
210 flowering time phenotype in all these lines (Extended Data Fig. 9). This result suggests that the
211 circadian clock regulation function of the TTG1 WDR proteins could have been acquired in early
212 diverging land plants. However, once again the more divergent MpWDR3 protein was only
213 partially able to rescue the mutant phenotype (Fig. 6; Extended Data Fig. 9). Despite high levels of
214 expression from the 35S promoter (Extended Data Fig. 8) we did not observe a rescue of the triple
215 mutant flowering time phenotype with this gene or of the circadian phenotype, which remained
216 arrhythmic (Fig. 6, Extended Data Fig. 9).
217
218 **Subfunctionalization within the TTG1 WDR family is present in the earliest diverging**
219 **flowering plants**
220 To explore whether the apparent subfunctionalization of the LWD1/LWD2 proteins occurred late
221 in the lineage including Arabidopsis, we analysed LWD protein function from a species that
222 diverged at the base of the flowering plant phylogenetic tree. The AmLWD protein of *Amborella*
223 *trichopoda*, the earliest diverging extant angiosperm[16], is a member of the LWD1 and LWD2

clade (Fig. 1A, Supplementary Fig. 1). Ectopic expression of *AmLWD* in the *ttg1-1* mutant and the *lwd1lwd2ttg1* triple mutant revealed that AmLWD did not rescue the epidermal defects of the *ttg1-1* mutant (Extended Data Fig. 10) but did rescue both the flowering time and circadian period defects arising from the loss of *LWD1* and *LWD2* (Fig. 6; Extended Data Fig. 9). This is equivalent to the result obtained with LWD1 and LWD2. These results show that LWD clade subfuntionalization is not an isolated event that happened in Arabidopsis but was already present in early diverging angiosperms.

Discussion

Unexpected evidence of a central role for the TTG1 WDR protein family in circadian regulation

Our results demonstrate that the TTG1 protein is capable of regulating the circadian clock and is able to partially rescue the loss of the other two TTG1 WDR proteins (LWD1 and LWD2). The arrhythmic phenotype of the *lwd1lwd2ttg1* triple mutant demonstrated the central role of this protein family in the regulation of the circadian clock. The complete arrhythmia of the triple mutants suggests a central role for TTG1 WDR proteins in the circadian oscillator, and that their importance might have been underestimated, concealed by the presence of the TTG1 protein in the *lwd1lwd2* double mutant. The complete loss of rhythms in the triple mutant, together with the strong late flowering phenotype, is comparable to loss-of-function of the PSEUDO RESPONSE REGULATOR (PRR) transcriptional regulators in the central circadian oscillator[17]; *prr5prr7prr9* mutants are also completely arrhythmic, like *lwd1lwd2ttg1*[15]. In simple terms, loss of *TTG1* function in an already *lwd1lwd2* mutant background completely incapacitates the circadian clock. Our data suggest that the *TTG1 WDR* gene family and the proteins these genes encode are required for rhythm generation. It is not yet clear whether this rhythm generation operates through the known interactions between TTG1-like proteins and transcription factors of the MYB and bHLH families, or whether it occurs through other pathways. Nevertheless, our analysis has shown that the presence of TTG1 in the *lwd1* mutant is detrimental for the circadian clock. The *lwd1ttg1* mutant has a wild type period whereas a single *lwd1* mutant is significantly different from the wild

type. The TTG1 protein is not only less efficient to regulate the circadian clock than LWD1 and LWD2, but its presence could also be detrimental in the absence of LWD1.

An ancient role for TTG1 WDR proteins in circadian regulation

The ability to drive circadian regulation does not seem to be a new function that has been acquired late in land plant evolution. Our transgenic rescue of the triple mutant *lwd1lwd2ttg1* with WDR TTG1 genes from across the land plant phylogenetic tree has shown that not only Arabidopsis proteins but also *Marchantia* and *Amborella* proteins possess the ability to restore flowering time in severely late flowering *lwd1lwd2ttg1* plants and to restore circadian rhythms. While the endogenous functions of the TTG1 WDR proteins in *Marchantia* are not yet known, an increasing body of evidence suggests that early diverging land plants, including the liverworts, have a circadian oscillator[18,19,20]. In Arabidopsis, LWD1 and LWD2 are proposed to act as a coregulator of *CCA1*[14] and bind the promoters of *PRR5, 9* and *1* (*TOC1*)[13]. In *Marchantia CCA1* itself appears to be absent, although related genes and the *PRR* genes are present[20]. This finding might explain the extreme degree of conservation of the TTG1 WDR proteins from across the plant tree of life, because the circadian clock is so essential to organismal fitness[21].

Functional conservation across 450 million years of evolution

If the circadian role of the TTG1 WDR proteins is an ancestral one, it might be expected that the function of TTG1 in specification of epidermal cell identity and pigment production is driven by protein characteristics or *cis* regulatory changes that have evolved more recently. However, when we attempted to rescue the *ttg1* mutant of Arabidopsis with the TTG1 WDR genes from *Marchantia polymorpha* we observed full rescue, including the production of pro-anthocyanidin in the seed coat, leaf anthocyanin synthesis, non-root hair cell determination and trichome differentiation. This strong transgenic rescue confirms that the ancestral TTG1 WDR proteins had the capacity to perform all of the combined functions of the Arabidopsis TTG1 and LWD proteins and suggests that subfunctionalization has occurred more recently.

One aspect of this transgenic rescue was particularly unexpected. Previous studies have suggested

that the role of TTG1 in specifying trichome and root hair development evolved very late[22,23]. The MYB members of the MBW involved in trichome development have only been described from Rosid genomes, and where the endogenous function of TTG1 orthologues from outside the Rosid clade has been studied, they have been found to regulate anthocyanin production but not epidermal cell differentiation[22]. It is therefore particularly surprising that two of the three TTG1 WDR proteins from *Marchantia* are capable of rescuing such newly evolved functions. Our data show that the WDR proteins have not evolved new capabilities in the Rosids, as has been previously hypothesized, but rather that the protein characteristics required for the WDR proteins to scaffold protein complexes that can specify trichomes and non-root hair cells were already present in early diverging land plants 450MYA. The key evolutionary change underpinning the evolution of trichomes and root hair development in Rosids is to be found in another protein in the MBW complex or in the presence of new target genes of the MBW complex.

Subfunctionalization in angiosperms through coding sequence change

Since the early diverging land plant TTG1 WDR proteins can rescue all aspects of the Arabidopsis *ttg1* and *lwd1lwd2* mutant phenotypes, but the 35:*LWD1* and *LWD2* constructs cannot rescue the pigmentation and epidermal patterning defects of the *ttg1* mutant, subfunctionalization must have occurred following the divergence of the TTG1 and LWD clades of proteins. This conclusion is supported by the observation that, despite its surprising role in circadian regulation, TTG1 is not as efficient as LWD1 and LWD2 in regulating the circadian clock. The ectopic expression of *TTG1* in the *lwd1lwd2ttg1* triple mutant is able to restore rhythmicity, but only at a shorter period equivalent to that observed in the *lwd1lwd2* mutant. Furthermore, our double mutant analyses show that, in all combinations, TTG1 is less effective at regulating *CCA1* expression and flowering time compared to LWD1 and LWD2. Since our transgenic rescue experiments all use the same strong promoter (and we confirmed expression in the transgenic lines), we can conclude that this subfunctionalization is a result of changes in the protein coding sequences of the TTG1 WDR genes, not a result of changes in their regulation. This is particularly surprising given the observation that clock gene expression is regulated in a circadian fashion and that constitutive

expression of some oscillator genes can inhibit rhythms. It is not possible to rule out rhythmic post-transcriptional modification of these proteins, which might also influence their function. The importance of the protein differences does not exclude that the expression of these genes has changed after the divergence between *TTG1* and *LWD1/LWD2*. Alignment of the genomic region upstream the transcription start codon shows longer blocks of similar sequence between *LWD1* and *LWD2*, although this might be expected because they have recently duplicated (Supplementary Fig. 2).

The two main roles of (i) circadian regulation and (ii) epidermal/anthocyanin specification have been divided between the LWD1/LWD2 clade proteins and the TTG1 clade, respectively. Our phylogenetic reconstruction of the TTG1 WDR protein family indicates that the duplication that gave rise to the TTG1 and LWD clades arose at the base of the seed plants, indicating that the subfunctionalization we observe in Arabidopsis must have occurred after this. We therefore investigated whether the subfunctionalization is ancestral to the angiosperms. The LWD protein from *Amborella trichopoda*, the monotypic member of the earliest diverging order of extant angiosperms[16], was able to rescue the circadian deficits arising from loss of *LWD1LWD2* function in our Arabidopsis mutants, but could not restore trichomes or seed coat pigmentation in the absence of *TTG1*. We therefore conclude that the subfunctionalization of these two protein clades, with small protein sequence divergence adapting each to a specific role, had occurred by the time the angiosperms arose, around 150 MYA.

Conclusion

Our data show that very small protein changes, hidden by extreme protein conservation, can be important in the evolution of gene function. We have shown that even when proteins are highly conserved across hundreds of millions of years their functional abilities can take unexpected twists. Detailed functional analysis is important to address how protein changes impact on protein functions, and in this case has revealed an unexpected role as a key circadian oscillator for the TTG1-WDR family of plant proteins.

Materials and Methods

Assembly, Alignment, and Phylogenetic Analysis

Homologs of TTG1 and LWD1/2 were downloaded as an orthogroup from all available fully sequenced genomes on Phytozome version 12. Further homologs of TTG1 and LWD/2 were obtained by blasting 1KP transcriptomes from gymnosperms, monilophytes, and major non-vascular plant lineages, using AtTTG1 and AtLWD1 as a reference sequence. All sequences were trimmed to include complete codons only, and then translated to amino acids. The dataset was aligned by MAFFT version 7.388, and phylogenetically analysed using FastTree version 1.0 with CAT optimization (to account for varying rates of evolution across sites), with the generation of FastTree support values with the Shimodaira-Hasegawa test (SH numbers). Phylogenetic trees were rooted using the Anthocerotophyta sequences as an outgroup, and the resultant polarised topology largely tracks the accepted organismal phylogeny. Analyses were run multiple times to confirm concordant topologies and visualized and annotated in FigTree version 1.4.3.

WDR transgenic plants

The WDR genes were amplified by RT-PCR from cDNA obtained from *Marchantia polymorpha*, *Amborella trichopoda* and *Arabidopsis thaliana* Columbia. A list of primers with restriction sites attached can be found in supplementary table 1. The restriction sites were used to insert cDNA fragments into a modified version of pGreen II 0029 with a double 35S promoter and a 35S terminator inserted using KpnI and BamHI/BglII (sequence in supplementary table 1). For transgenic rescue of Arabidopsis T-DNA insertion mutants that carry Kan resistance, the same 35S promoter/terminator was added to phosphinothricin (BASTA) resistant pGreenII0229 using KpnI and BamHI/BglII.

Marchantia gene MpWDR1 corresponds to Mapoly0259s0004.1, *MpWDR2* to Mapoly0161s0021.1, *MpWDR3* to Mapoly0027s0005.1.

Agrobacterium tumefaciens strain GV3101 was used to transform Arabidopsis. Arabidopsis

transformation was performed using the floral dip method[24].

Semi quantitative RT-PCR was performed to check the level of transgene expression in the transgenic plants. 5ul of the RT-PCR product was run for varying numbers of cycles as indicated in Extended data Fig. 4 and 8 to ensure that the PCR reaction was not at plateau phase. The list of primers used can be found in supplementary table 1.

The *ttg1-1* mutant in Landsberg erecta accession was used for single mutant transgenic rescue studies.

Triple mutant analysis in Arabidopsis

The *lwd1lwd2* (SALK_006874 (*lwd1*) and SALK_072182 (*lwd2*)) CCA1:LUC line was kindly provided by Shu-Hising Wu from the Institute of Plant and Microbial Biology, Academia Sinica, Taipei[13]. This double mutant is in the Col background. To obtain a triple mutant we crossed the double to a GABI-Kat T-DNA insertion mutant *ttg1-21* GK-580A05 also in Col. All experiments involving the triple mutant *lwd1lwd2 ttg1* contain this *ttg1-21* allele. Genotyping of the segregating population was conducted with PCR using the set of primers in supplementary table 1. RT-PCR was performed on the mutant and the complete absence of transcripts of *TTG1*, *LWD1* and *LWD2* in the mutant was verified.

Anthocyanin quantification

Arabidopsis thaliana plants were grown on liquid ½ MS media with 3% sucrose for 14 days. They were subsequently harvested and frozen in liquid nitrogen and ground, acidic methanol extraction was conducted as in ref. [25] but using 35000 as coefficient for anthocyanins and MW of 647 (Cyanin – cyanidin 3,5 diglucoside) resulting in the subsequent formula Conc = Abs/35000 x dil factor x 647x 1000/mg of sample extracted = conc mg.g-1 DW or FW.

Measuring circadian rhythms of *CCA1:LUC* and leaf movement

Circadian rhythms were measured by imaging the movement of leaves and bioluminescence of a fusion between the CCA1 promoter and the firefly luciferase reporter gene (CCA1:LUC). Plants

were grown and assayed for luciferase activity exactly as described in ref.[26] using a Photek ICCD25 camera system. Leaf movement was measured as reported in ref.[26]. The data are normalized to the mean luminescence for each trace. Circadian period estimates in constant light were calculated using FFT-NLLS analysis curve fitting tool with the BRASS V3.0 package for MS excel[26] or in Biodare2[27]. A good fit to a cosine curve is described by a relative amplitude error (RAE) of 0, whereas no fit = 1. In practice RAE > 0.5 is indicative of no detectable rhythm[26]. Supplementary table 2 summarises period length and RAE for all CCA1::LUC experiments.

qRT-PCR

qRT-PCR was performed using the LUNA qPCR mastermix from NEB and the CFX384 Bio-rad machine. Three biological replicates were performed from different plant pools grown in the same conditions. Primers used can be found in supplementary table 1. Relative expression was calculated using the delta delta Ct method, correcting for primer efficiency.

Supplementary Table 1 includes all primers used in this article.

Data availability

All data that support the findings of this study are available in the University of Cambridge data repository, with the identifier https://doi.org/10.17863/CAM.44078

Corresponding authors

Correspondence to Beverley J Glover (bjg26@cam.ac.u) or Alex A.R. Webb (aarw2@cam.ac.uk).

Acknowledgements

We thank Matthew Dorling for excellent lab and greenhouse support and Edwige Moyroud for helpful discussions. Elsa Matthus and Julia Davies provided help with root hair analysis, Matt Stancombe and Xiao Wang provided help with the circadian clock experiments, and Nick Albert and Kevin Davies provided guidance on anthocyanin extraction. We thank Shu-Hising Wu

(Institute of Plant and Microbial Biology, Academia Sinica, Taipei) for providing the *lwd1lwd2 CCA1:LUC* line. We thank the Cambridge University Botanic Garden for supplying *Amborella trichopoda* tissue for RT-PCR. Carlos A. Lugo provided support with the statistical analysis and Qi Wang with figure and dot plot presentation. CAA acknowledges support from the Cambridge University Botanic Garden Research Fund. TJH was supported by BBSRC UK grant BB/M006212/1 awarded to AARW.

Author contributions

B.J.G., C.A.A. and A.A.R.W. conceived the project and designed experiments. C.A.A. and T.J.H conducted all experiments. S.F.B conducted all phylogenetic analyses. B.J.G., A.A.R.W. and C.A.A. wrote the manuscript. All authors commented before submission.

Competing Interests

The authors declare no competing interests.

References

1. Jacob, F. Evolution and tinkering. *Science* **196**, 1161–6 (1977).

2. Wray, G. A. The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* **8**, 206–216 (2007).

3. Carroll, S. B. Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution. *Cell* **134**, 25–36 (2008).

4. Smith, T. F., Gaitatzes, C., Saxena, K. & Neer, E. J. The WD repeat: A common architecture for diverse functions. *Trends Biochem. Sci.* **24**, 181–185 (1999).

5. Ramsay, N. a & Glover, B. J. MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. *Trends in Plant Science* **10**, 63–70 (2005).

6. Kornneef, M. The complex syndrome of ttg mutants. *Arab. inf.* **18**, 45–51 (1981).

7. Pattanaik, S., Patra, B., Singh, S. K. & Yuan, L. An overview of the gene regulatory network controlling trichome development in the model plant, Arabidopsis. *Front. Plant Sci.* **5**,

448 259 (2014).
449 8. Schiefelbein, J., Huang, L. & Zheng, X. Regulation of epidermal cell fate in Arabidopsis
450 roots: the importance of multiple feedback loops. *Front. Plant Sci.* **5**, 47 (2014).
451 9. Xu, W., Dubos, C. & Lepiniec, L. Transcriptional control of flavonoid biosynthesis by
452 MYB-bHLH-WDR complexes. *Trends Plant Sci.* **20**, 176–185 (2015).
453 10. Pesch, M. *et al.* TTG1 and GL1 compete for binding to GL3 in Arabidopsis thaliana. *Plant*
454 *Physiol.* **168**, pp.00328.2015 (2015).
455 11. Zhang, B., Chopra, D., Schrader, A. & Hülskamp, M. Evolutionary comparison of
456 competitive protein-complex formation of MYB, bHLH, and WDR proteins in plants. *J. Exp. Bot.*
457 **70**, 3197–3209 (2019).
458 12. Wu, J.-F., Wang, Y. & Wu, S.-H. Two New Clock Proteins, LWD1 and LWD2, Regulate
459 Arabidopsis Photoperiodic Flowering. *Plant Physiol.* **148**, 948–959 (2008).
460 13. Wang, Y. *et al.* LIGHT-REGULATED WD1 and PSEUDO-RESPONSE REGULATOR9
461 Form a Positive Feedback Regulatory Loop in the Arabidopsis Circadian Clock. *Plant Cell* **23**,
462 486–498 (2011).
463 14. Wu, J.-F. *et al.* LWD-TCP complex activates the morning gene CCA1 in Arabidopsis. *Nat.*
464 *Commun.* **7**, 13181 (2016).
465 15. Nakamichi, N., Kita, M., Ito, S., Yamashino, T. & Mizuno, T. PSEUDO-RESPONSE
466 REGULATORS, PRR9, PRR7 and PRR5, Together play essential roles close to the circadian
467 clock of Arabidopsis thaliana. *Plant Cell Physiol.* **46**, 686–698 (2005).
468 16. The Angiosperm Phylogeny Group III. An update of the Angiosperm Phylogeny Group
469 classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* **161**, 105–
470 121 (2009).
471 17. McClung, C. R. Wheels within wheels: new transcriptional feedback loops in the
472 Arabidopsis circadian clock. *F1000Prime Rep.* **6**, 2 (2014).
473 18. Okada, R. *et al.* Functional characterization of CCA1/LHY homolog genes, PpCCA1a and
474 PpCCA1b, in the moss Physcomitrella patens. *Plant J.* **60**, 551–563 (2009).
475 19. Holm, K., Källman, T., Gyllenstrand, N., Hedman, H. & Lagercrantz, U. Does the core

476 circadian clock in the moss Physcomitrella patens (Bryophyta) comprise a single loop? *BMC*
477 *Plant Biol.* **10**, (2010).
478 20. Linde, A.-M. *et al.* Early evolution of the land plant circadian clock. *New Phytol.* (2017).
479 doi:10.1111/nph.14487
480 21. Dodd, A. N. Plant Circadian Clocks Increase Photosynthesis, Growth, Survival, and
481 Competitive Advantage. *Science (80-.).* **309**, 630–633 (2005).
482 22. Serna, L. & Martin, C. Trichomes: different regulatory networks lead to convergent
483 structures. *Trends Plant Sci.* **11**, 274–280 (2006).
484 23. Yang, C. & Ye, Z. Trichomes as models for studying plant cell differentiation. *Cellular*
485 *and Molecular Life Sciences* (2012). doi:10.1007/s00018-012-1147-6
486 24. Clough, S. J. & Bent, A. F. Floral dip: a simplified method forAgrobacterium-mediated
487 transformation ofArabidopsis thaliana. *Plant J.* **16**, 735–743 (1998).
488 25. Albert, N. W. *et al.* Light-induced vegetative anthocyanin pigmentation in Petunia. *J. Exp.*
489 *Bot.* **60**, 2191–2202 (2009).
490 26. Martí Ruiz, M. C. *et al.* Circadian oscillations of cytosolic free calcium regulate the
491 Arabidopsis circadian clock. *Nat. Plants* **4**, (2018).
492 27. Zielinsky T, Moore AM, Troup E, Halliday KJ, Millar AJ. Strengths and Limitations of
493 Period Estimation Methods for Circadian Data. PLoS ONE 9(5): e96462. (2014)
494
495
496
497
498
499 **Figure Legends**
500 **Fig. 1. Evolution of the TTG1 and LWD1/LWD2 clades. a,** Simplified phylogenetic tree of the
501 TTG1 WDR proteins (detailed tree available in Supplementary Fig. 1). AthTTG1, AthLWD1,
502 AthLWD2 refer to proteins from *Arabidopsis thaliana*, AmLWD refers to a protein from the early
503 diverging angiosperm *Amborella trichopoda*. MpWDR1, MpWDR2, MpWDR3 refer to proteins

from the liverwort *Marchantia polymorpha*. The black star indicates a gene duplication event at the base of the seed plant lineage. **b, c, Transgenic rescue of *ttg1-1* mutant ectopically expressing different WDR genes. b,** Seed coat pigments are absent in the *ttg1-1* mutant and in plants ectopically expressing *LWD1* and *LWD2*. Seed coat pro-anthocyanidins are visible in the *ttg1-1* mutant overexpressing *TTG1*. **c,** Picture of the leaf blade, trichomes are absent in the *ttg1-1* mutant and in the plants ectopically expressing *LWD1* and *LWD2*, but present in the *ttg1-1* mutant overexpressing *TTG1*. The same phenotypes were observed in multiple independent lines (*ttg1* 35S:TTG1 = 7, *ttg1* 35S:LWD1 = 10, *ttg1* 35S:LWD2 = 12).

Fig. 2. *lwd1lwd2ttg1* has arrhythmic *CCA1:LUC* and leaf movement. a, Mean normalized leaf movement rhythms measured across 6 days in constant light from Col-0 (green, n = 68), *lwd1lwd2ttg1* (pink, n = 83), *ttg1* (blue, n = 66) and *lwd1lwd2* (yellow, n = 59) seedlings. Mean normalized leaf pixel position measured from 14 day old seedlings shown with SEM. Constant illumination supplied as 70 μmol $m^{-2}s^{-1}$ cool white fluorescent light. FFT-NLLS analysis of leaf movement circadian period estimates shown with relative amplitude error (RAE) for rhythmic individuals of Col-0 (green, number rhythmic=46/68), *ttg1* (blue, number rhythmic =41/66), *lwd1lwd2* (yellow, number rhythmic =36/59) and *lwd1lwd2ttg1* (pink, number rhythmic =23/83). **b,** Mean normalized *CCA1:LUC* rhythms measured across 2 12:12 LD (light/dark) cycles and 5 days in constant light from Col-0 (green, n = 24) and *lwd1lwd2ttg1* (pink, n = 24) seedlings. Mean normalized *CCA1:LUC* luminescence (counts $800s^{-1}$) measured from 11 day old seedlings grown on 0.5 MS media shown with SEM. Diel and constant light supplied as monochromatic red and blue 70 μmol $m^{-2}s^{-1}$ LED illumination. FFT-NLLS analysis of *CCA1:LUC* circadian period estimates shown with relative amplitude error (RAE) for individual replicates Col-0 (green, number rhythmic=24) and *lwd1lwd2ttg1* (pink, number rhythmic =21/24).

Fig. 3. *CCA1:LUC* rhythms in WDR family single, double and triple mutant lines. **a,** Mean normalized *CCA1:LUC* rhythms measured across LD and four days in constant light from Col-0

(green, n = 11), or *lwd1lwd2ttg1* (pale pink, n = 5), *lwd1lwd2* (yellow, n = 5), *lwd1ttg1* (light orange, n = 11), *lwd2ttg1* (brown, n = 11), *ttg1* (blue, n = 11), *lwd1* (magenta, n = 11) and *lwd2* (purple, n = 11) seedlings. Mean normalized *CCA1:LUC* luminescence (counts 800 s^{-1}), measured from 11 day old seedlings are plotted with SEM. Diel and constant light supplied as monochromatic red and blue 70 μmol $m^{-2}s^{-1}$ LED illumination. **b,** FFT-NLLS circadian period analysis performed on *CCA1:LUC* rhythms. Best fitted period length and relative amplitude error (RAE) shown for individual replicates of mean rhythms shown in **a**.

Fig. 4. Two of the three *Marchantia polymorpha* WDR genes rescue the *ttg1-1* mutant. a, Seed coat pigments are present in the *ttg1-1* mutant lines ectopically expressing *MpWDR1* and *MpWDR2*. Seed coat pro-anthocyanidins are not visible in the *ttg1-1* mutant expressing *MpWDR3*. **b,** Leaf trichomes are absent in the *ttg1-1* mutant ectopically expressing *MpWDR3* but they are present in plants expressing *MpWDR1* and *MpWDR2*. The same phenotypes were observed in multiple independent lines *ttg1* 35S:MpWDR1 = 20, *ttg1* 35S:MpWDR2 = 15, *ttg1* 35S:MpWDR3 = 18.

Fig. 5. TTG1 and LWD1/LWD2 proteins show differential ability to rescue the circadian defect of the *lwd1lwd2ttg1* triple mutant. a, Mean normalized *CCA1:LUC* rhythms measured across LD and four days in constant light from Col-0 (green, n = 12), *lwd1lwd2* (yellow, n = 12) and *lwd1lwd2ttg1* (pink, n = 12), expressing either 35S:*TTG1 (TTG1,* n = 12, *TTG1-3* n = 6, *TTG1-4* n = 6), 35S:*LWD1* (n = 12), 35S:*LWD2* (n = 8). Mean normalized *CCA1:LUC* luminescence (counts 800 s^{-1}), measured from 11 day old seedlings are plotted with SEM. Diel and constant light supplied as monochromatic red and blue 70 μmol m^{-2} s^{-1} LED illumination. **b,** FFT-NLLS circadian period analysis performed on *CCA1:LUC* rhythms. Best fitted period length and relative amplitude error (RAE) shown for individual replicates of mean rhythms shown in **a**.

Fig. 6. Different WDR proteins show differential ability to rescue the circadian defect of the *lwd1lwd2ttg1* triple mutant. a, Mean normalized *CCA1:LUC* rhythms measured across LD and four days in constant light from triple mutants expressing 35S:*MpWDR1* (n = 12), 35S:*MpWDR2* (n = 12), 35S:*MpWDR3* (n = 12) and 35S:*AmLWD* (n = 11). Mean normalized *CCA1:LUC* luminescence (counts 800 s^{-1}), measured from 11 day old seedlings are plotted with SEM. Diel and constant light supplied as monochromatic red and blue 70 μmol m^{-2} s^{-1} LED illumination. **b,** FFT-NLLS circadian period analysis performed on *CCA1:LUC* rhythms. Best fitted period length and relative amplitude error (RAE) shown for individual replicates of mean rhythms shown in **a**.

Extended data Fig. 1.. Alignment of TTG1-WDR proteins. The alignment was performed with the MAFFT algorithm using the cloud-based informatics platform benchling. The colours indicate the degree of amino acid conservation between the six proteins from dark red for the most conserved to blue for the least conserved.

Extended Data Fig. 2. Transgenic rescue of the *ttg1-1* phenotype. a,b, Amount of anthocyanin in mg/g of dry weight, bars represent standard deviation of up to three different extractions. Comparison between WT, *ttg1-1* mutant and *ttg-1-1* mutant ectopically expressing **a,** *TTG1, LWD1, LWD2, MpWDR1, MpWDR2*. **b,** *MpWDR3, AmLWD***.** Graphs show mean values for two biological replicates, each replicate pooled several seedlings.

Extended Data Fig. 3. Transgenic rescue of the *ttg1-1* phenotype. a,b,c, Boxplots of root hair count in 2.5 mm of the first 5 mm of the root in the same genotypes used for the anthocyanin assay in Extended Data Fig. 2. Number of plants analyzed in a is WT = 13, *ttg1* = 7, *ttg1* 35S:TTG1 = 9, *ttg1* 35S:LWD1 = 8, *ttg1* 35S:MpWDR1 = 10; in b is WT = 13, *ttg1* = 10 *ttg1* 35S:TTG1 = 11, *ttg1* 35S:LWD2 = 14, *ttg1* 35S:MpWDR2 = 12; in c is WT = 9, *ttg1* = 6, *ttg1* 35S:AmLWD = 6, *ttg1* 35S:MpWDR3 = 6. Additional details about the statistics can be found in Supplementary table 3. **d**, Table illustrating p values for pairwise comparisons. p values were

584 calculated using a non parametric anova using the Kruskal-Wallis test, followed by a *post hoc*
585 analysis of the means using the Conover test. *ttg1* 35S:*TTG1*/*WDR1*/*WDR2* are all significantly
586 different from the *ttg1* mutant. *ttg1* 35S:*LWD1*/*AmLWD* are significantly different from the WT
587 and *ttg1* 35S:*WDR3* is significantly different from both, with higher support to be different from
588 the WT. Additional details about the statistics can be found in Supplementary table 3.
589
590 **Extended Data Fig. 4. Semi quantitative RT-PCR in plants ectopically expressing TTG1**
591 **WDR genes in the *ttg1-1* mutant and in the triple mutant *lwd1lwd2ttg1*.** The figure shows for
592 each sample 5 ul of the same amplification reaction after 20-25-35 PCR cycles. DNA ladder is
593 1kb hyperladder (Bioline). **a**, PCR of WT and *ttg1* overexpressing lines (35S:*TTG1*, *LWD1*,
594 *LWD2*, *MpWDR1*, *MpWDR2*) and reference gene (*EUKARYOTIC TRANSLATION INITIATION*
595 *FACTOR 4A1* (*EIF4A1*)). **b**, PCR of WT, triple mutant *lwd1lwd2ttg1* and *ttg1* overexpressing
596 lines (35S:*AmLWD*, *TTG1*, *LWD1*, *LWD2*, *MpWDR1*, *MpWDR2)*. Negative control samples
597 indicated with "-". Given the big differences observed, this experiment was performed only once,
598 using multiple lines for most of the transgenic plants.
599
600 **Extended Data Fig. 5. Flowering time phenotype of single and double and triple mutant**
601 **combinations. a,** WT and triple mutant plants grown in the same tray in long day conditions show
602 a dramatically different flowering time. **b, c,** The graphs represent the mean number of rosette and
603 cauline leaves at bolting in different mutant combinations, error bars represent standard deviation.
604 Number of plants in b is WT = 57, *lwd1lwd2ttg1* = 47 in c is WT = 22, lwd1ttg1 = 23, *lwd1lwd2* =
605 16, *lwd2ttg1* = 17, *ttg1* = 14.
606
607 **Extended Data Fig. 6. Comparisons of *lwd1* mutant with *lwd1ttg1* double mutant. a**,
608 *CCA1:LUC* luminescence measured from Col-0, *lwd1*, *ttg1* and *lwd1ttg1* seedlings. Seedlings
609 were entrained in 12:12 light dark cycle and transferred to camera chamber on day 9.
610 Luminescence was measured for one 12:12 light dark cycle and 96 hours in constant light. Mean

luminescence shown with SEM for n = 7, except *ttg1* where n = 3. FFT-NLLS was used to estimate period values implemented using Biodare 2. Student's t test was used to identify whether genotypes were significantly different for period values with RAE<0.5. * denotes p<0.05. Details about the statistical analysis can be found in Supplementary table 3. **b**, Expression analysis of LWD2 in *ttg1lwd1* mutant. The graph shows mean relative expression of *LWD2* in the *ttg1lwd1* double mutant compared to WT in three biological replicas, data obtained by qRT-PCR with *LWD2* specific primers and reference gene *UBQ10*. Error bars represent standard deviation on three biological replicas.

Extended Data Fig. 7 *lwd1lwd2ttg1* triple mutant phenotype. **a**, Rosettes of Arabidopsis plants with mutant combinations of different TTG1 WDR genes. Plants were germinated at the same time and grown in the same LD conditions. In the triple mutant *lwd1lwd2ttg1* leaf morphology is perturbed, whereas single and double mutant combinations have wild type leaf morphologies. These lines were grown repeatedly with no variations on these observations. **b**, Leaf margins of *ttg1* mutant, *lwd1ttg1* double mutant and the triple mutant *lwd11lwd2ttg1*. These differences were observed in a minimum of 6 plants in each of at least three independent batches. **c**, Boxplot of trichome numbers on the leaf edge of the *ttg1* mutant, double mutants and the triple mutant. Data represent total trichome number on a plant with 9 leaves (number of plants counted *lwd1lwd2ttg1* 17, *ttg1* 18, *ttg1lwd1* 13, *ttg1lwd2* 11). **d,** Table illustrating p values for pairwise leaf trichome number comparisons. p values were calculated using a non parametric anova using the Kruskal-Wallis test, followed by a post hoc analysis of the means using the Conover test. Additional details about the statistics can be found in Supplementary table 3.

Extended Data Fig. 8. Semi quantitative RT-PCR in plants ectopically expressing TTG1 WDR genes in the triple mutant *lwd1lwd2ttg1*. DNA ladder is 1kb hyperladder (Bioline). The figure shows for each sample 5ul of the same amplification reaction after 20-25-35 or 22-27-35 PCR cycles. **a,** PCR with gene specific primers for *TTG1* and reference gene (*EUKARYOTIC*

638 *TRANSLATION INITIATION FACTOR 4A1* (*EIF4A1*)) on cDNA of triple mutant *lwd1lwd2ttg1*
639 plants overexpressing *TTG1* in the triple mutant *lwd1lwd2ttg1*. **b,** PCR with gene specific primers
640 for *MpWDR3* and housekeeping gene on *ttg1* and triple mutant *lwd1lwd2ttg1* plants expressing
641 *Marchantia polymorpha* gene *MpWDR3*. Data for *MpWDR2* are included for comparison to a
642 gene expression level that was capable of transgenic rescue. Negative control samples indicated
643 with "-". The figure show all lines analyzed in semi quantitative RT-PCR.
644
645 **Extended Data Fig. 9. Flowering time of WT, triple mutant and triple mutant plants**
646 **overexpressing TTG1-like WDR proteins. a,** Plants in each panel were sown at the same time
647 and grown alongside each other in long day conditions. WT, triple mutant plants *lwd1lwd2ttg1*,
648 triple mutant overexpressing *LWD1, TTG1, MpWDR1, MpWDR2, MpWDR3* and *AmLWD*. All
649 plant we have analyzed (WT 66, *lwd1lwd2* 41, triple 48, triple 35S:*AmLWD1* 26, triple 35S:*LWD1*
650 39, triple 35S:*MpWDR1* 23, triple 35S:*MpWDR2* 30, triple 35S:*MpWDR3* 54, triple 35S:*TTG1* 66)
651 show the same pattern, with small variations that are reported quantitatively in b. **b,** Mean number
652 of rosette and cauline leaves at bolting (error bars indicate standard deviation) and relative p
653 values obtained using Post hoc pairwise test for multiple comparisons of mean rank sums (Dunn's
654 test) used after Kruskal-Wallis one-way analysis of variance by ranks to do pairwise comparisons.
655 Triple 35S:*LWD1* and triple 35S:*MpWDR1* have the same flowering time as wild type; triple
656 35S:*TTG1*, triple 35S:*WDR2* and triple 35S:*AmLWD* flower slightly earlier than the wild type and
657 triple 35S:*MpWDR3* flowers later than the WT. 35S:*MpWDR3* flowers at the same time as the
658 triple mutant (p value 1). *lwd1lwd2* flowering time is significantly different from the wild type and
659 most of the transgenics but the p value is higher when we compare *lwd1lwd2* to triple 35S:*TTG1*
660 and triple 35S:*AmLWD* (p value respectively 0.09, 0.1).
661 Number of plants in the analysis: WT 66, *lwd1lwd2* 41, triple 48, triple 35S:*AmLWD1* 26, triple
662 35S:*LWD1* 39, triple 35S:*MpWDR1* 23, triple 35S:*MpWDR2* 30, triple 35S:*MpWDR3* 54, triple
663 35S:*TTG1* 66. Additional details about the statistics can be found in Supplementary table 3.
664

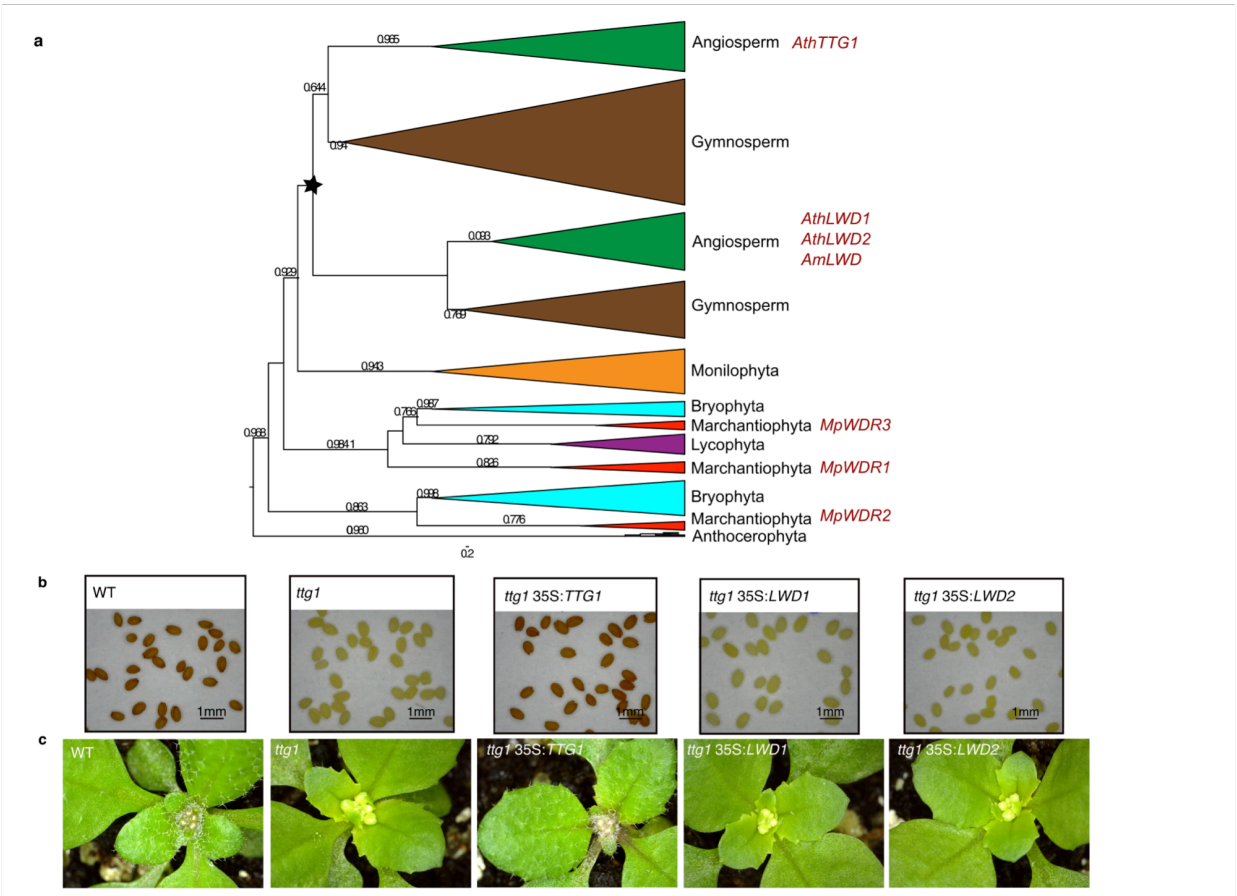
Extended Data Fig. 10. Transgenic rescue of *ttg1-1* mutant with 35S:*AmLWD*. AmLWD is not able to rescue the seedcoat and trichome phenotypes of the *ttg1-1* mutant. 6 of 6 independent transgenic lines all showed the same phenotype.

Supplementary Fig. 1. Phylogenetic tree of TTG1 WDR proteins from across the plant kingdom. Green = angiosperms, brown = gymnosperms, orange = monilophyta, purple = lycophyta, red = marchantiophyta, turquoise = bryophyta, black = anthocerotophyta. Functionally characterised genes are highlighted in red.

Supplementary Fig. 2. Pairwise alignments of the genomic region upstream of the start codon of the Arabidopsis thaliana genes LWD1, LWD2 and TTG1.

677 **Fig.1**

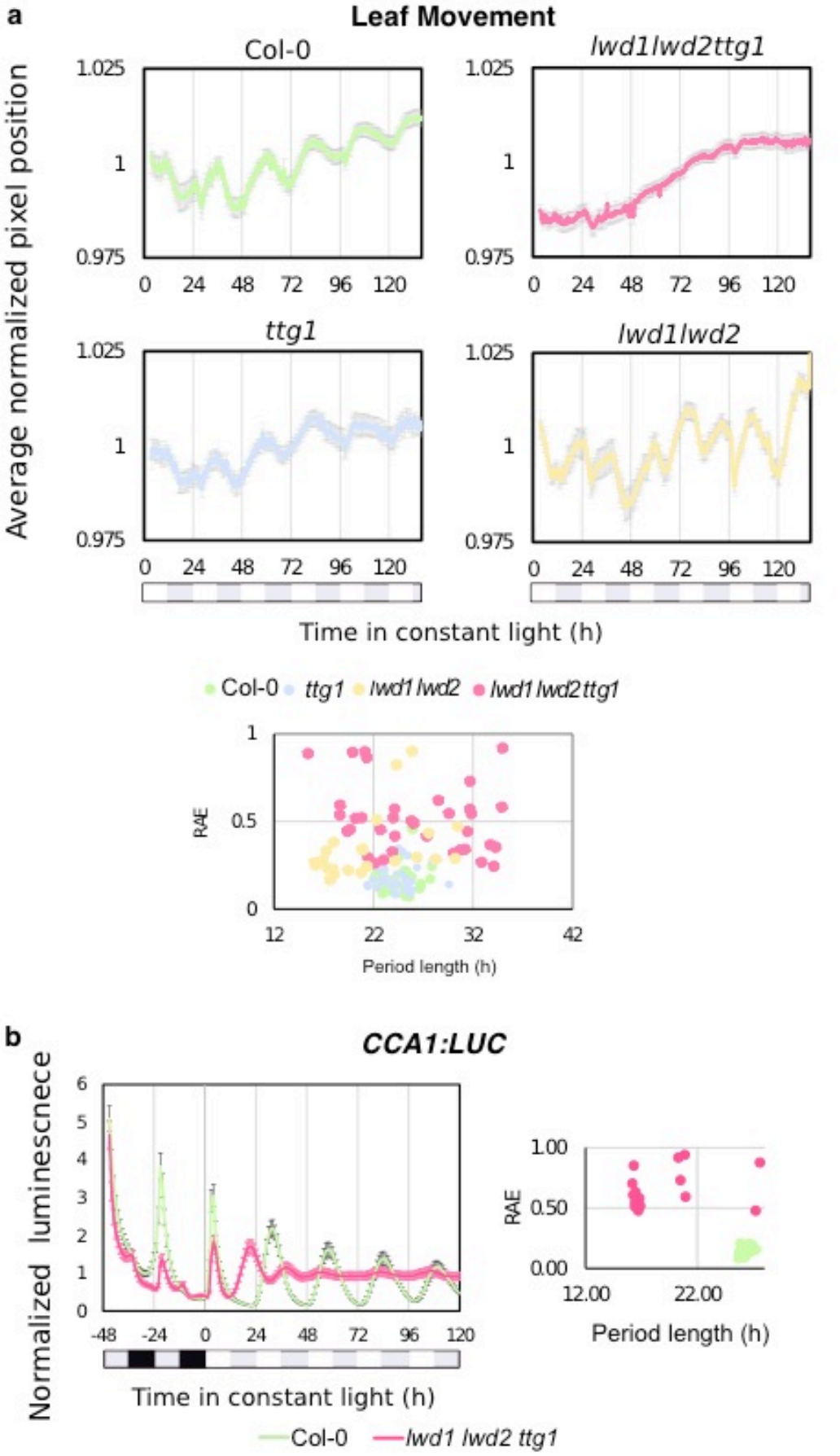
678



679

680 **Fig.2**

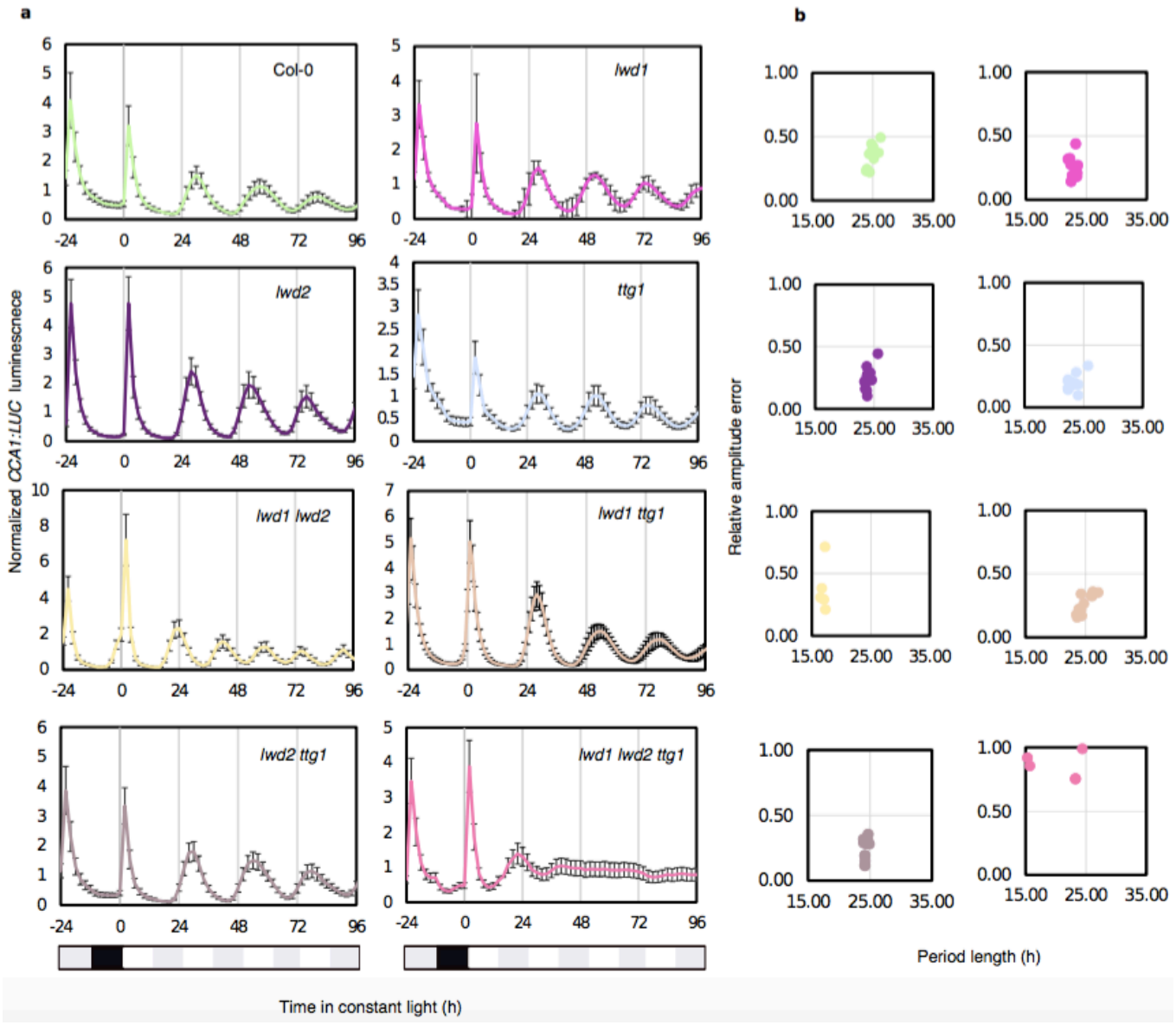
681



682

683 **Fig.3**

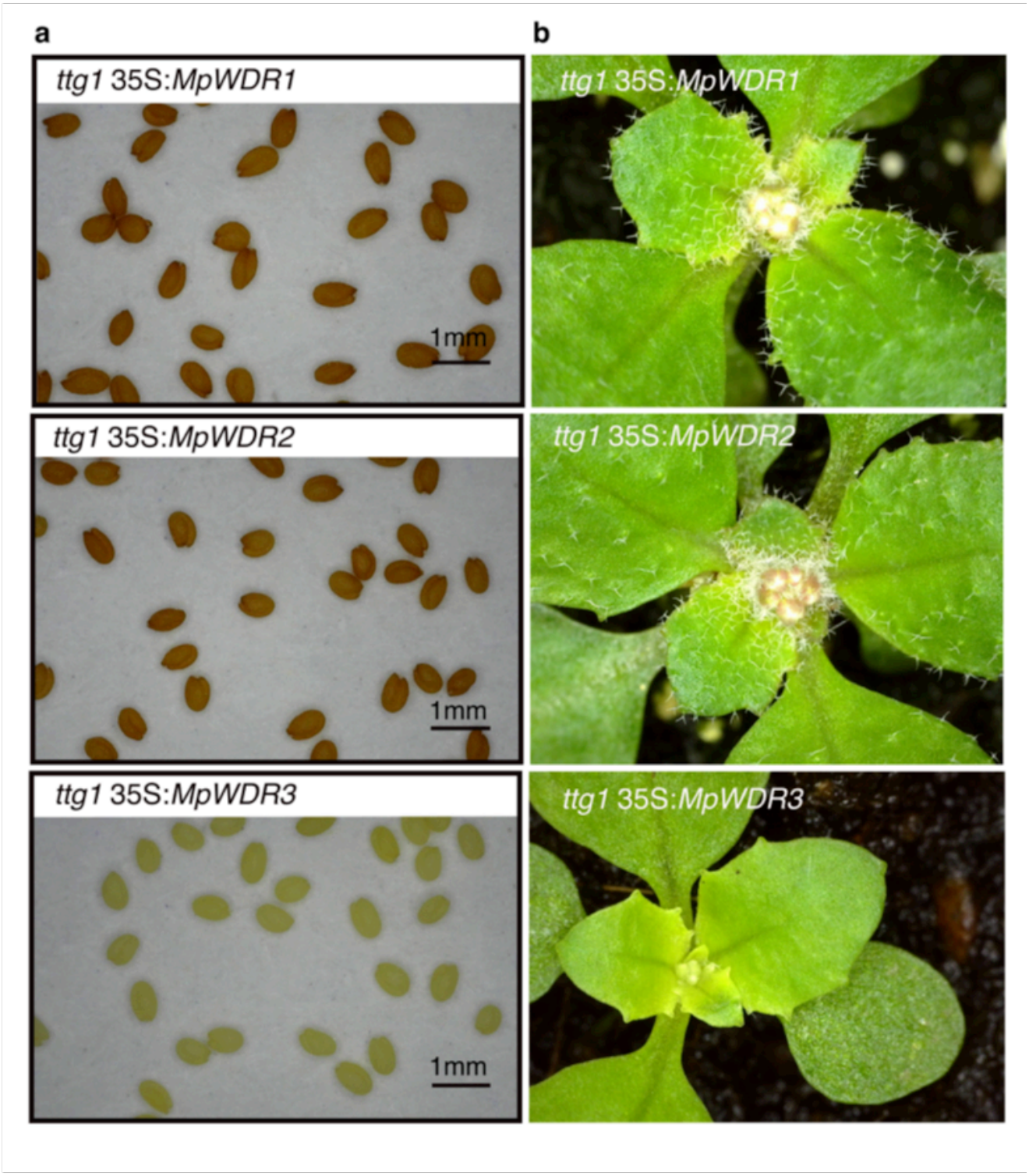
684



685

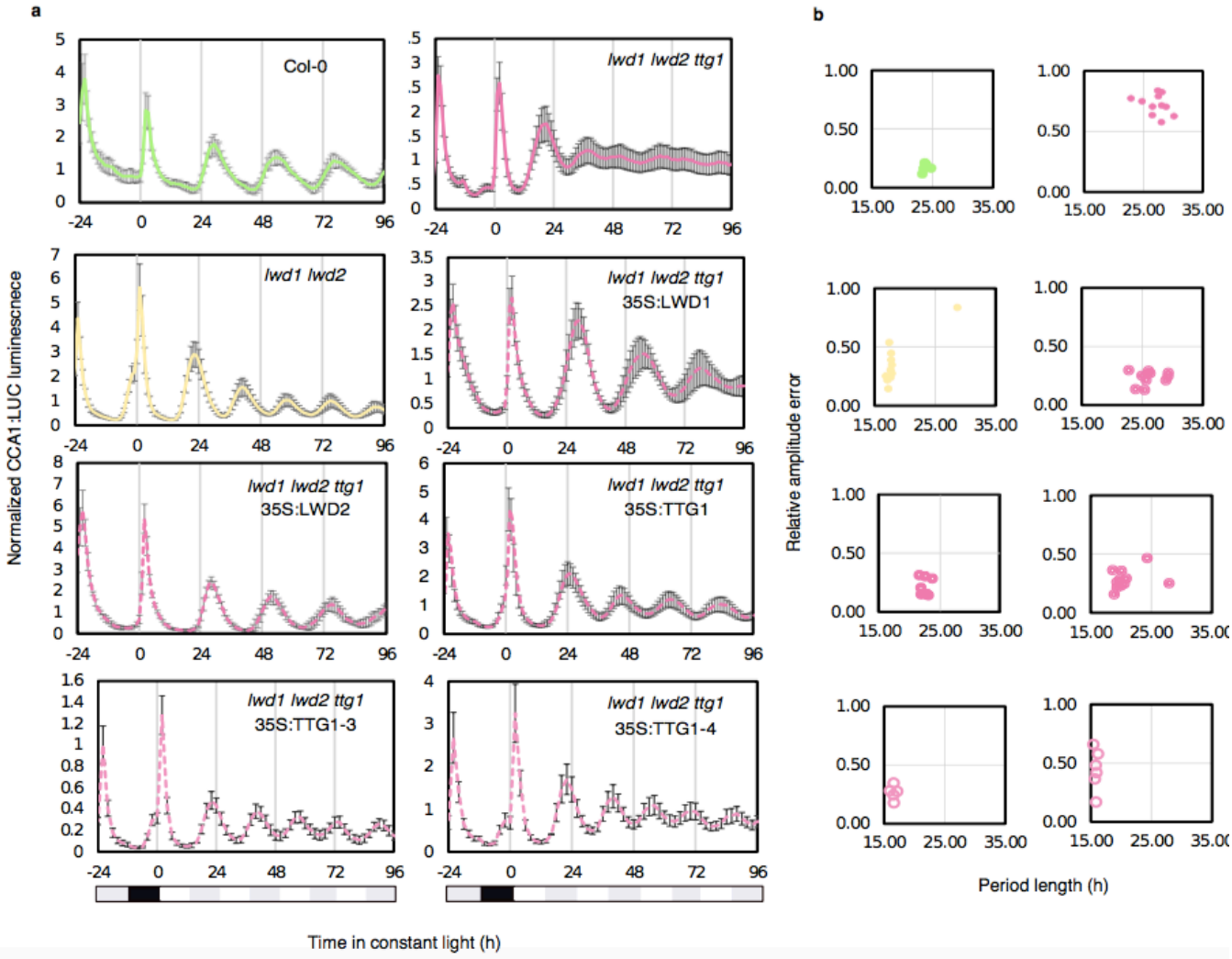
686 **Fig.4**

687


a
b
ttg1 35S:MpWDR1
1mm
ttg1 35S:MpWDR1
ttg1 35S:MpWDR2
1mm
ttg1 35S:MpWDR2
ttg1 35S:MpWDR3
1mm
ttg1 35S:MpWDR3

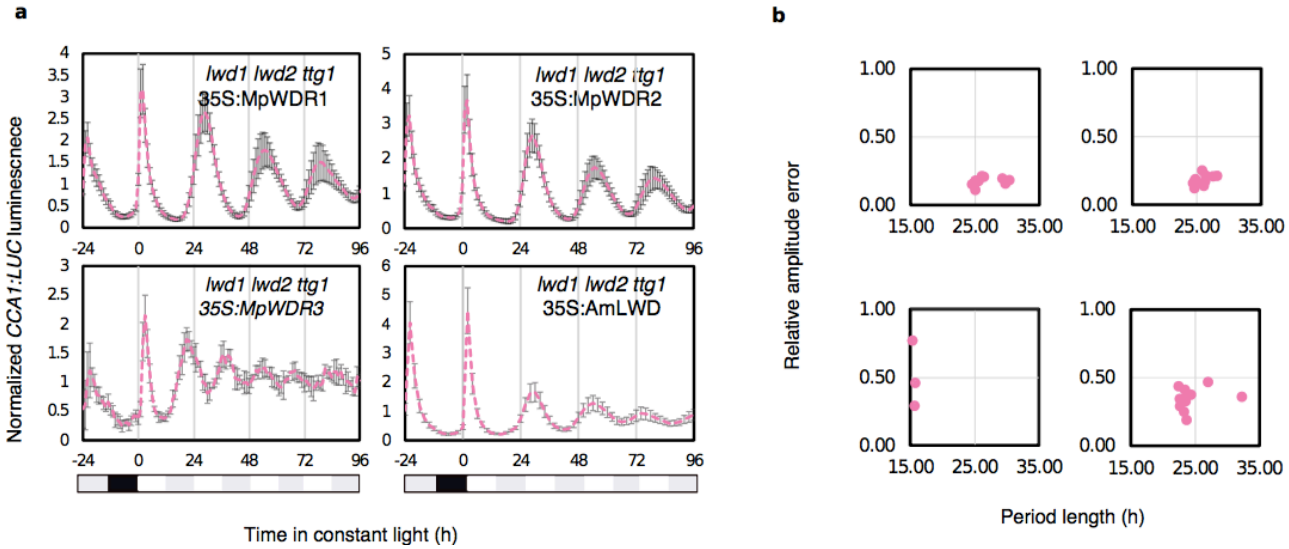
688 **Fig.5**

689



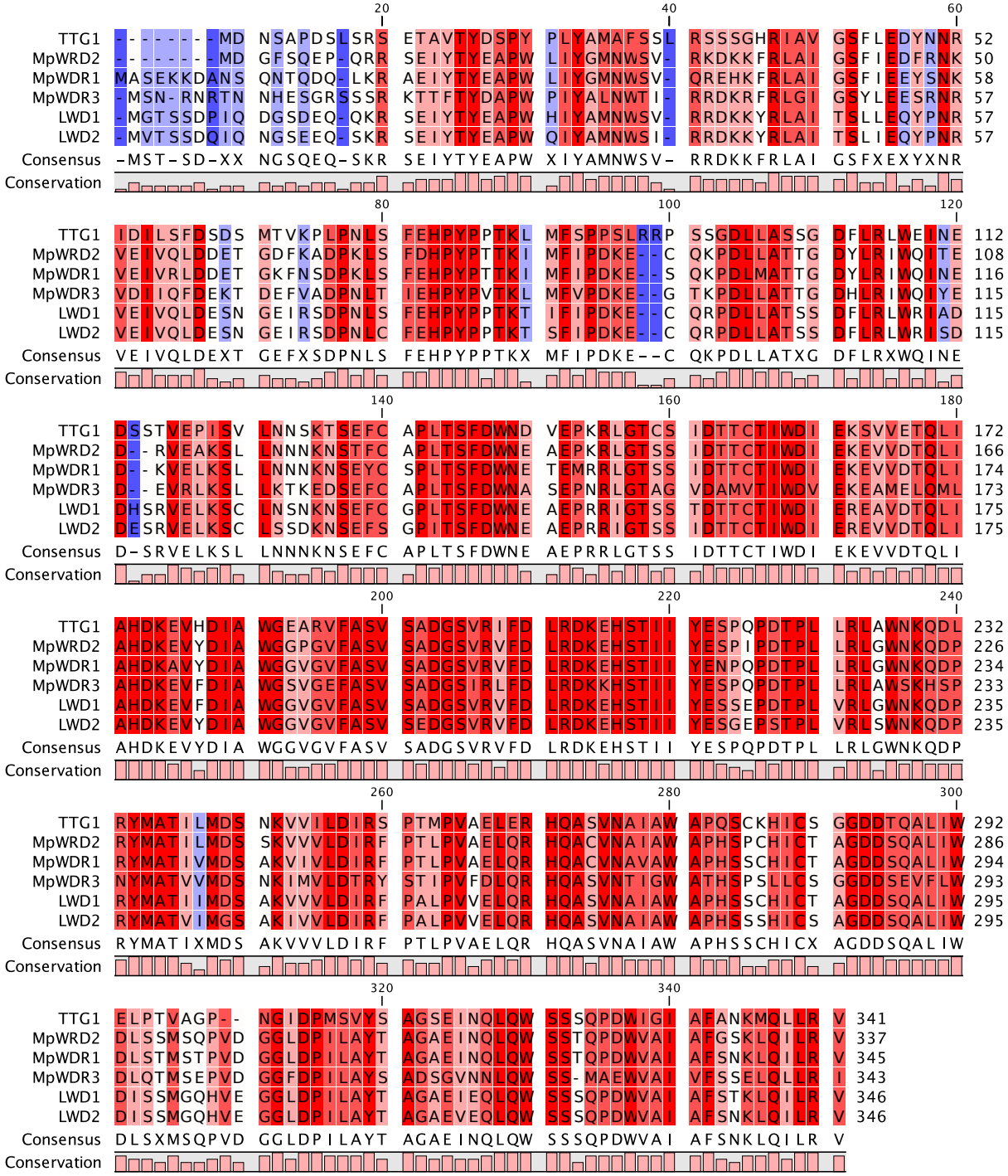
690 **Fig.6**

691

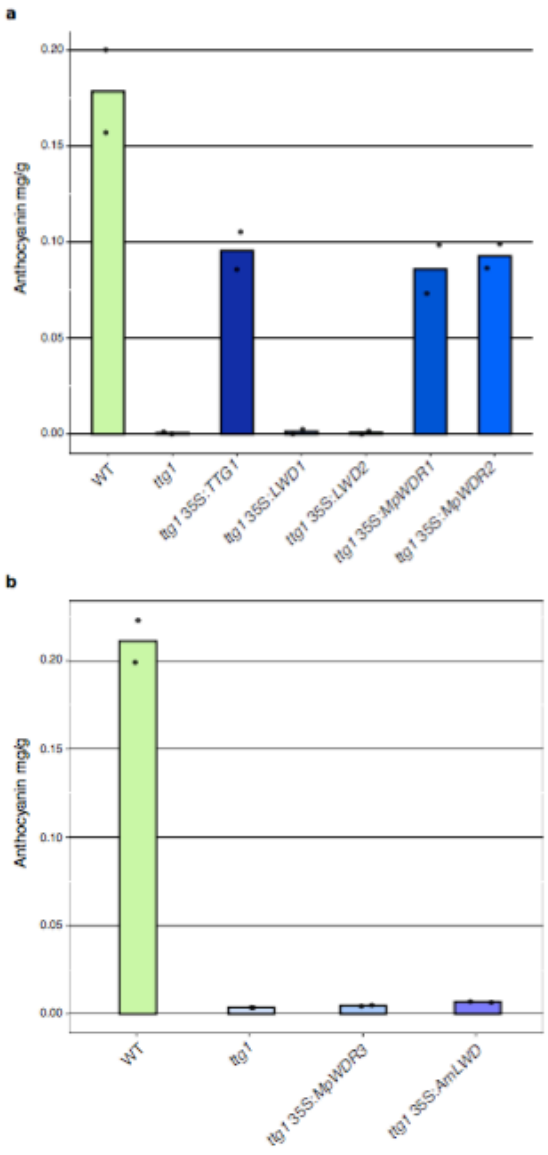


692

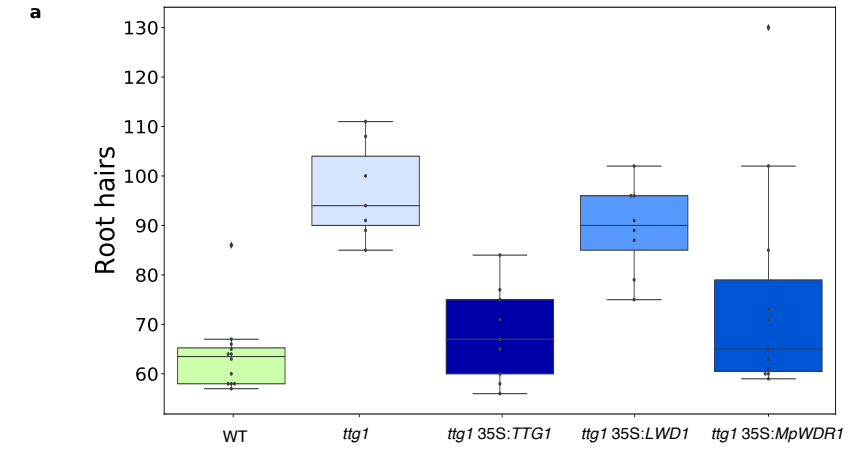
Extended Data Fig. 1

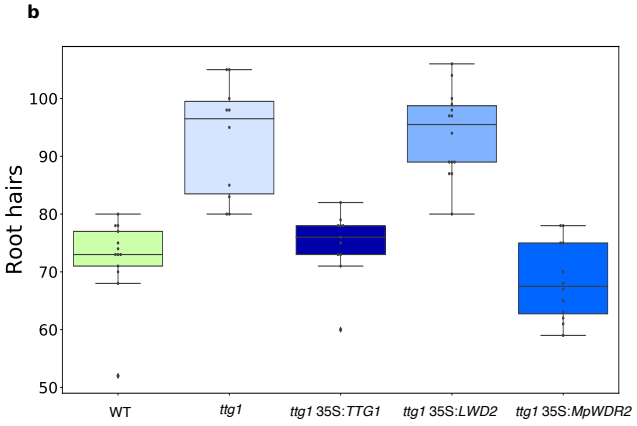


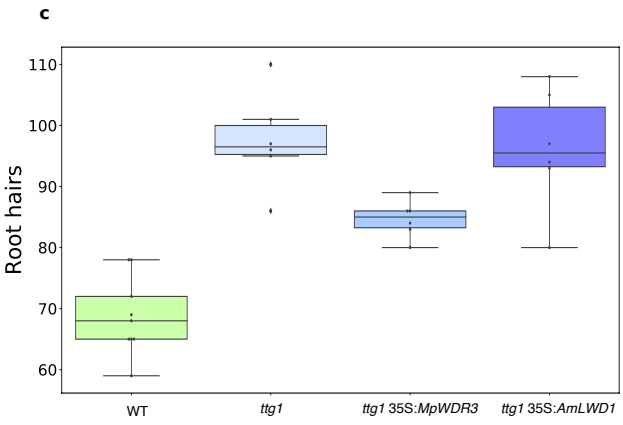
Extended Data Fig. 2



Extended Data Fig. 3







d P values

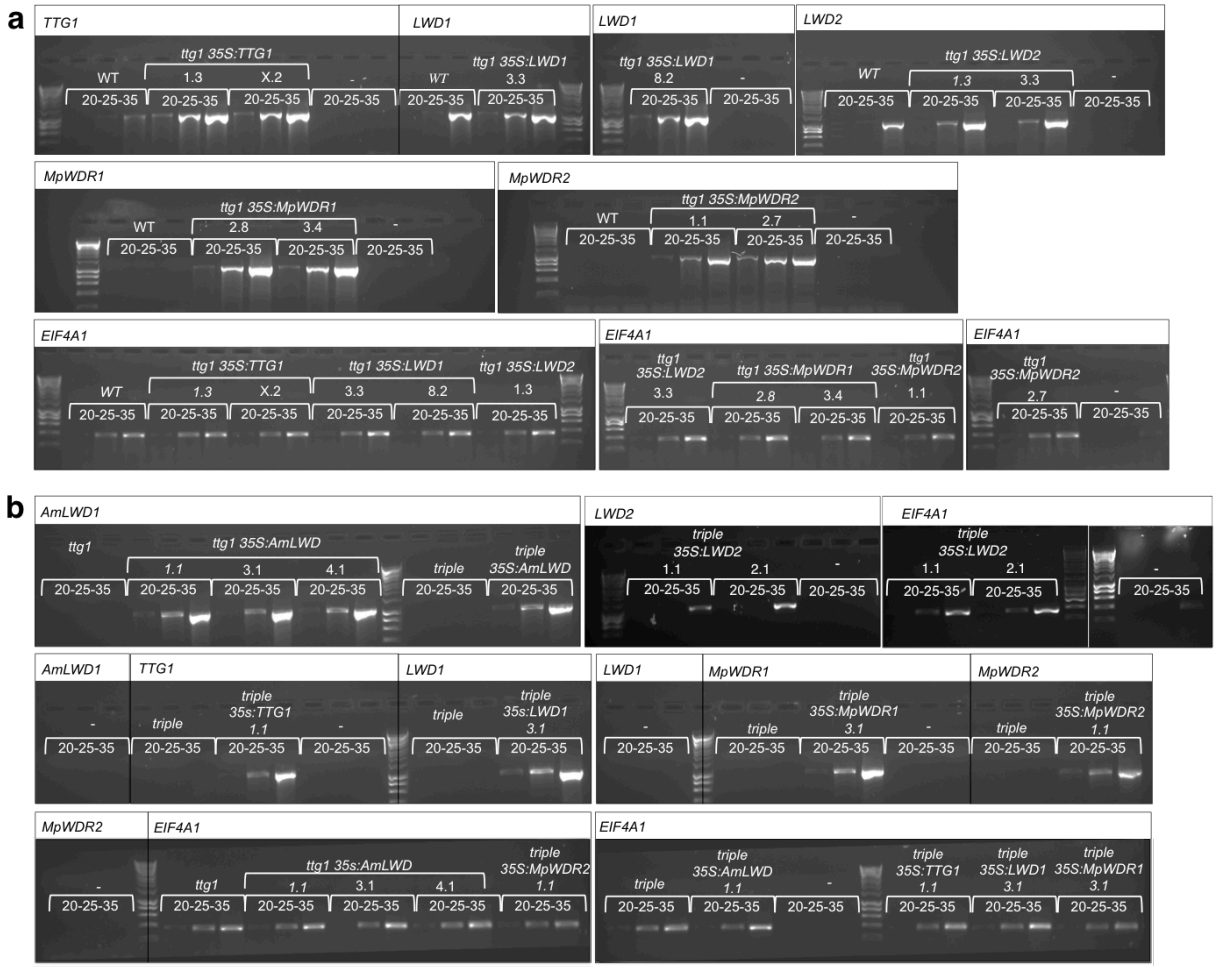
	ttg1 35S:LWD1	*ttg1 35S:TTG1*	*ttg1 35S:WRD1*	*ttg1*
WT	1.15E-04	9.10E-01	2.03E-01	2.00E-05
	ttg1 35S:LWD1	*ttg1 35S:TTG1*	*ttg1 35S:WRD1*	WT
ttg1	9.10E-01	8.15E-04	4.36E-03	2.00E-05

	ttg1 35S:LWD2	*ttg1 35S:TTG1*	*ttg1 35S:WRD2*	*ttg1*
WT	2.08E-10	4.94E-01	2.91E-01	5.52E-09
	ttg1 35S:LWD2	*ttg1 35S:TTG1*	*ttg1 35S:WRD2*	WT
ttg1	8.35E-01	6.91E-07	3.15E-11	5.52E-09

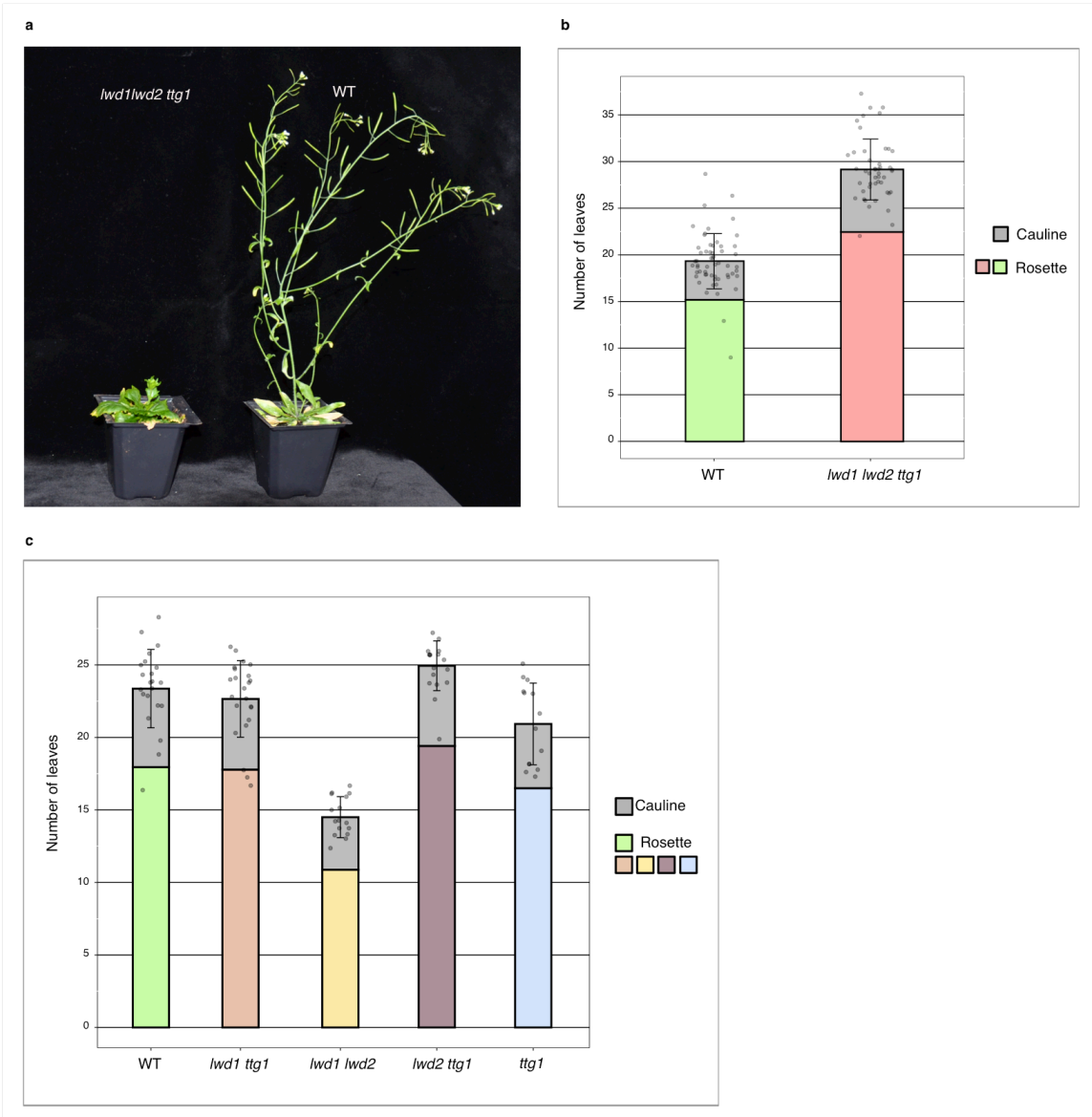
	ttg1 35S:AmLWD	*ttg1 35S:WDR3*	*ttg1*
WT	5.13E-07	8.66E-04	1.34E-07
	ttg1 35S:AmLWD	*ttg1 35S:WDR3*	WT
ttg1	5.23E-01	4.73E-03	1.34E-07

p<0.001
p<0.01
p<0.05
NS

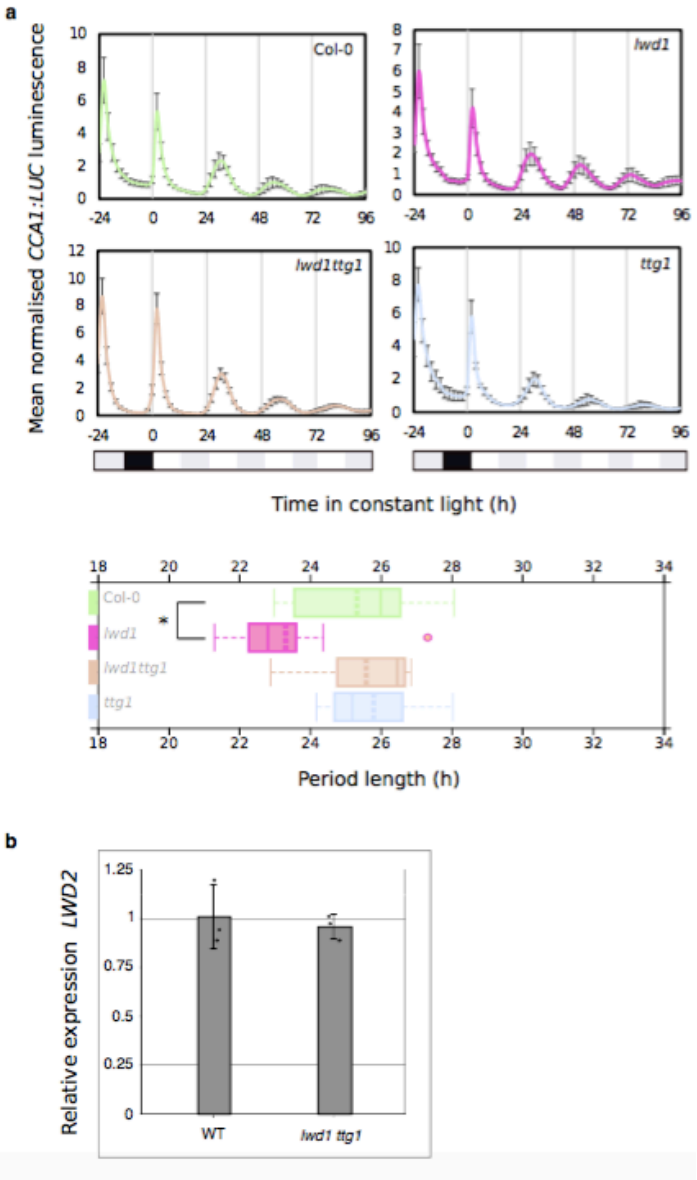
Extended Data Fig. 4



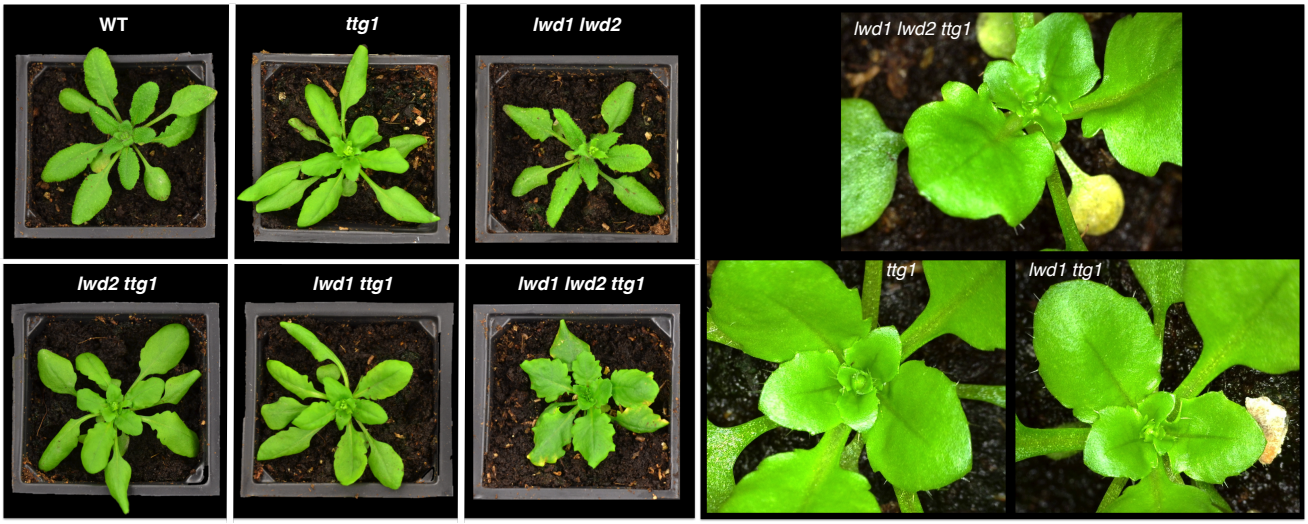
Extended Data Fig. 5

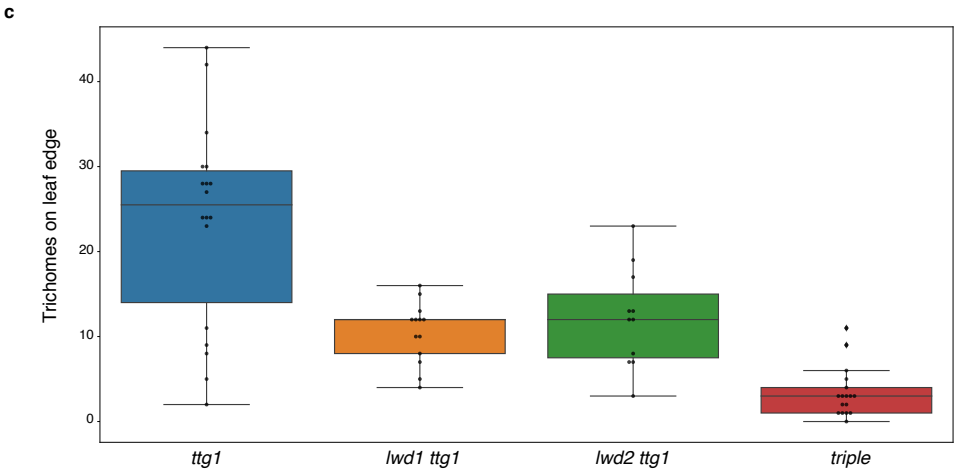


Extended Data Fig. 6



Extended Data Fig. 7



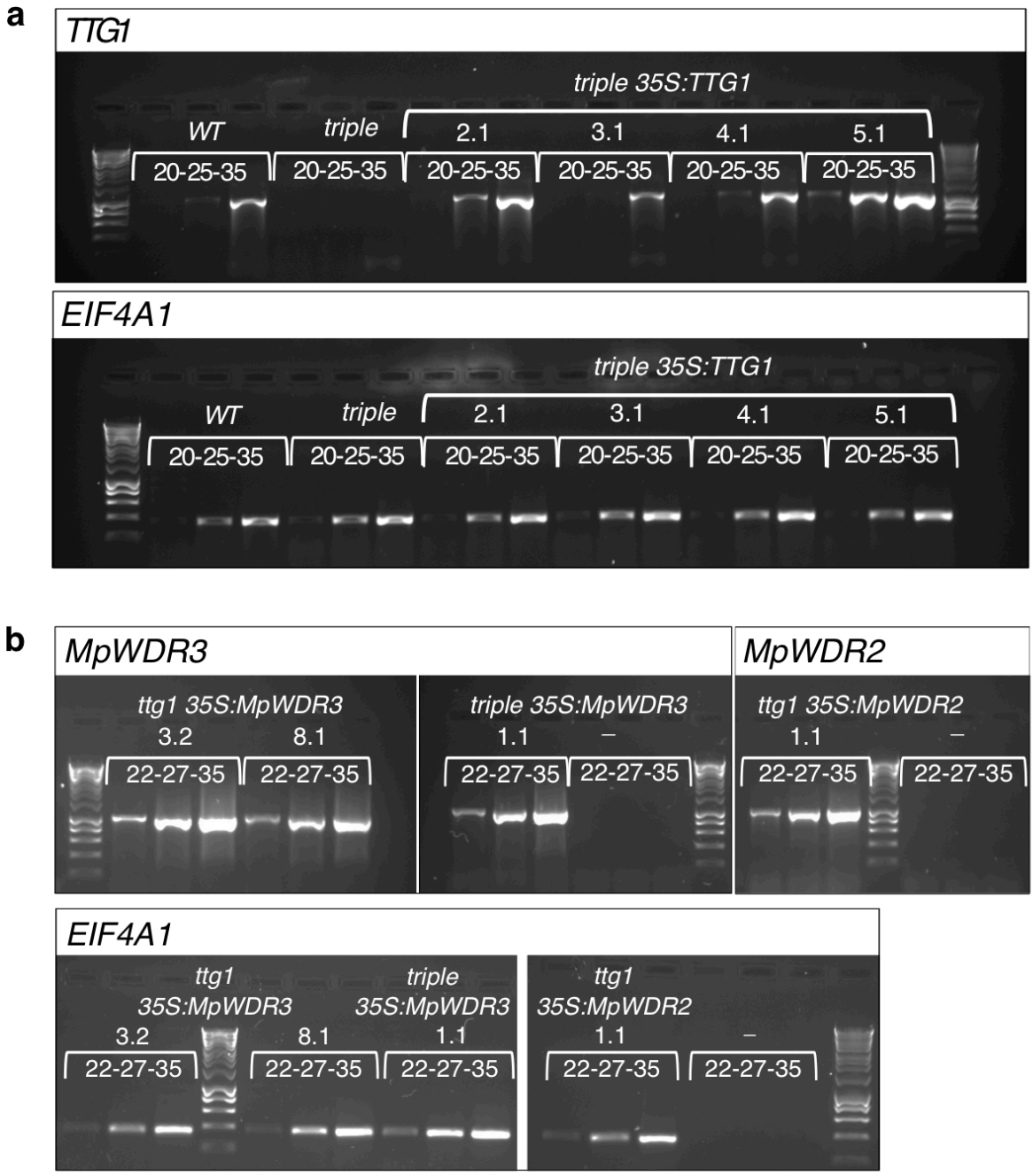


d

	triple	*ttg1*	*lwd1 ttg1*
ttg1	2.17E-10		
lwd1 ttg1	1.26E-04	1.03E-02	
lwd2 ttg1	6.69E-05	3.56E-02	0.663969

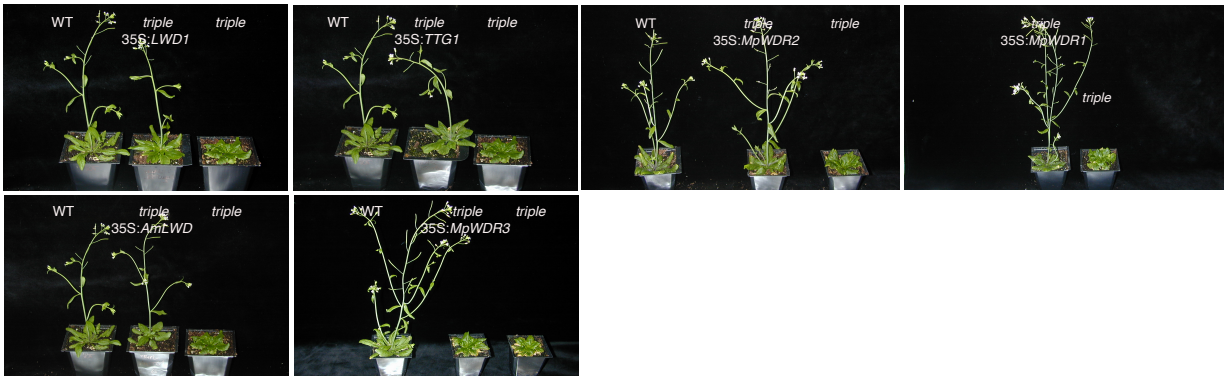
p<0.001
p<0.01
p<0.05

Extended Data Fig. 8

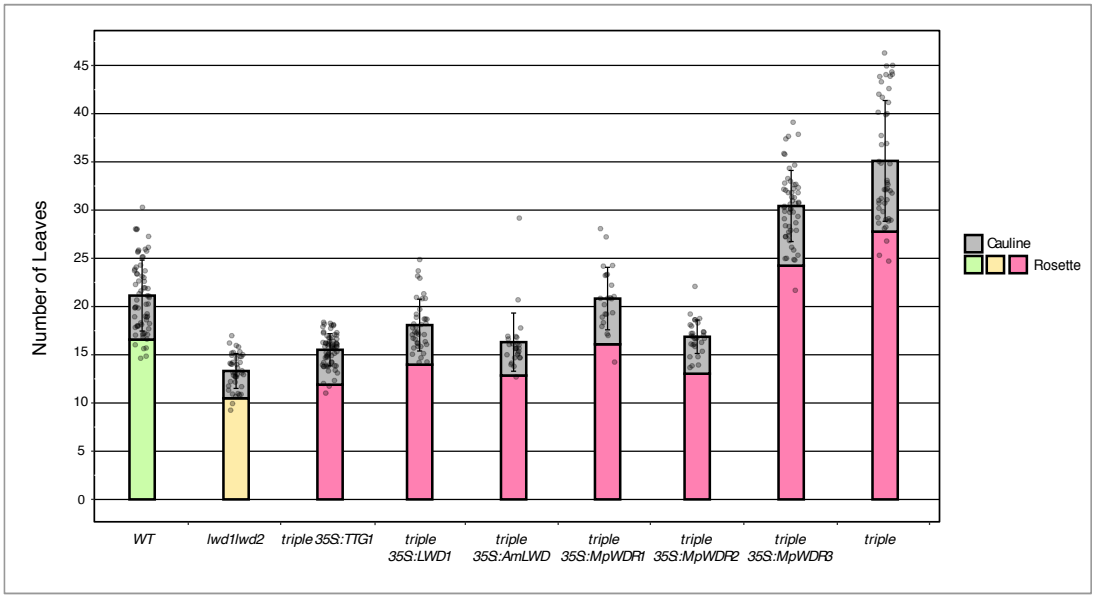


Extended Data Fig. 9

a



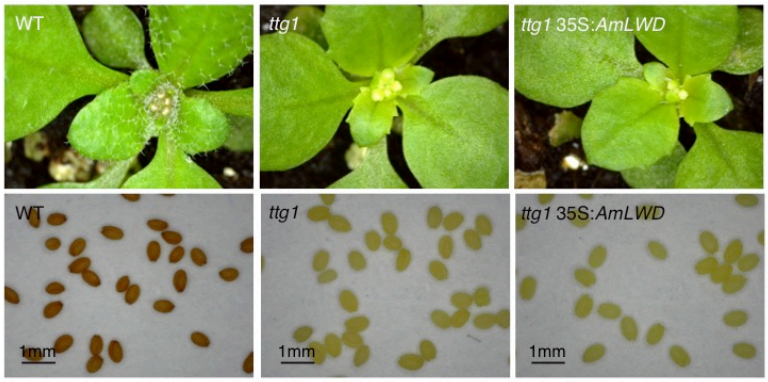
b



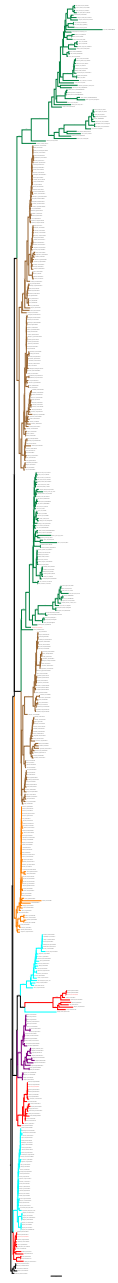
p<0.001
p<0.01
p<0.05

	WT	*lwd1 lwd2*	*triple*	*triple 35S:AmLWD*	*triple 35S:LWD1*	*triple 35S:MpWDR1*	*triple 35S:MpWDR2*	*triple 35S:MpWDR3*	*triple 35S:TTG1*
WT		5.20E-15	1.60E-07	2.24E-04	1.27E-01	1.00E+00	1.28E-02	2.56E-05	4.48E-09
lwd1 lwd2	5.20E-15		2.25E-36	1.27E-01	7.64E-06	8.66E-09	2.63E-03	9.45E-33	9.39E-02
triple	1.60E-07	2.25E-36		1.28E-16	3.46E-12	2.39E-04	1.13E-13	1.00E+00	8.26E-30
triple 35S:AmLWD1	2.24E-04	1.27E-01	1.28E-16		3.35E-01	6.37E-03	1.00E+00	4.39E-14	1.00E+00
triple 35S:LWD1	1.27E-01	7.64E-06	3.46E-12	3.35E-01		4.23E-01	1.00E+00	9.78E-10	3.23E-02
triple 35S:MpWDR1	1.00E+00	8.66E-09	2.39E-04	6.37E-03	4.23E-01		9.85E-02	5.11E-03	1.01E-04
triple 35S:MpWDR2	1.28E-02	2.63E-03	1.13E-13	1.00E+00	1.00E+00	9.85E-02		2.99E-11	5.06E-01
triple 35S:MpWDR3	2.56E-05	9.45E-33	1.00E+00	4.39E-14	9.78E-10	5.11E-03	2.99E-11		3.67E-26
triple 35S:TTG1	4.48E-09	9.39E-02	8.26E-30	1.00E+00	3.23E-02	1.01E-04	5.06E-01	3.67E-26	

Extended Data Fig. 10

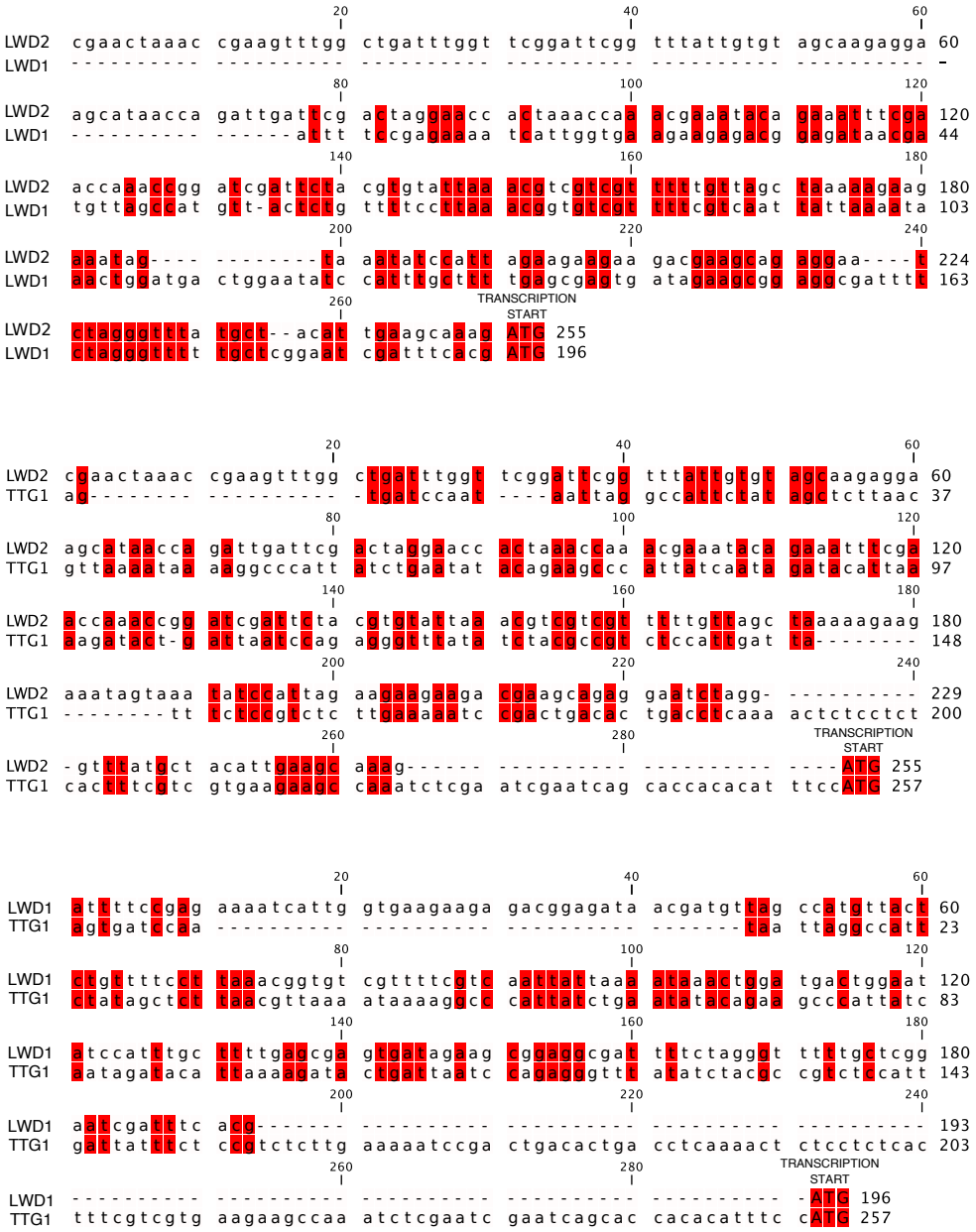


Supplementary fig. 1



730 **Supplementary fig. 2**

731



732

733 **Supplementary Table 1**

734

Primer name	*Description*	*Sequence*
AC1	cloning AtTTG1 for + PstI	**GGGGCTGCAG**ATGGATAATTCAGCTCCAGATTCG
AC2	cloning AtTTG1 rev + EcoRI	**GGGGGAATTC**TCAAACTCTAAGGAGCTGC
AC3	cloning MpWDR2 for + HindIII	**GGGGAAGCTT**ATGGACGGGTTCTCACAAGAACC
AC4	cloning MpWDR2 rev + EcoRI	**GGGGGAATTC**TCACACTCGTAGAATTTGGAG
AC5	cloning MpWDR1 for + PstI	**GGGGCTGCAG**ATGGCGAGCGAGAAGAAGGATG
AC6	cloning MpWDR1 rev + PstI	**GGGGGAATTC**TCACACCCTGAGAATCTGCAAC
AC7	cloning AtLWD1 for + PstI	**GGGGCTGCAG**ATGGGAACGAGCAGCGATC
AC8	cloning AtLWD1 rev + EcoRI	**GGGGGAATTC**TCAAACCCTGAGAATTTGCAG
AC9	cloning AtLWD2 for + HindIII	**GGGGAAGCTT**ATGGTTACGAGCAGCGATC
AC10	cloning AtLWD2 rev + EcoRI	**GGGGGAATTC**TCAGACCCGGAGAATCTGC
AC123	cloning MpWDR3 for + BamHI	**GGGGGATCC**ATGTCGAACCGAAATAGAACC
AC124	cloning MpWDR3 Rev + SmaI	**GGGCCCGGG**TCATATTCTTAAAAGCTGCAGCTC
AC97	cloning AmLWD for + PstI	**GGGGCTGCAG**ATGGCCAATGACCAGAGCC
AC98	cloning AmLWD rev + EcoRI	**GGGGGAATTC**TTAGACCCTCAAAATCTGAAGC
AC29	LWD1 genotyping forward	GTTCCGATGAGCAGCAGAAGC
AC30	LWD1 genotyping reverse	GCTCAATCTCAGCACCAGCAG
AC31	LWD2 genotyping for	CAAAATGGTTCTGAAGAGCAATC
AC32	LWD2 genotyping rev	TACTAGAAGTACCAATTCGTCTAG
AC33	ttg21 genotyping for	CAGCACCACACATTTCCATG
AC34	ttg21 genotyping rev	AGTCTCAACAACAGACTTCTC
AC35	GABI KAT 8409 LB	ATATTGACCATCATACTCATTGC
EIF4A1F	EIF4A1F	TCTTGGTGAAGCGTGATGAG
EIF4A1R	EIF4A1R	GCTGAGTTGGGAGATCGAAG
AC286	LWD2 qPCR for	TCTTCCCATATCTGCTCCGCTG
AC287	LWD2 qPCR rev	TCAACCTCAGCGCCGGCTG
AC288	UBQ10 qPCR for	GGCCTTGTATAATCCCTGATGAATAAG
AC289	UBQ10 qPCR rev	GAAAGAGATAACAGGAACGGAAACATAG

>2X35S promoter + multiple cloning site + 35S terminator
GGTACCCCTACTCCAAAAATGTCAAAGATACAGTCTCAGAAGACCAAAGGGCTATTGAGACTTTTCA
ACAAAGGGTAATTTCGGGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTCACTTCATCGAAAGG
ACAGTAGAAAAGGAAGGTGGCTCCTACAAATGCCATCATTGCGATAAAGGAAAGGCTATCATTCAA
GATGCCTCTGCCGACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGA
AGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGACATCTCCACTGACGTAAGGGATGA
CGCACAATCCCACCCCTACTCCAAAAATGTCAAAGATACAGTCTCAGAAGACCAAAGGGCTATTGA
GACTTTTCAACAAAGGGTAATTTCGGGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTCACTTC
ATCGAAAGGACAGTAGAAAAGGAAGGTGGCTCCTACAAATGCCATCATTGCGATAAAGGAAAGGCT
ATCATTCAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTG
GAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGACATCTCCACTGACGTAA
GGGATGACGCACAATCCCACTATCCTTCGCAAGACCCTTCCTCTATATAAGGAAGTTCATTTCATTT
GGAGAGGACAGCCCAAGCTTGGCTGCAGGTCGACGGATCCCCGGGAATTCGGTACGCTGAAATCA
CCAGTCTCTCTCTACAAATCTATCTCTCTCTATTTTCTCCATAAATAATGTGTGAGTAGTTTCCCGATA
AGGGAAATTAGGGTTCTTATAGGGTTTCGCTCATGTGTTGAGCATATAAGAAACCCTTAGTATGTAT
TTGTATTTGTAAAATACTTCTATCAATAAAATTTCTAATTCCTAAAACCAAAATCCAGTACTAAAATCC
AGATCC

Supplementary Table 1. List of primers used in this study and sequence of the 2X35S promoter cassette used for the overexpression analyses.

735 **Supplementary Table 2**

736

Genotype containing CCA1:LUC	***n=***	**Number of plants for which FFT-NLLS provided an estimate of the period (Number of plants with RAE <0.5)**	**Mean period length (± SEM)**	**Mean RAE (± SEM)**
Data in figure 2				
Col-0	24	24 (24)	26.4 ± 0.2 h	0.14 ± 0.01
lwd1 lwd2 ttg1	24	21 (0)	19.5 ± 1.0 h	0.65 ± 0.03
Data in figure 3				
Col-0	11	11 (11)	25.0 ± 0.8 h	0.35 ± 0.03
lwd1	11	11 (11)	23.0 ± 0.6 h	0.25 ± 0.03
lwd2	11	11 (11)	24.1 ± 0.2 h	0.25 ± 0.03
ttg1	11	11 (11)	23.4 ± 0.3 h	0.20 ± 0.02
lwd1 lwd2	5	5 (5)	17.0 ± 0.2 h	0.38 ± 0.09
lwd1 ttg1	12	12 (12)	24.9 ± 0.3 h	0.25 ± 0.03
lwd2 ttg1	11	11 (11)	24.3 ± 0.1 h	0.24 ± 0.03
lwd1 lwd2 ttg1	5	5(0)	20.4 ± 2.0 h	0.85 ± 0.05
Data in figure 5				
Col-0	12	12 (12)	26.2 ± 0.3 h	0.23 ± 0.01
lwd1 lwd2 ttg1	12	11 (0)	27.2 ± 0.6 h	0.72 ± 0.03
lwd1 lwd2	12	12 (10)	18.4 ± 0.9 h	0.36 ± 0.05
lwd1 lwd2 ttg1 35S::LWD1	12	12 (12)	26.3 ± 0.6 h	0.24 ± 0.02
lwd1 lwd2 ttg1 35S::TTG1	12	12 (12)	20.7 ± 0.8 h	0.28 ± 0.02
lwd1 lwd2 ttg1 35S::TTG1-3	5	5(5)	16.6 ± 0.2 h	0.26 ± 0.03
lwd1 lwd2 ttg1 35S::TTG1-4	6	6 (4)	15.8 ± 0.1 h	0.44 ± 0.07
Data in figure 6				
Col-0	7	7 (7)	24.8 ± 0.2 h	0.17 ± 0.01
lwd1 lwd2 ttg1	4	3 (0)	16.4 ± 0.6 h	0.68 ± 0.07
lwd1 lwd2	8	8 (8)	16.9 ± 0.2 h	0.23 ± 0.02
lwd1 lwd2 ttg1 35S::MpWDR1	12	12 (12)	26.5 ± 0.6 h	0.18 ± 0.01
lwd1 lwd2 ttg1 35S::MpWDR2	12	12 (12)	25.7 ± 0.4 h	0.20 ± 0.02
lwd1 lwd2 ttg1 35S::MpWDR3	3	3 (1)	15.6 ± 0.1 h	0.51 ± 0.14
lwd1 lwd2 ttg1 35S:: AmLWD	11	11 (11)	24.4 ± 0.9 h	0.35 ± 0.02
Data in extended data 6				
Col-0	7	7 (7)	25.3 ± 0.7 h	0.07 ± 0.01
lwd1	7	7 (7)	23.3 ± 0.8 h	0.07 ± 0.01
lwd1 ttg1	7	7 (7)	25.6 ± 0.6 h	0.09 ± 0.00
ttg1	3	3 (3)	25.8 ± 1.2 h	0.06 ± 0.02

Supplementary Table 2. Summary of period and RAE data from circadian clock experiments. Period values and statistics from FFT-NLLS analysis.

Supplementary Table 3

Extended data figure 3

a:

genotype	count	mean	std	min	25%	50%	75%	max
WT	12.0	63.833333	7.814129	57.0	58.00	63.5	65.25	86.0
ttg1	7.0	96.857143	9.822229	85.0	90.00	94.0	104.00	111.0
ttg1 35S::LWD1	8.0	89.375000	9.022789	75.0	85.00	90.0	96.00	102.0
ttg1 35S::TTG1	9.0	68.111111	9.439868	56.0	60.00	67.0	75.00	84.0
ttg1 35S::WDR1	10.0	72.700000	21.700230	59.0	60.25	64.0	72.50	130.0

Kruskal-Wallis rank sum test
Kruskal-Wallis chi-squared (H) = 25.385, df = 4, p-value = 4.21e-05

epsilon.squared	lower.ci	upper.ci
0.81	0.748	0.964

Freeman.theta
0.697

	gts	n	Median	Conf.level	Percentile.lower	Percentile.upper
1	ttg1	7	94.0	0.95	89	108.0
2	ttg1 35S::LWD1	8	90.0	0.95	79	96.0
3	ttg1 35S::TTG1	9	67.0	0.95	58	77.0
4	ttg1 35S::WDR1	10	64.0	0.95	60	78.0
5	WT	12	63.5	0.95	58	65.5

######################
b:

genotypes	count	mean	std	min	25%	50%	75%	max
WT	13.0	72.461538	7.042909	52.0	71.00	73.0	77.00	80.0
ttg1	10.0	92.900000	9.960477	80.0	83.50	96.5	99.50	105.0
ttg1 35S::LWD2	14.0	94.000000	7.379806	80.0	89.00	95.5	98.75	106.0
ttg1 35S::TTG1	11.0	74.818182	5.844967	60.0	73.00	76.0	78.00	82.0
ttg1 35S::WDR2	12.0	68.416667	6.748176	59.0	62.75	67.5	75.00	78.0

Kruskal-Wallis rank sum test
Kruskal-Wallis chi-squared = 43.892, df = 4, p-value = 6.756e-09

epsilon.squared	lower.ci	upper.ci
0.776	0.669	0.94

Freeman.theta
0.674

	gts	n	Median	Conf.level	Percentile.lower	Percentile.upper
1	ttg1	10	96.5	0.95	83.0	102
2	ttg1 35S::LWD2	14	95.5	0.95	89.0	99
3	ttg1 35S::TTG1	11	76.0	0.95	73.0	78
4	ttg1 35S::WDR2	12	67.5	0.95	62.5	75
5	WT	13	73.0	0.95	71.0	77

##########################
c:

genotype	count	mean	std	min	25%	50%	75%	max
WT	9.0	68.777778	6.320162	59.0	65.00	68.0	72.0	78.0
ttg1	6.0	97.500000	7.867655	86.0	95.25	96.5	100.0	110.0
ttg1 35S::AmLWD	6.0	96.166667	9.948199	80.0	93.25	95.5	103.0	108.0
ttg1 35S::WDR3	6.0	84.666667	3.076795	80.0	83.25	85.0	86.0	89.0

Kruskal-Wallis rank sum test

```
Kruskal-Wallis chi-squared = 20.746, df = 3, p-value = 0.0001189

epsilon.squared     lower.ci     upper.ci
0.913               0.808        1

Freeman.theta
        0.737

              gts n Median Conf.level Percentile.lower Percentile.upper
1            ttg1 6   96.5       0.95             90.5            106.0
2 ttg1 35S::AmLWD 6   95.5       0.95             86.5            106.0
3  ttg1 35S::WDR3 6   85.0       0.95             81.5             87.5
4              WT 9   68.0       0.95             65.0             78.0
```

Extended data figure 6

```
  n=     min          q1         med      q3     max   mean   stdev  n=
Col-     22.99       23.56         26     26.555 28.08  25.33  1.95   7
lwd1     21.3         22.275       22.81  23.61  27.33  23.32  2      7
lwd1ttg1 22.89       24.775       26.47   26.685 26.86  25.59  1.58   7
ttg1     24.19       24.695       25.2    26.62  28.04  25.81  2      3
```

T= 3.01798 P=0 .01169 df=12
The t-test was two sided

Extended data figure 7

```
          count       mean        std  min   25%   50%   75%   max
genotype
triple     17.0   3.411765   2.938087  0.0   1.0   3.0   4.0  11.0
ttg1       18.0  23.388889  11.985967  2.0  14.0  25.5  29.5  44.0
ttg1lwd1   13.0  10.461538   3.619746  4.0   8.0  12.0  12.0  16.0
ttg1lwd2   11.0  12.181818   5.862051  3.0   7.5  12.0  15.0  23.0


Kruskal-Wallis rank sum test
chi-squared (H-test) = 32.374, df = 3, p-value = 4.364e-07

Effects size (epsilon square 95% confidence)
 epsilon.squared    lower.ci     upper.ci
         0.722      0.632        0.944

       gts  n Median Conf.level Percentile.lower Percentile.upper
1   triple 17    3.0       0.95               1               4
2     ttg1 18   25.5       0.95              17              29
3 ttg1lwd1 13   12.0       0.95               8              12
4 ttg1lwd2 11   12.0       0.95               7              17
```

Extended data figure 9

```
genotype            count     mean       std   min   25%   50%    75%   max
WT                  66.0  21.136364  3.678990  15.0  18.0  21.0  24.00  30.0
lwd1 lwd2           41.0  13.317073  1.808854   9.0  12.0  13.0  15.00  17.0
triple              48.0  35.104167  6.264522  25.0  30.0  33.0  41.25  46.0
triple 35S::AmLWD1  26.0  16.269231  3.040496  13.0  15.0  16.0  16.75  29.0
triple 35S::LWD1    39.0  18.076923  2.689009  14.0  16.0  17.0  19.00  25.0
triple 35S::MpWDR1  23.0  20.826087  3.242504  14.0  19.0  21.0  23.00  28.0
triple 35S::MpWDR2  30.0  16.866667  1.736690  14.0  16.0  17.0  17.75  22.0
triple 35S::MpWDR3  54.0  30.425926  3.688850  22.0  28.0  30.5  32.00  39.0
triple 35S::TTG1    66.0  15.515152  1.666340  11.0  14.0  16.0  17.00  18.0


Kruskal-Wallis rank sum test
```

```
Kruskal-Wallis chi-squared = 315.32, df = 8, p-value < 2.2e-16

epsilon.squared     lower.ci     upper.ci
0.294               0.258        0.42

Freeman.theta
        0.377
```

	gts	n	Median	Conf.level	Percentile.lower	Percentile.upper
1	lwd1 lwd2	41	13.0	0.95	13.0	14.0
2	triple	48	33.0	0.95	31.0	37.5
3	triple 35S::AmLWD1	26	16.0	0.95	15.0	16.0
4	triple 35S::LWD1	39	17.0	0.95	17.0	19.0
5	triple 35S::MpWDR1	23	21.0	0.95	19.0	22.0
6	triple 35S::MpWDR2	30	17.0	0.95	16.0	17.0
7	triple 35S::MpWDR3	54	30.5	0.95	29.5	31.5
8	triple 35S::TTG1	66	16.0	0.95	15.0	16.0
9	WT	66	21.0	0.95	19.0	22.0

Additional details:

In Extended data figure 3, 7. The Kruskal-Wallis sum rank were used to test the null hypothesis that the population median of all of the groups are equal. This is a one way non-parametric version of ANOVA. The test works on 2 or more independent samples, which may have different sizes. Note that rejecting the null hypothesis does not indicate which of the groups differs. Post-hoc comparisons between groups are required to determine which groups are different. Dunn's test was employed as the post-hoc pairwise analysis except in 3 and 7 where the Conover test was used. The inflation ratio for the p values used in both cases was the Holm step-down method.

Supplementary Table 3. Details of statistical analysis conducted in Extended Data Figures 3, 6, 7 and 9.