# Supporting Information

## Rugnone et al. 10.1073/pnas.1302170110

### SI Materials and Methods

**Mutant Allele Confirmation of *Night Light–Inducible and Clock-Regulated* Genes.** Plants were grown in soil for 3 wk. Samples from four plants per genotype were collected to reduce biological variation. RNA was obtained using TRIzol reagent (Invitrogen). One microgram of RNA was treated with RQ1 RNase-Free DNase (Promega) and subjected to retrotranscription with Moloney Murine Leukemia Virus Reverse Transcriptase (M-MLV RT) (Invitrogen) and oligo-dT according to manufacturer's instructions. Transcript abundance of night light–inducible and clock-regulated genes (*LNK*) genes was determined by PCR to confirm the null or reduced expression in each line. *ACT2* was used as an expression control. Primers used for *LNK* cDNA expression analysis were as follows: *lnk1-1,* forward: AGGTGACTTAGGGTGGTTCTCTTC, reverse: CTGGCGAAAACTTGTTGCTTCC; *lnk1-2,* forward: TGGAGTTGCTAGTGGAAGAATC, reverse: TGTTCCCGTATCACGAGAAAC; *lnk1-3*, forward: AGCTTACAAGGTCCAACTGTTGAT, reverse: TGCGAGCCTGCGCCTTTCTC; *lnk2-1/lnk2-2/lnk2-3,* forward: GGTGAGAGGGTTGTTGAAGCATC, reverse: AAAGGAGCAGTCAGGAGATTGTTTC.

**Physiological Measurements.** Flowering time was estimated by counting the number of rosette leaves at the time of bolting. For hypocotyl length measurements, WT, *lnk1*, *lnk2,* and *lnk1;lnk2* seedlings were grown on 0.8% agar under complete darkness, continuous white light (LL), short day (8-h light/16-h dark) photoperiods, continuous red (100 $\mu mol \cdot m^{-2} \cdot s^{-1}$), or continuous blue light (10 $\mu mol \cdot m^{-2} \cdot s^{-1}$), and the final length of the hypocotyls was measured after 4 d. Light effects on hypocotyl elongation were calculated normalizing hypocotyl length under each light regime relative to hypocotyl length of the same genotypes under constant dark conditions. For leaf movement analysis, plants were grown under 16-h light/8-h dark cycles and transferred to continuous 20 $\mu mol \cdot m^{-2} \cdot s^{-1}$ white fluorescent light at 22 °C, and the position of the first pair of leaves was recorded every 2 h for 6 d using digital cameras and determined using Image J software.

**Subcellular Localization of LNK1.** Subcellular localization of LNK1 was determined by analyzing T1 transgenic plants of the WT Columbia accession transformed with the *35S:LNK1:YFP* construct. Hypocotyl tissue from 10-d-old plants was analyzed. Five independent transgenic lines were observed showing similar results. Transgenic plants were obtained using the floral dip transformation method (1). The 35S:LNK1:YFP construct was assembled as follows. The *AT5G64170.2* coding sequence, obtained from the *Arabidopsis* Information Resource (TAIR10), was synthesized de novo by GeneScript Corpoporation and then introduced into the pearly gate101 destination vector (2) using Gateway technology (Invitrogen). Imaging was performed using a LSM 510 META confocal microscope equipped with a 405-nm diode and an argon ion (488 nm) excitation laser system and a 40×, NA 1.3 objective. Images were processed with the LSM image browser software.

**Quantitative RT-PCR.** For time course analysis, 15-d-old plants were grown under 12-h light/12-h dark cycles at 22 °C and then transferred for 3 d to continuous white light at 22 °C. Samples were collected every 4 h for 2 d, starting 24 h after transfer to constant conditions. Total RNA was obtained from these samples using TRIzol reagent (Invitrogen). One microgram of RNA was treated with RQ1 RNase-Free DNase (Promega) and subjected to retrotranscription with M-MLV (Invitrogen) and oligo-dT according to the manufacturer's instructions. Synthesized cDNAs were amplified with FastStart Universal SYBR Green Master (Roche) using the Mx3000P Real Time PCR System (Agilent Technologies) cycler. The *PP2A* (*AT1G13320*) transcript was used as a housekeeping gene. Quantitative RT-PCR (qRT-PCR) quantification was conducted using the standard curve method as described in the Methods and Applications Guide from Agilent Technologies. Primer sequences and PCR conditions are available on request.

**Phylogenetic Analysis.** Homologs of *Arabidopsis thaliana LNK1* (splice variant 2) were identified using TBLASTN (www.phytozome.net/). Protein sequences were aligned using the Clustal Omega program. A maximum likelihood phylogenetic tree was built using SeaView Version 4 (3), with 1,000 boostrap replicates.

**Microarray Analysis.** Total RNA was extracted from the entire aerial structure of 15-d-old WT plants grown under 12-h light/12-h dark cycles at 22 °C. Triplicate samples were collected after a 1-h white light treatment (70 $\mu mol \cdot m^{-2} \cdot s^{-1}$) in the middle of the night or subjective day, and control samples kept in darkness were collected at the same time. Each replicate consisted of 10–12 plants to reduce biological variation. Total RNA was processed and hybridized to Affymetrix GeneChip *Arabidopsis* ATH1 Genome Arrays, according to the manufacturer's instructions. Data were analyzed using MAS5, and ANOVA was used to identify differentially expressed genes [q value $< 0.0005$ (4); fold change $\geq 2$]. Genes were then classified into different groups according to the relative effect of the light pulse given during the night compared with the effect of the same treatment given during the subjective day (Dataset S1). These groups include those in which the effect of light, i.e., induction or repression of gene expression, was at least twice as large during the subjective day than at night, those that showed a response that was at least twice as large during the night compared with subjective day, and finally, those in which the difference between the effect during the subjective day and night was not larger than twofold (Dataset S1). Microarray data have been deposited in Gene Expression Omnibus (GEO; accession nos. GSE46741 for the superseries and GSE46621 for the subseries corresponding to the microarray data set).

**Functional Category Enrichment Analysis.** Functional categories associated with specific groups of light-regulated genes were identified using the BioMaps tool from the virtual plant software (http://virtualplant.bio.nyu.edu/cgi-bin/vpweb). This tool allowed us to determine which functional categories were statistically overrepresented in particular lists of genes compared with the entire genome (5).

**Growth Conditions and Protocol Used for cDNA Library Preparation and High-Throughput Sequencing.** Seeds were sown onto Murashige and Skoog medium containing 0.8% agarose, stratified for 4 d in the dark at 4 °C, and then grown at 22 °C in continuous light or LD. Whole plants were harvested after 9 d,and total RNA was extracted with RNeasy Plant Mini Kit (QIAGEN) following the manufacturer's protocols. To estimate the concentration and quality of samples, NanoDrop 2000c (Thermo Scientific) and the Agilent 2100 Bioanalyzer (Agilent Technologies) with the Agilent RNA 6000 NanoKit were used, respectively. Libraries were prepared following the TruSeq RNA Sample Preparation Guide

(Illumina). Briefly, 3 μg of total RNA was polyA-purified and fragmented, and first-strand cDNA synthesized by reverse transcriptase (SuperScript II; Invitrogen) and random hexamers. This was followed by RNA degradation and second-strand cDNA synthesis. End repair process and addition of a single A nucleotide to the 3′ ends allowed ligation of multiple indexing adapters. Then, an enrichment step of 12 cycles of PCR was performed. Library validation included size and purity assessment with the Agilent 2100 Bioanalyzer and the Agilent DNA 1000 kit (Agilent Technologies). Samples were pooled to create 17 multiplexed DNA libraries, which were single-end sequenced with an Illumina Genome Analyzer II kit on the Illumina GAIIx platform, providing 100-bp single-end reads.

**Processing of RNA Sequencing Reads.** RNA sequencing (RNA-seq) reads were analyzed using Illumina Pipeline version 1.3. Reads were quality-filtered using the standard Illumina process and demultiplexed with two allowed barcode mismatches. Sequence files were generated in FASTQ format. Table S1 provides a summary table of main read count statistics. Sequence data have been deposited in GEO (accession no. GSE43865). The TopHat suite (6) was used to map reads to the *A. thaliana* TAIR10 reference genome. Along with the prebuilt *A. thaliana* index, the reference genome was downloaded from ENSEMBL (December 2012). Default values for TopHat parameters were used with the exception of maximun intron length parameter, which was set to a value of 5,000 nt following estimated values reported in ref. 7.

**RNA-seq Data Processing and Differential Expression Analysis.** Several packages from the Bioconductor library (version 2.11) of the R (version 2.15) statistical analysis framework were used to quantify gene differential expression signals. Default values were used unless explicitly stated otherwise. Package easyRNAsEq. (8) was used to generate read count tables at the gene level from Binary Alignment Map (BAM) files. A nonspecific prefiltering step was then conducted to filter out genes with less than two counts per million reads present along the whole set of samples, resulting in 21,143 (22,628) of 33,602 genes that were considered for further analysis in the WT vs. *lnk1;lnk2* (time course) experiment. The subsequent normalization and statistical analysis of read count data were performed using the package edgeR (version 3.04) following general guidelines provided in refs. 9 and 10. First, differences in RNA composition for each library were taken into account through a normalization step using the trimmed mean of M values (TMM) methodology. Then, estimates of the dispersion parameter for each transcript were obtained in a two-step procedure using the functionality implemented in estimateGLMTrendedDisp and estimateGLMTagwiseDisp functions (10). To assess differential expression, a negative binomial generalized log-linear model was fitted to each gene read counts using the glmFit function. Finally we used glmLRT to conduct genewise statistical tests for the coefficient contrasts of interest.

Following this analysis pipeline, we found 806 genes differentially expressed between *lnk1;lnk2* and WT Col conditions [Benjamini-Hochberg false discover rate (FDR)-adjusted $P < 0.05$]. For the more complex time course experiment, statistical significance tests were performed for mean differences between WT and *lnk1;lnk2* mutant time courses, along with genotype–time interaction contrasts for time points 6, 10, 14, 18, and 22 h. We then focus our attention on genes that simultaneously fulfilled the following two conditions: transcripts should present large (fold change >1.5) and significant (Benjamini-Hochberg FDR-adjusted $P < 0.0001$) changes in *lnk1;lnk2* vs. WT mean expression level along the time course and at least one significant (Benjamini-Hochberg FDR-adjusted $P < 0.0001$) genotype–time interaction contrast. In this way, a subset of 387 transcripts were identified and considered for follow-up analysis.

**Expression Profile Clustering.** To analyze patterns of coordinated gene expression behavior, read counts were log-transformed after a minimal offset (offset level = 1e−6) was added to avoid zero count values. Then, for each transcript, an extended expression profile was defined concatenating the WT and *lnk1;lnk2* time course profiles. A correlation-based similarity measure between extended expression profiles $T_i$ and $T_j$ was considered, and the distance metric $d_{ij} = 0.5[1 - cor(T_i, T_j)]$ was used to perform a hierarchical clustering (complete linkage) of gene profiles. Finally, clusters of coordinated expression were obtained from the dendrogram structure with the aid of the dynamicTreeCut (11) package.

**ChIP Analysis.** ChIP assays were performed essentially as described in Huang et al. (12).


1. Clough SJ, Bent AF (1998) Floral dip: A simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J* 16(6):735–743.
2. Earley KW, et al. (2006) Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J* 45(4):616–629.
3. Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* 27(2):221–224.
4. Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* 100(16):9440–9445.
5. Katari MS, et al. (2010) VirtualPlant: A software platform to support systems biology research. *Plant Physiol* 152(2):500–515.
6. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* 25(9):1105–1111.
7. Hong X, Scofield DG, Lynch M (2006) Intron size, abundance, and distribution within untranslated regions of genes. *Mol Biol Evol* 23(12):2392–2404.
8. Delhomme N, Padioleau I, Furlong EE, Steinmetz LM (2012) easyRNASeq: A bioconductor package for processing RNA-Seq data. *Bioinformatics* 28(19):2532–2533.
9. McCarthy DJ, Chen Y, Smyth GK (2012) Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* 40(10):4288–4297.
10. McCarthy D, Chen Y, Smyth G (2012) *edgeR: Differential Expression Analysis of Digital Gene Expression Data User's Guide* Available at http://www.bioconductor.org/packages/release/bioc/vignettes/edgeR/inst/doc/edgeRUsersGuide.pdf.
11. Langfelder P, Zhang B, Horvath S (2008) Defining clusters from a hierarchical cluster tree: The Dynamic Tree Cut package for R. *Bioinformatics* 24(5):719–720.
12. Huang W, et al. (2012) Mapping the core of the Arabidopsis circadian clock defines the network structure of the oscillator. *Science* 336(6077):75–79.

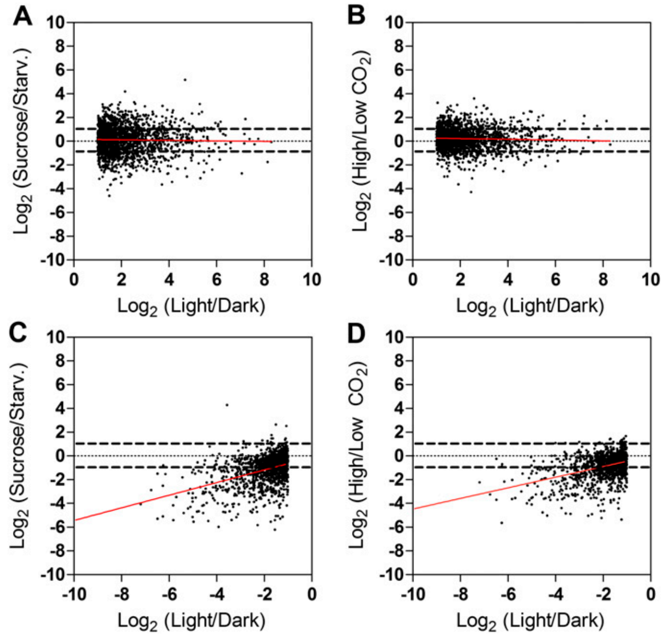


**Fig. S1.** Genomewide analysis of sucrose and $CO_2$ effects on gene expression of light-regulated genes. Comparative analysis of the effect on gene expression of a light pulse given during subjective day time (*x* axis) vs. the effect of sucrose added to plants undergoing starvation (*y* axis) or the effect of enhanced photosynthetic activity resulting from increased $CO_2$ levels. (*A* and *B*) Light-induced genes. (*C* and *D*) Light-repressed genes. Data for changes in gene expression resulting from added sucrose or enhanced photosynthetic activity were obtained from Osuna et al. (1). Data for light effects during the subjective day correspond to results described in Dataset S1.

1. Osuna D, et al. (2007) Temporal responses of transcripts, enzyme activities and metabolites after adding sucrose to carbon-deprived Arabidopsis seedlings. *Plant J* 49(3):463–491.

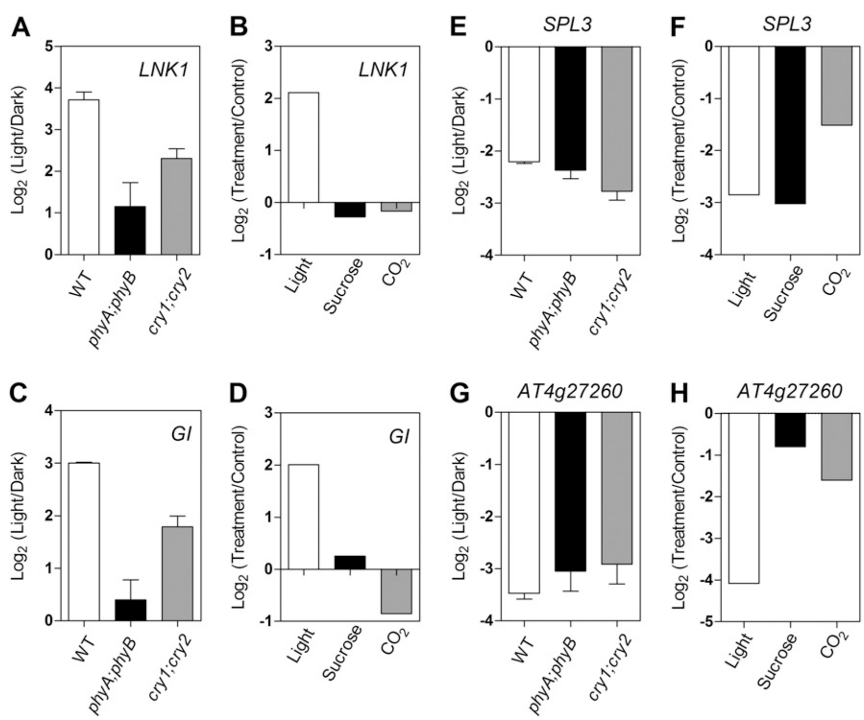


**Fig. S2.** Light effects on gene expression during the subjective day in plant photoreceptor mutants and comparison with the effect of sucrose and photosynthesis. Changes in gene expression measured by qRT-PCR in two light-induced (*A–D*) and two light-repressed (*E–H*) genes shown as fold change. (*A, C, E*, and *G*) Effect of a light pulse given in the middle of the subjective day on *AT5G64170* (*LNK1*) (*A*), *AT1G22770* (*GI*) (*C*), *AT2G33810* (*SPL3*) (*E*), and *At4g27260* (*G*) expression in WT, *phyA;phyB,* and *cry1;cry2* mutant plants. Data represent average ± SEM ($n$ = 3). (*B, D, F*, and *H*) Effect of sucrose or $CO_2$ addition on the expression of the same genes indicated above. Data for changes in gene expression resulting from added sucrose or enhanced photosynthetic activity were obtained from Osuna et al. (1).

1. Osuna D, et al. (2007) Temporal responses of transcripts, enzyme activities and metabolites after adding sucrose to carbon-deprived Arabidopsis seedlings. *Plant J* 49(3):463–491.

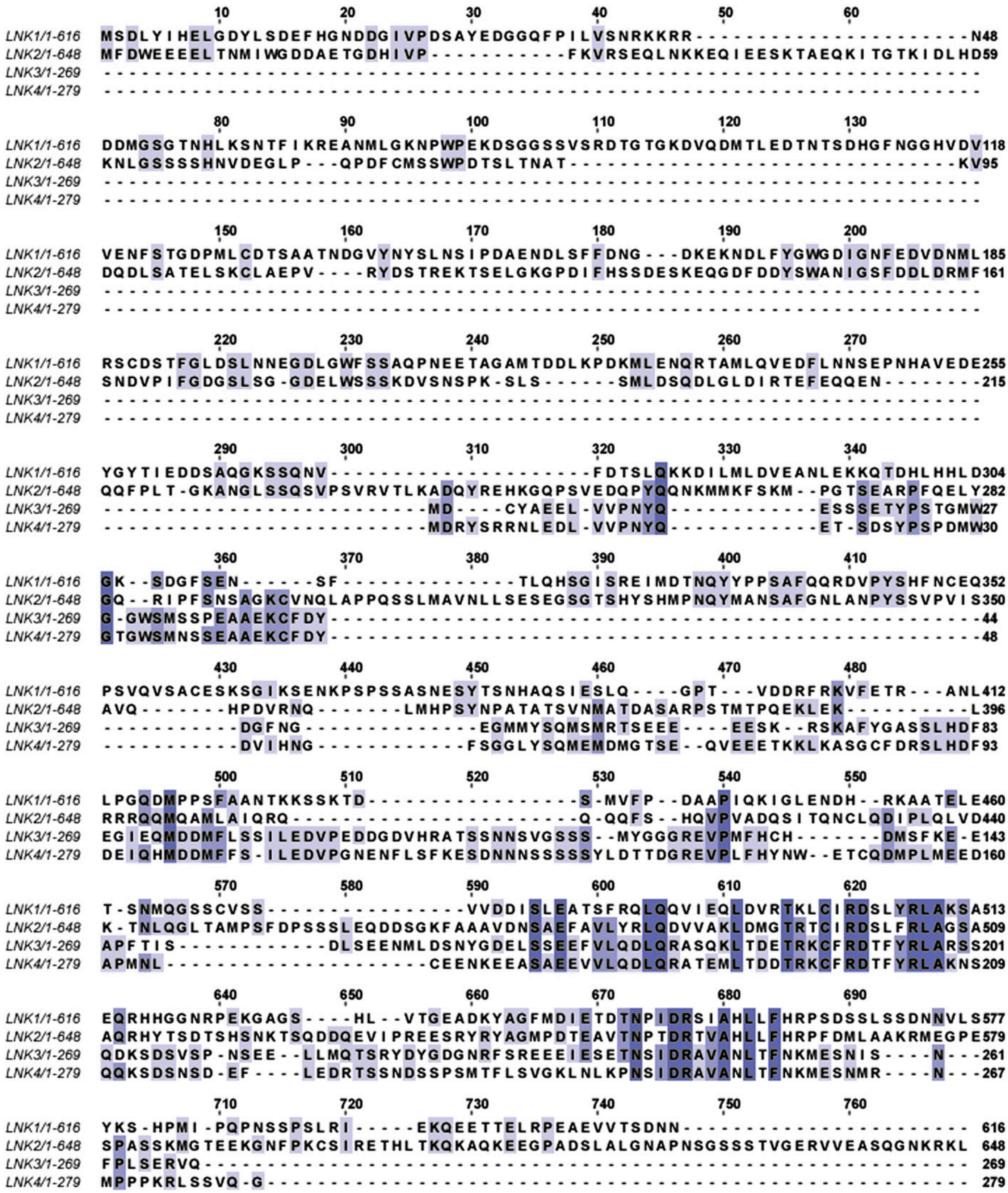


**Fig. S3.** Sequence alignment of *LNK* homologs in *Arabidopsis thaliana*. The length of the sequence aligned is shown, and degree of similarity between amino acids is highlighted (darker blue indicates higher similarity).

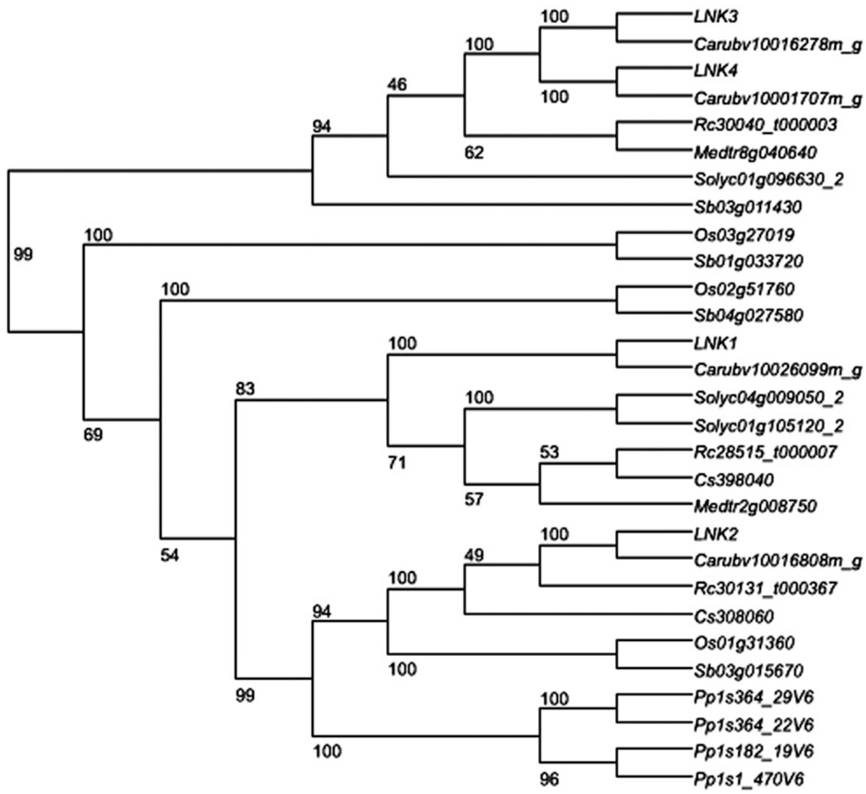


**Fig. S4.** Cladogram displaying *LNK1* homologs in a broad range of species from embryophyta group. *Arabidopsis thaliana LNK1* has three paralogs. Where multiple homologs were identified within a single species, the annotated gene model code is provided. Rc, *Ricinus communis*; Pp, *Physcomitrella patens*; Solyc, *Solanum lycopersicum*; Carubv, *Capsella rubella*; Os, *Oryza sativa*; Cs,*Cucumis sativus*; Medtr, *Medicago truncatula; Sb, Sorghum bicolor*. Percentage bootstrap values are presented for each node.

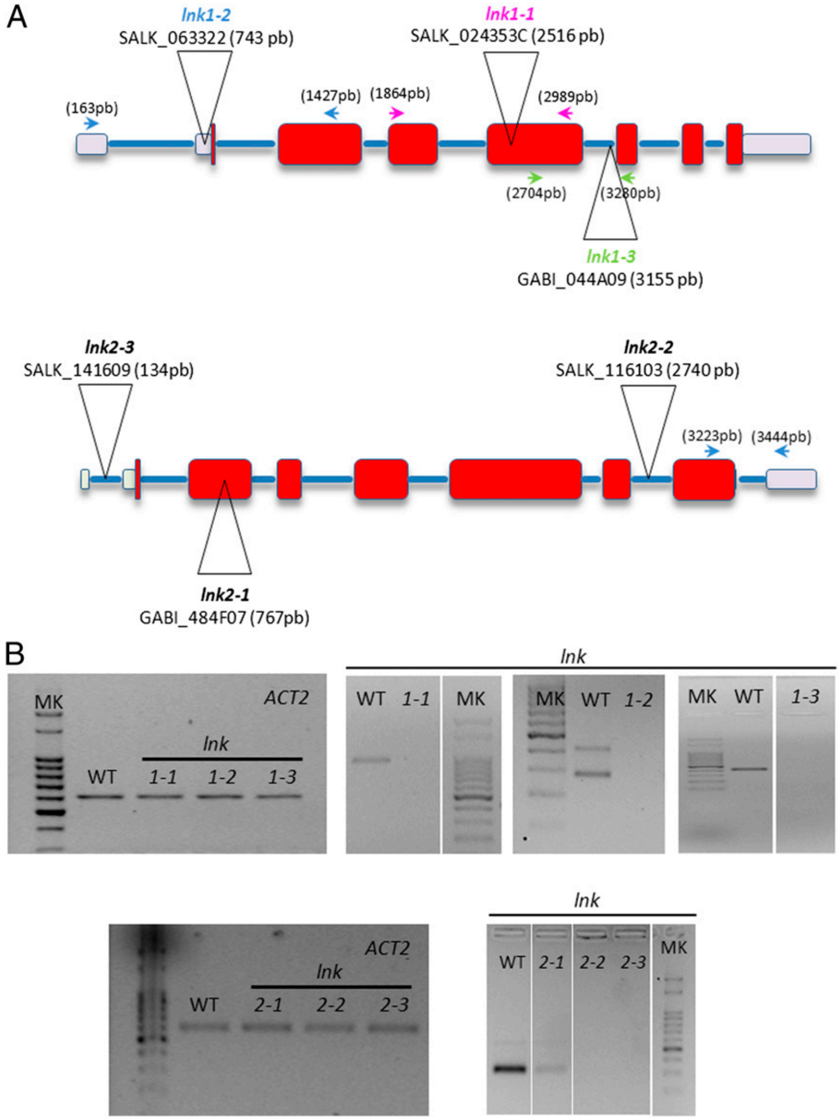


**Fig. S5.** *LNK1* and *LNK2* expression in different mutant backgrounds. (*A*) Scheme of *LNK1* and *LNK2* showing the site of T-DNA insertions in the different mutant alleles. (*B*) All mutant alleles have strongly reduced expression of the full-length mRNA, evaluated using primers flanking the T-DNA insertion. Plants were grown in soil for 3 wk in continuous light conditions. Samples harvested were processed until cDNA synthesis. Transcript presence was determined by PCR.

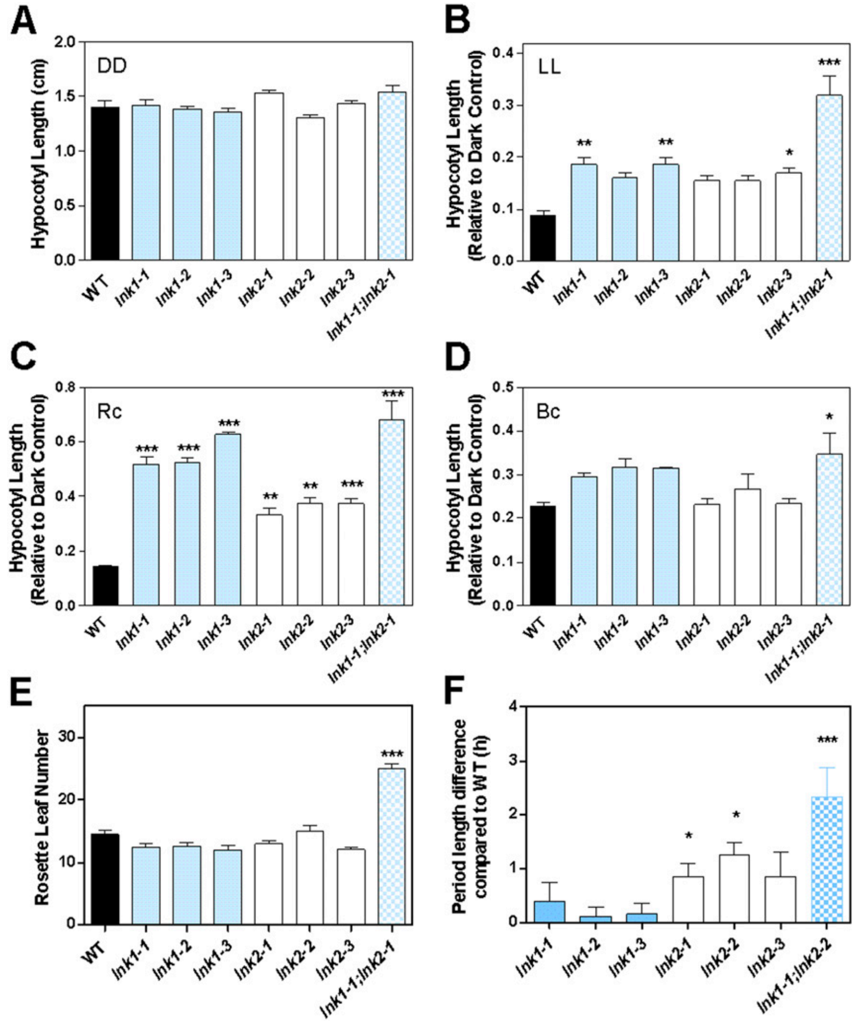


**Fig. S6.** Physiological characterization of different *lnk1* and *lnk2* mutant lines. (*A–D*) Hypocotyl length of plants grown for 4 d under different light conditions. (*A*) DD, continuous darkness. (*B*) LL, continuous white light. (*C*) Rc, continuous red light. (*D*) Bc, continuous blue light. Hypocotyl length under different light conditions is expressed relative to the hypocotyl length of each genotype under continuous darkness. (*E*) Flowering time measured as the number of rosette leaves at bolting in LL. (*F*) Period length differences between mutant and WT plants in the circadian rhythm of leaf movement (period length of WT plants = 24.2 ± 0.16; $n = 7$). Period length was calculated by BRASS 3.0 software. ANOVA followed by a Tukey's multiple comparison test was used to evaluate statistical significance of the difference with WT plants. Error bars indicate ±SEM (***$P < 0.001$, **$P < 0.01$, *$P < 0.05$).

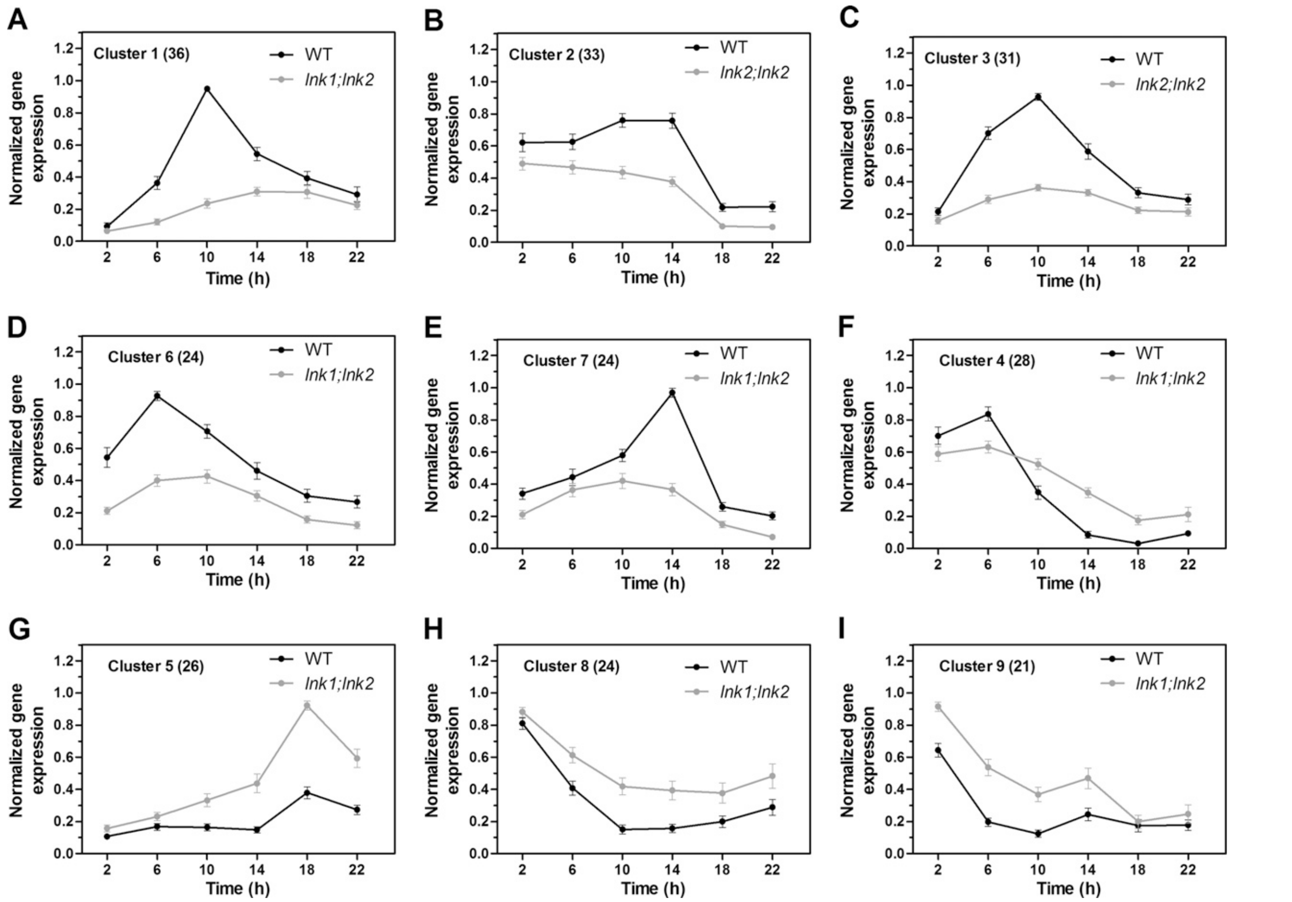


**Fig. S7.** Gene clusters identified by RNA-seq analysis. (*A*–*I*) Clusters of genes with similar expression patterns detected using a correlation-based distance metric and a hierarchical clustering procedure followed by a hybrid adaptive dendrogram cut step. Data sets represent the average of normalized expression level for all genes within each cluster. Number of genes in each cluster is indicated between parentheses. Plants were grown and harvested in 16-h light/8-h dark cycles. Clusters obtained for genes down-regulated (*A*–*E*) or up-regulated (*G*–*I*) in *lnk1;lnk2* mutant. (*F*) Cluster formed by genes with a significant alteration in the temporal pattern of expression but without large differences in expression levels between WT and *lnk1;lnk2*.

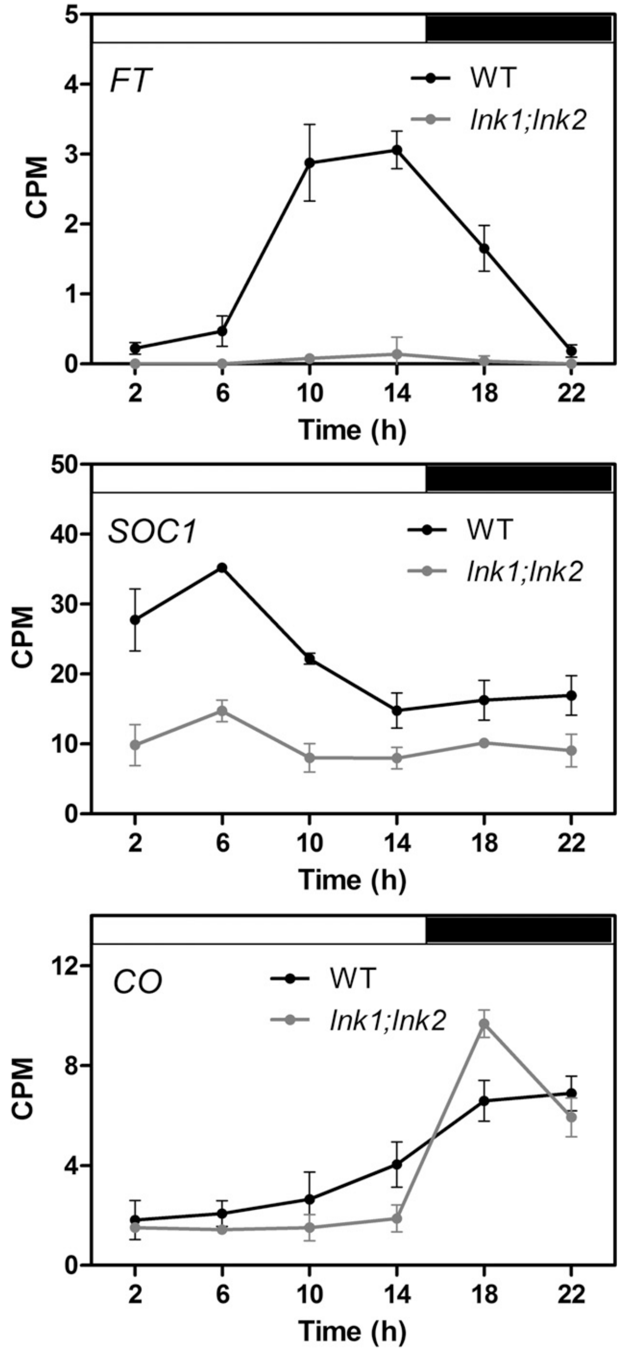


**Fig. S8.** Expression levels of flowering time genes in WT and *lnk1;lnk2* mutant plants. Data were from the RNA-seq experiment. cpm, counts per million.

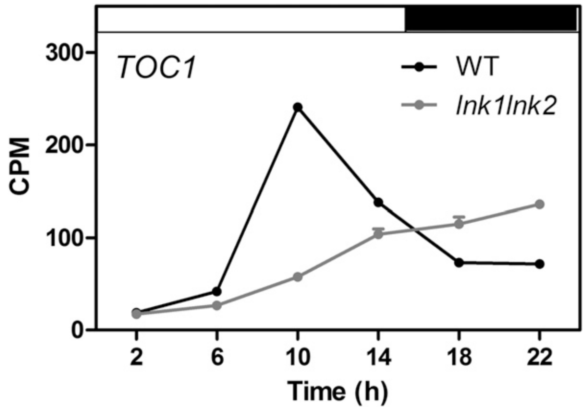


**Fig. S9.** *TOC1* expression under LD conditions in WT and *lnk1;lnk2* mutant plants. Data were from the RNA-seq experiment. cpm, counts per million.

**Table S1. Read counts summary stats**

| Experimental condition | Minimum library size | Maximum library size | Median library size | Correlation between replicates | Correlation between conditions |
|---|---|---|---|---|---|
| LL | 14,074,859 | 18,297,931 | 15,855,443 | 0.9893, 0.9986 | 0.9677, 0.9986 |
| Time course | 9,176,985 | 13,408,078 | 11,350,789.67 | 0.8758, 0.9995 | 0.827, 0.9995 |

**Dataset S1. Genes differentially regulated by light during the middle of the subjective day or night**

Dataset S1 (XLSX)

**Dataset S2. GO enrichment analysis of genes differentially regulated by light**

Dataset S2 (XLSX)

**Dataset S3. Genes differentially regulated in *lnk1-1;lnk2-1* mutants compared with WT plants**

Dataset S3 (XLSX)