nature genetics

Domestication selected for deceleration of the circadian clock in cultivated tomato

Niels A Müller[1], Cris L Wijnen[1,7], Arunkumar Srinivasan[1], Malgorzata Ryngajllo[1], Itai Ofner[2], Tao Lin[3,4], Aashish Ranjan[5,7], Donnelly West[5], Julin N Maloof[5], Neelima R Sinha[5], Sanwen Huang[3,4], Dani Zamir[2] & José M Jiménez-Gómez[1,6]

The circadian clock is a critical regulator of plant physiology and development, controlling key agricultural traits in crop plants[1]. In addition, natural variation in circadian rhythms is important for local adaptation[2–4]. However, quantitative modulation of circadian rhythms due to artificial selection has not yet been reported. Here we show that the circadian clock of cultivated tomato (*Solanum lycopersicum*) has slowed during domestication. Allelic variation of the tomato homolog of the *Arabidopsis* gene *EID1* is responsible for a phase delay. Notably, the genomic region harboring *EID1* shows signatures of a selective sweep. We find that the *EID1* allele in cultivated tomatoes enhances plant performance specifically under long day photoperiods, suggesting that humans selected slower circadian rhythms to adapt the cultivated species to the long summer days it encountered as it was moved away from the equator.

The circadian clock is an endogenous timekeeper that allows anticipation of daily and annual seasonal changes, which is key to synchronizing internal processes with the external environment. The occurrence of circadian timekeeping in all domains of life and the pervasiveness of its control indicate that it is of central importance for all living beings[5,6]. A functional circadian clock enhances fitness in animals, bacteria and plants[7–10]; fitness is further enhanced if the clock is tuned according to the specific environment. Such local adaptation is inferred from latitudinal clines of circadian variables[2–4]. Because crop plants are often spread over a broader geographical range than their ancestors were, it seems plausible that clock optimization could similarly enhance agricultural productivity[11]. Mutations in homologs of *Arabidopsis* clock genes in diverse crop plants have been found to change the two clock outputs flowering time and photoperiod sensitivity[12–18]. However, these mutations either do not change circadian rhythms[12–14,19] or almost completely abolish them[16,17]. To our knowledge, quantitative modulation of circadian rhythms due to domestication or breeding has not yet been reported.

We chose to study the circadian clock in the day-neutral crop plant tomato and its wild relatives to assess potential variation in circadian rhythms independently from photoperiodic flowering. To this end, we monitored circadian leaf movements of 102 diverse tomato accessions, including 34 varieties of the cultivated species *S. lycopersicum*, 44 accessions of its wild ancestor *Solanum pimpinellifolium* and 24 accessions for the more distantly related green-fruited wild tomato species. We observed marked differences in circadian period and phase among the three groups (**Fig. 1a**,**b**, **Supplementary Table 1** and **Supplementary Note**). The period (time for one complete cycle) of the cultivated tomato varieties was 2 h longer, and their phase (time of the highest peak) was delayed by more than 3 h, compared to the ancestral *S. pimpinellifolium* accessions. The distant tomato relatives showed a similar phase to that of *S. pimpinellifolium* but an even shorter period (**Fig. 1c**). Gene expression profiles of two putative core clock genes showed similar differences, suggesting that leaf movements reflect the molecular circadian clock (**Supplementary Figs. 1** and **2**). An RNA sequencing (RNA-seq) time course confirmed a deceleration of the molecular clock of cultivated tomato (**Supplementary Fig. 3a–c**). Additionally, lower gene expression amplitudes suggest that the clock is weaker in cultivated tomato than in the wild species (**Supplementary Figs. 3d** and **4–6**). The fact that period and phase together separate the cultivated tomato varieties from their wild ancestors in a dichotomous way demonstrates that the circadian clock of cultivated tomato has changed during domestication (**Fig. 1b**).

Tomato domestication probably began in the Andean region of Ecuador and Peru and was completed in Mesoamerica[20]. To assess when during domestication the circadian clock was changed, we compared tomatoes representing three sequential domestication

[1]Department of Plant Breeding and Genetics, Max Planck Institute for Plant Breeding Research, Cologne, Germany. [2]Institute of Plant Sciences, Hebrew University of Jerusalem, Rehovot, Israel. [3]Key Laboratory of Biology and Genetic Improvement of Horticultural Crops of the Ministry of Agriculture, Sino-Dutch Joint Laboratory of Horticultural Genomics, Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing, China. [4]Agricultural Genomic Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China. [5]Department of Plant Biology, University of California, Davis, California, USA. [6]Institut Jean-Pierre Bourgin, INRA, AgroParisTech, CNRS, Université Paris-Saclay, Versailles, France. [7]Present addresses: Laboratory of Genetics, Wageningen University, Wageningen, the Netherlands (C.L.W.); National Institute of Plant Genome Research, New Delhi, India (A.R.). Correspondence should be addressed to J.M.J.-G. (jmjimenez@mpipz.mpg.de).

Received 6 August; accepted 21 October; published online 16 November 2015; doi:10.1038/ng.3447

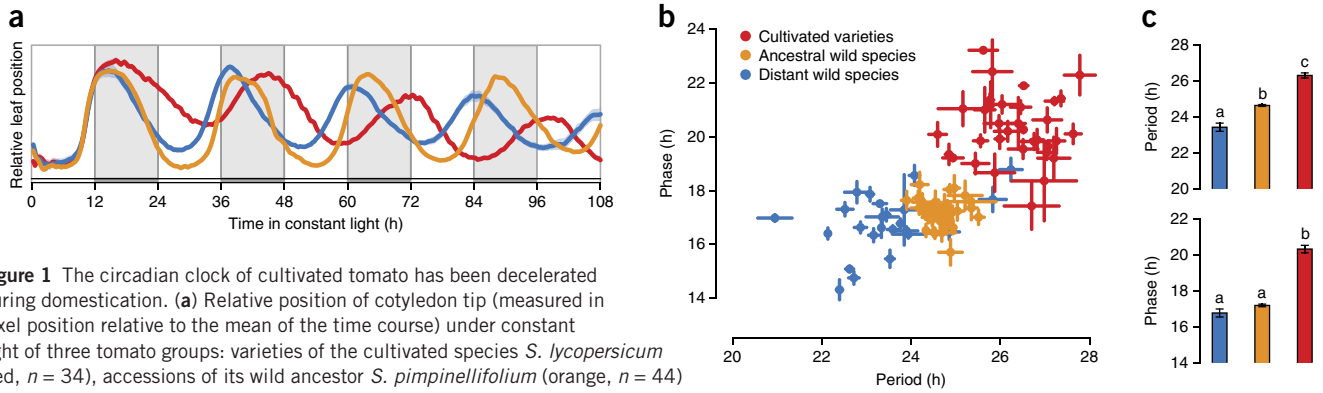

Figure 1 The circadian clock of cultivated tomato has been decelerated during domestication. (**a**) Relative position of cotyledon tip (measured in pixel position relative to the mean of the time course) under constant light of three tomato groups: varieties of the cultivated species *S. lycopersicum* (red, $n = 34$), accessions of its wild ancestor *S. pimpinellifolium* (orange, $n = 44$) and accessions of the distantly related green-fruited tomato species (blue, $n = 24$). Lines and shading show mean and s.e.m.; shaded areas indicate subjective nights. Data are from eight independent experiments. (**b**) Circadian period and phase estimates (mean ± s.e.m., $n = 2$–5) of the 102 genotypes shown in **a**. (**c**) Circadian period and phase estimates for the same three groups in **a**. Different letters indicate significant differences ($P < 0.05$, one-way ANOVA and Tukey's *post hoc* HSD test).

steps with *S. pimpinellifolium* and with 'modern' cultivars. The earliest domesticated types, the Ecuadorian cherry tomatoes, showed a delayed phase; however, their period was similar to that of their wild ancestor (**Supplementary Fig. 7a**). Plants with long periods arose during the next step of domestication and can be found among Mesoamerican cherry tomatoes and cultivars (**Supplementary**

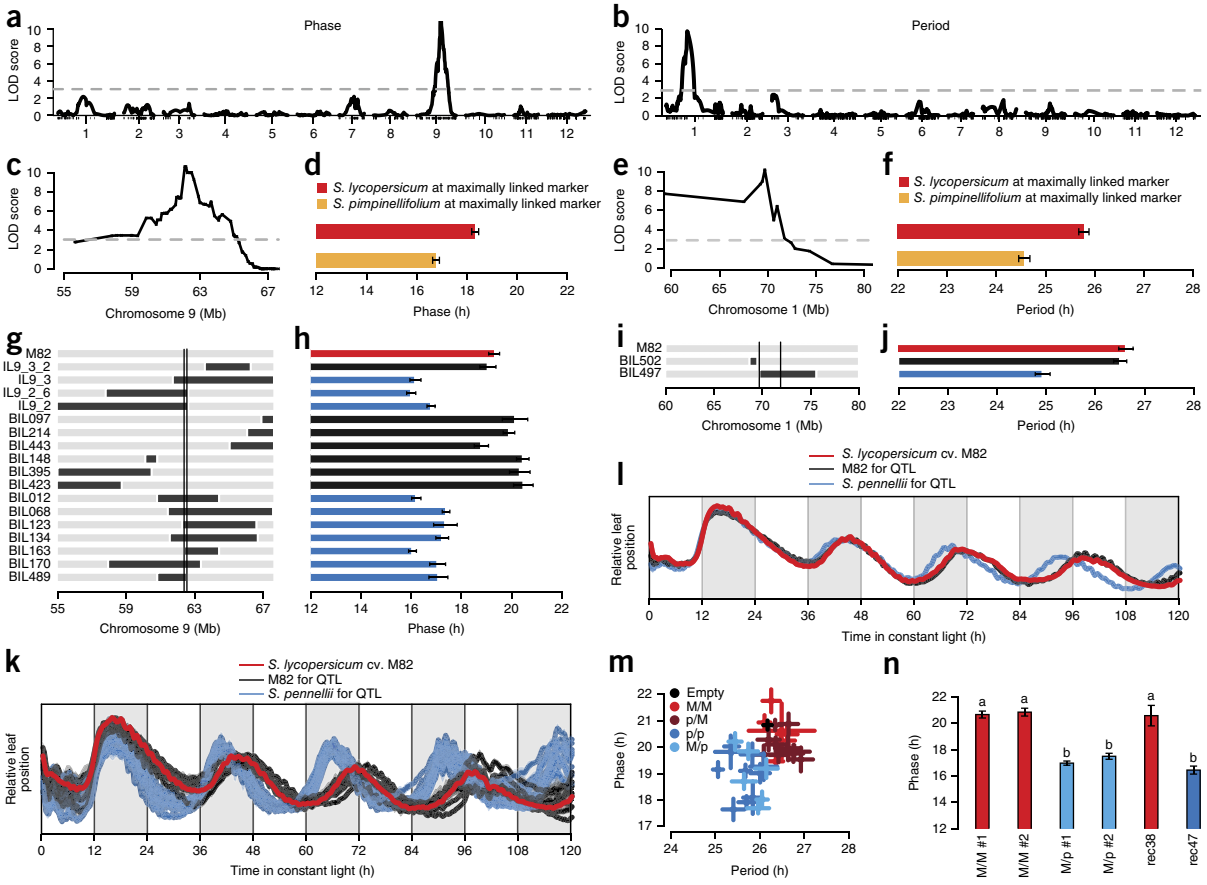

Figure 2 Two loci are responsible for variation in circadian rhythms between wild and cultivated tomato. (**a**,**b**) Logarithm of odds (LOD) scores for phase (**a**) and period (**b**) along the genetic map. Dashed line indicates the 5% significance threshold. (**c**–**f**) LOD scores and significance threshold for phase and period along the physical map at the QTL regions (**c**,**e**) and phase and period estimates (mean ± s.e.m.) (**d**,**f**) of all RILs grouped by their genotype at the QTL. Data are from five independent experiments. (**g**–**j**) Genotypes (**g**,**i**) and phase and period estimates (**h**,**j**) of selected ILs and BILs. Gray indicates *S. lycopersicum*; black, *S. pennellii*; and vertical lines outline the QTL in **g** and **i**; lines are colored according to genotype at the QTL in **h** and **j**. $n = 10$–22, mean ± s.e.m. (**h**,**j**). Each line was analyzed in at least two independent experiments. (**k**,**l**) Mean relative position of cotyledon tip of the lines shown in **g**–**j**. Shaded areas indicate subjective nights. (**m**) Period and phase estimates of independent transgenic T2 populations transformed with different *EID1* promoter–cDNA combinations ($n = 4$–12) or empty vector ($n = 31$; mean ± s.e.m.). M, *S. lycopersicum* cv. M82; p, *S. pennellii*. Data are from three independent experiments. (**n**) Phase estimates of four homozygous transgenic lines and two recombinants (mean ± s.e.m., $n = 7$–12). Different letters indicate significant differences ($P < 0.05$, one-way ANOVA and Tukey's *post hoc* HSD test).

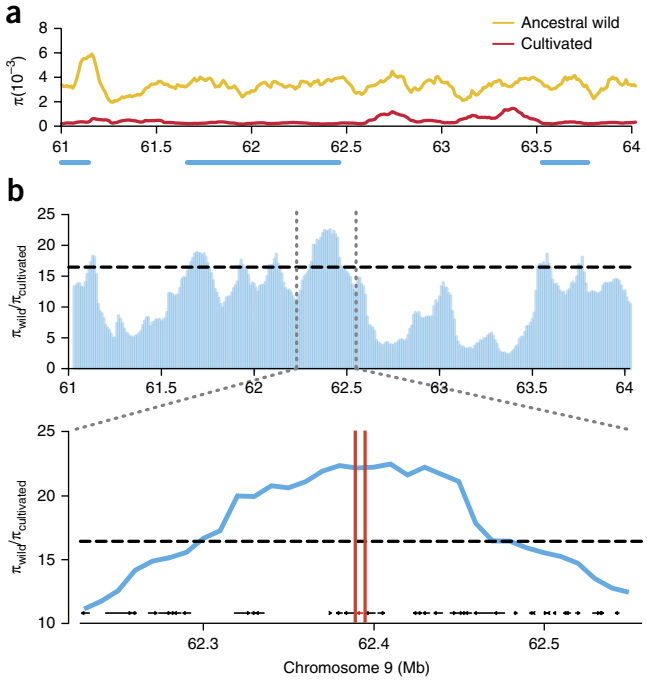

Figure 3 The *EID1* region shows evidence of a selective sweep. (**a**) Genetic diversity (π) of 166 *S. lycopersicum* varieties (cultivated) and 53 *S. pimpinellifolium* accessions (wild). Blue lines (bottom) indicate putative selective sweeps as determined by the π ratio. (**b**) π ratio for a region on chromosome 9 in the wild tomato ancestor *S. pimpinellifolium* (π_{wild}) and the cultivated species *S. lycopersicum* ($\pi_{cultivated}$). Horizontal dashed lines indicate genome-wide top 5% ratio cutoff; horizontal arrows (bottom) indicate locations of genes; vertical red lines (bottom) delimit *EID1*.

Fig. 7b,**c**). Notably, the changes in phase and period became fixed in 'modern' cultivars that originated after the tomato was brought to Europe, about 500 years ago. These results show that the deceleration of the circadian clock was a stepwise process that started very early in domestication.

Although variation in circadian rhythms is observable only under artificial constant conditions, it has been shown in various organisms to affect fitness under natural diurnal conditions[8,10] and has been proposed to be of adaptive significance[2,3]. This fitness effect can be explained by subtle changes in the timing of rhythmic processes under diurnal conditions caused by variation in circadian variables such as period and phase. To evaluate whether the differences in circadian rhythms have an effect on tomato under natural conditions, we analyzed the diurnal transcriptomes of the cultivated tomato variety M82 and the distant wild relative *Solanum pennellii*. The phase distribution of the 2,368 genes cycling in both species was bimodal, as described for other species[21]. Notably, the peaks of this distribution occurred 1–2 h earlier in the wild species (**Supplementary Fig. 8**), indicating that differences in circadian rhythms have a profound effect on the tomato transcriptome under natural light–dark cycles, potentially affecting many downstream processes.

To dissect the genetic architecture of the circadian rhythm differences and identify the underlying genes, we performed quantitative trait locus (QTL) analysis in two populations generated from crosses of the wild species *S. pimpinellifolium* and *S. pennellii* with the cultivated varieties Moneymaker and M82, respectively[22,23]. These analyses revealed a simple genetic architecture, with two QTLs shared among populations—one involved in phase and one in period (**Fig. 2a**,**b**). We fine-mapped the phase QTL to a region of ~0.3 Mb and the period QTL to a region of ~2 Mb using *S. pennellii* backcross inbred lines (BILs) (**Fig. 2c–l** and **Supplementary Table 2**).

We confirmed the phase QTL, which slightly affects period as well (**Supplementary Fig. 9**), and narrowed it further to a region containing 13 annotated genes with two recombinant lines (**Supplementary Fig. 10** and **Supplementary Table 3**). One of the genes is annotated as encoding a phytochrome A–associated F-box protein (Solyc09g075080) and is homologous to *EID1* in *Arabidopsis*. The *Arabidopsis* gene is a negatively acting component of the phytochrome signaling cascade[24], which has an important role in circadian clock control[25]. To test whether *EID1* was the causative gene underlying the phase QTL, we cloned the *S. pennellii* and the *S. lycopersicum* alleles under the control of each native promoter and transformed the four constructs into the tomato cultivar M82. Transgenic plants carrying the wild species coding sequence (cds) had a significantly more advanced phase with both promoter sequences ($P = 1.68 \times 10^{-14}$, two-way analysis of variance (ANOVA)), demonstrating that variation in the coding region of *EID1* is responsible for the phase QTL (**Fig. 2m**,**n** and **Supplementary Fig. 11**). We used published genome sequences to compare *EID1* across 214 cultivated varieties and 77 wild accessions[26,27] (**Supplementary Table 4**). This comparison revealed a single polymorphism that was present in the coding sequence in all of the cultivated varieties but none of the wild species accessions. This polymorphism (a 3-bp deletion) is located in a region highly conserved among several *Solanaceae* species and probably represents the causal mutation underlying the phase QTL (**Supplementary Fig. 12**).

To assess whether the altered circadian clock of cultivated tomato confers an adaptive advantage and has been selected during tomato breeding or is merely the result of a founder effect and genetic drift, we scanned the genomic region surrounding the *EID1* locus for signatures of selection. We used genome sequences of 166 big-fruited *S. lycopersicum* varieties and 53 *S. pimpinellifolium* accessions[27]. Notably, we found that *EID1* is located in a region that shows evidence of a selective sweep, exhibiting very low genetic diversity in cultivated tomato (**Fig. 3**). These findings indicate that the cultivated allele of *EID1* has been under positive selection during tomato domestication or improvement, suggesting that an altered circadian phase confers an adaptive advantage.

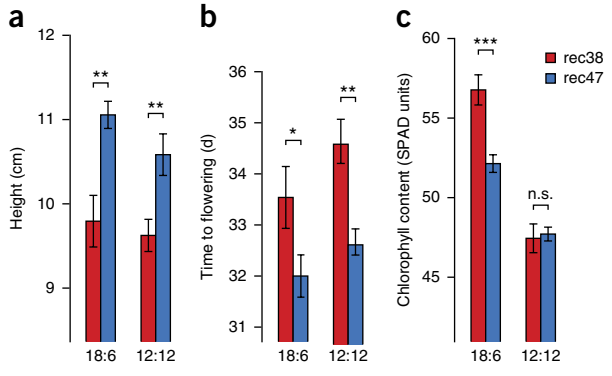

Figure 4 Differences in circadian phase affect overt phenotypes under diurnal conditions. (**a**,**b**) Height 4 weeks after germination (**a**), days to flowering (**b**) and relative chlorophyll content 4 weeks after germination in recombinant lines differing only in the *EID1* locus and thus also in circadian phase (**Supplementary Fig. 10**) grown under long days (18 h light/6 h dark (18:6)) and neutral days (12:12). $n = 16–18$ (**a**,**c**) or 11–15 (**b**), mean ± s.e.m.; $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, *t*-test; n.s., not significant.

Why would there be an advantage in having a decelerated circadian clock? For *Arabidopsis* it has been suggested that a circadian clock with a longer period is advantageous at higher latitudes[2]. Tomato originates from the equatorial region. During domestication, however, it was moved to Mesoamerica and later to Europe, where it faced long summer days[20]. To test whether the slower clock of cultivated tomato leads to better synchronization with longer external light phases, we grew two lines, differing only in their genotype at the *EID1* locus, under 12- and 18-h days. Temporal gene expression waveforms generated with qRT-PCR on different clock genes showed that allelic variation in *EID1* causes a slight phase shift also under diurnal conditions (**Supplementary Fig. 13**). Additionally, plants carrying the cultivated allele of *EID1* were shorter, flowered later and had higher chlorophyll content than those bearing the wild allele ($P < 0.05$, *t*-test; **Fig. 4**). Whereas differences in height and flowering time were independent of day length, differences in chlorophyll content occurred specifically under long days (**Fig. 4**). We hypothesize that the slower clock represents an adaptation to long photoperiods that manifests as higher chlorophyll content, which may in turn enhance overall crop performance. Tomato is damaged when light is present a time when the plant expects darkness[28]. The lower chlorophyll content we observed in plants carrying the wild *EID1* allele under long days could represent the first signs of light-induced injury. Notably, *CAB13*, which encodes light-harvesting chlorophyll a/b–binding protein 13 and has been reported to confer tolerance to light-induced injury[29], was more than fourfold upregulated at the end of a long day in plants carrying the cultivated *EID1* allele (**Supplementary Fig. 13**), providing a possible mechanistic link between changes in temporal gene expression and plant performance. Additionally, the most significantly enriched Gene Ontology (GO) term among genes presenting a delayed diurnal phase in cultivated tomato is 'chlorophyll binding', again indicating that an altered circadian clock may be important for shifting the expression of light-harvesting proteins (**Supplementary Table 5**). In the wild species *S. pimpinellifolium*, chlorophyll content was lower under long days than under short days (**Supplementary Fig. 14c**), suggesting that circadian asynchrony may be even more pronounced when an early phase is combined with a short period. In conclusion, a decelerated clock appears to delay subjective night, which in turn may prevent circadian asynchrony and its associated side effects under long summer days.

Here we provide an example of enhancement of crop performance through modulation of the circadian clock. It will be exciting to see studies of how domestication and breeding have affected the circadian system in other crop plants. Such studies will help to evaluate the potential benefits of manipulating the circadian clock for crop improvement. More comparisons of crop plants with their wild ancestors may also reveal promising target genes for such manipulation and may thereby help to satisfy increasing demand for agricultural production.

URLs. BRASS software, http://millar.bio.ed.ac.uk/Downloads.html; Picard software, http://broadinstitute.github.io/picard/; Sol Genomics network, http://solgenomics.net; NCBI Short Read Archive, ftp://ftp.sra.ebi.ac.uk.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. BioProject: RNA sequencing data have been deposited under accession code PRJNA295848.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank M. Koornneef and A. de Montaigu for helpful discussions and critical reading of the manuscript, U. Tartler and M. Pohe for technical assistance, members of the department, members of the Plan Breeding and Genetics Department at the Max Planck Institute for Plant Breeding Research and S.J. Davis for comments and advice, E.M. Willing for technical advice, and A.W. van Heusden for providing seeds and genotype information for the *S. pimpinellifolium* RIL population. We acknowledge funding from a core grant from the Max Planck Society. N.A.M., A.S. and C.L.W. were funded from a core grant from the Max Planck Society. M.R. was supported by funding from the German Research Foundation under the German-Israeli Project Cooperation program (DFG DIP project number FE552/12-1 awarded to J.M.J.-G.). D.Z. received a European Research Council Advanced grant (YIELD).

AUTHOR CONTRIBUTIONS

J.M.J.-G. and N.A.M. conceived and designed the research. J.M.J.-G., N.A.M., A.S. and M.R. performed and analyzed the RNA sequencing experiment. D.Z., N.R.S., J.N.M., I.O., A.R. and D.W. contributed the *S. pennellii* BIL population. S.H. and T.L. identified signatures of selection. C.L.W. isolated recombinant lines. N.A.M. performed all other experiments and analyzed the data. N.A.M. and J.M.J.-G. wrote the manuscript and all authors revised it.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Hsu, P.Y. & Harmer, S.L. Wheels within wheels: the plant circadian system. *Trends Plant Sci.* **19**, 240–249 (2014).
2. Michael, T.P. *et al.* Enhanced fitness conferred by naturally occurring variation in the circadian clock. *Science* **302**, 1049–1053 (2003).
3. Pittendrigh, C.S. & Takamura, T. Latitudinal clines in the properties of a circadian pacemaker. *J. Biol. Rhythms* **4**, 217–235 (1989).
4. Joshi, D.S. Latitudinal variation in locomotor activity rhythm in adult *Drosophila ananassae*. *Can. J. Zool.* **77**, 865–870 (1999).
5. Edgar, R.S. *et al.* Peroxiredoxins are conserved markers of circadian rhythms. *Nature* **485**, 459–464 (2012).
6. Nagel, D.H. & Kay, S.A. Complexity in the wiring and regulation of plant circadian networks. *Curr. Biol.* **22**, R648–R657 (2012).
7. Beaver, L.M. *et al.* Loss of circadian clock function decreases reproductive fitness in males of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **99**, 2134–2139 (2002).
8. Woelfle, M.A., Ouyang, Y., Phanvijhitsiri, K. & Johnson, C.H. The adaptive value of circadian clocks: an experimental assessment in cyanobacteria. *Curr. Biol.* **14**, 1481–1486 (2004).
9. Green, R.M., Tingay, S., Wang, Z.Y. & Tobin, E.M. Circadian rhythms confer a higher level of fitness to *Arabidopsis* plants. *Plant Physiol.* **129**, 576–584 (2002).
10. Dodd, A.N. *et al.* Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage. *Science* **309**, 630–633 (2005).
11. McClung, C.R. Wheels within wheels: new transcriptional feedback loops in the *Arabidopsis* circadian clock. *F1000Prime Rep.* **6**, 2 (2014).
12. Turner, A., Beales, J., Faure, S., Dunford, R.P. & Laurie, D.A. The pseudo-response regulator Ppd-H1 provides adaptation to photoperiod in barley. *Science* **310**, 1031–1034 (2005).
13. Murphy, R.L. *et al.* Coincident light and clock regulation of pseudoresponse regulator protein 37 (PRR37) controls photoperiodic flowering in sorghum. *Proc. Natl. Acad. Sci. USA* **108**, 16469–16474 (2011).
14. Pin, P.A. *et al.* The role of a pseudo-response regulator gene in life cycle adaptation and domestication of beet. *Curr. Biol.* **22**, 1095–1101 (2012).
15. Zakhrabekova, S. *et al.* Induced mutations in circadian clock regulator Mat-a facilitated short-season adaptation and range extension in cultivated barley. *Proc. Natl. Acad. Sci. USA* **109**, 4326–4331 (2012).
16. Faure, S. *et al.* Mutation at the circadian clock gene *EARLY MATURITY 8* adapts domesticated barley (*Hordeum vulgare*) to short growing seasons. *Proc. Natl. Acad. Sci. USA* **109**, 8328–8333 (2012).
17. Weller, J.L. *et al.* A conserved molecular basis for photoperiod adaptation in two temperate legumes. *Proc. Natl. Acad. Sci. USA* **109**, 21158–21163 (2012).
18. Gao, H. *et al. Days to heading 7*, a major quantitative locus determining photoperiod sensitivity and regional adaptation in rice. *Proc. Natl. Acad. Sci. USA* **111**, 16337–16342 (2014).
19. Campoli, C., Shtaya, M., Davis, S.J. & von Korff, M. Expression conservation within the circadian clock of a monocot: natural variation at barley Ppd-H1 affects circadian expression of flowering time genes, but not clock orthologs. *BMC Plant Biol.* **12**, 97 (2012).
20. Blanca, J. *et al.* Variation revealed by SNP genotyping and morphology provides insight into the origin of the tomato. *PLoS ONE* **7**, e48198 (2012).

21. Filichkin, S.A. *et al.* Global profiling of rice and poplar transcriptomes highlights key conserved circadian-controlled pathways and cis-regulatory modules. *PLoS ONE* **6**, e16907 (2011).
22. Kazmi, R.H. *et al.* Complex genetics controls natural variation among seed quality phenotypes in a recombinant inbred population of an interspecific cross between *Solanum lycopersicum* x *Solanum pimpinellifolium*. *Plant Cell Environ.* **35**, 929–951 (2012).
23. Eshed, Y. & Zamir, D. An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics* **141**, 1147–1162 (1995).
24. Dieterle, M., Zhou, Y.C., Schafer, E., Funk, M. & Kretsch, T. EID1, an F-box protein involved in phytochrome A–specific light signaling. *Genes Dev.* **15**, 939–944 (2001).
25. Somers, D.E., Devlin, P.F. & Kay, S.A. Phytochromes and cryptochromes in the entrainment of the *Arabidopsis* circadian clock. *Science* **282**, 1488–1490 (1998).
26. Aflitos, S. *et al.* Exploring genetic variation in the tomato (*Solanum* section *Lycopersicon*) clade by whole-genome sequencing. *Plant J.* **80**, 136–148 (2014).
27. Lin, T. *et al.* Genomic analyses provide insights into the history of tomato breeding. *Nat. Genet.* **46**, 1220–1226 (2014).
28. Velez-Ramirez, A.I., van Ieperen, W., Vreugdenhil, D. & Millenaar, F.F. Plants under continuous light. *Trends Plant Sci.* **16**, 310–318 (2011).
29. Velez-Ramirez, A.I. *et al.* A single locus confers tolerance to continuous light and allows substantial yield increase in tomato. *Nat. Commun.* **5**, 4549 (2014).

ONLINE METHODS

Plant material. Wild and cultivated tomato accessions used for this study are listed in **Supplementary Table 1** and were obtained from the University of California Davis C.M. Rick tomato genetics resource center (TGRC), the Institute for Conservation and Improvement of Valentian Agrodiversity (COMAV) at the Polytechnic University of Valencia, Spain, or kindly donated by R. Finkers (Wageningen University).

The *S. pennellii* introgression line population was obtained from the University of California Davis C.M. Rick tomato genetics resource center (TGRC) and has been described in detail[23,30]. The *S. pimpinellifolium* recombinant inbred line (RIL) population and respective genotype data were kindly provided by S. van Heusden[22]. The *S. pennellii* BILs used for the fine-mapping of the phase QTL were generated by crossing the wild species *S. pennellii* LA0716 with the cultivar M82 and backcrossing the resulting hybrid to M82 2–3 times. Six or seven rounds of consecutive selfing gave rise to the final BIL population consisting of approximately 700 lines. To genotype these lines, DNA was isolated from leaf tissue using the DNeasy Plant Mini Kit (QIAGEN) and digested with the restriction enzyme NlaIII. Subsequently, reduced representation genomic libraries were prepared using the RESCAN method[31]. These libraries were sequenced at the University of California Berkeley Genomics Sequencing Laboratory on the Illumina HiSeq2000 platform (Illumina, Inc., San Diego, CA) to generate 100-bp single-end reads. Reads were pre-processed, quality filtered and barcode sorted using the Fastx_toolkit and custom perl scripts. The filtered reads were mapped to the tomato reference genome (v2.40) using the BWA parameters: -e 15 -i 10 -k 1 -l 25 -n 0.05 (ref. 32). Samtools (with the '–bq 1' option) was used to retain the reads that mapped uniquely to the reference genome. SNPs were called for parents and individual BILs using VarScan (v2.3.1) (ref. 33). For this, the pileup2snp command was used (parameters: -min-coverage 4 -min-reads 2 -min-avg-qual 20 -minvar- freq 0.9 -p-value 0.05). Finally, SNPs of individual BILs were compared to parental SNPs to identify the *S. pennellii* introgression region and to define the introgression boundaries (**Supplementary Table 2**).

Growth conditions. All tomato seeds were treated with saturated tri-sodium phosphate (Na_3PO_4) for 15 min to kill viruses on the seed coat and to enhance germination efficiency. After three washes with distilled water, seeds were kept in water for 3 d in the dark. On the third day they were sown on standard soil. We positioned genotypes on the basis of a completely randomized design.

For all circadian experiments we used 7 cm × 7 cm pots. For leaf movement analyses seedlings were entrained in an controlled environment chamber (Elbanton) for 2–4 d under cool white fluorescent tubes (~100 µmol m^{-2} s^{-1}) in 12 h light/12 h dark (12:12) and 20:18 °C temperature cycles. On the last day of entrainment a polystyrene ball was attached to the tip of one cotyledon of each seedling using petroleum jelly[34]. At the dark-light transition, we transferred the seedlings to an identical chamber set to constant light and temperature (~100 µmol m^{-2} s^{-1} and 25 °C) and started the image capture. For the quantitative reverse transcription PCR (qRT-PCR) and RNA sequencing experiments, seedlings were entrained for 6–7 d under cool white fluorescent tubes and incandescent bulbs (~100 µmol m^{-2} s^{-1}) in 12:12 or 18:6 (light/dark) and 25:18 °C temperature cycles. For the diurnal RNA-seq time course, we collected leaf tissue every 4 h during 1 day and 1 night. This setting was repeated for the diurnal qRT-PCR experiment, but we collected leaf tissue every 2 h. For the circadian time course, we set the controlled environment chamber to constant conditions (~100 µmol m^{-2} s^{-1} and 25 °C) and collected leaf samples every 4 h for 3 d for the qRT-PCR experiment, and for 2 d for the RNA-seq experiment.

For the phenotyping of the two recombinants (rec38 and rec47), the cultivar Moneymaker and the wild species S. *pimpinellifolium* under different photoperiods we used 9 cm × 9 cm square pots. We grew 18 individuals of each genotype under cool white fluorescent tubes (~100 µmol m^{-2} s^{-1}) with 12:12 or 18:6 light/dark and 24:18 °C temperature cycles. Trays were shuffled within and between chambers at least three times a week to eliminate potential positional and chamber effects.

Image capture and analysis. Leaf movements were studied by analyzing pictures of seedlings taken at an interval of 20 min for 4–5 d using Pentax Optio WG-1 digital cameras triggered by their internal intervalometers. We determined the vertical position of the cotyledon tip over time by measuring the centroid of the polystyrene balls using the software ImageJ (ref. 35). Estimates for the circadian period, phase and relative amplitude error (RAE) per plant were obtained via fast Fourier transform nonlinear least-squares analysis[36] using the biological rhythms analysis software system BRASS. Following common practice, we excluded the first 24 h from the analysis to remove potential noise caused by the transfer from the entrainment chamber to the imaging chamber. The RAE, estimated by BRASS, is a measure of the robustness of a rhythm and can theoretically have values between 0 and 1, where a value of 0 indicates a perfect rhythm and a value of 1 indicates a rhythm that is not statistically significant[36]. Only seedlings with RAE <0.25 were further analyzed, which corresponds to the 95% quantile as determined for the *S. pennellii* IL population (385 individual measurements). Additionally, extreme period or phase outliers were removed.

Quantitative trait locus (QTL) analysis. For QTL mapping we used the R-package 'qtl' (ref. 37). We analyzed 90 *S. pimpinellifolium* RILs that have been genotyped for 1,969 markers. We estimated the genetic map on the basis of recombination frequencies using the Kosambi map function. LOD scores were calculated with a single-QTL model using standard interval mapping with a 1-cM grid. The 5% significance threshold was determined by 10,000 permutations.

Markers for fine-mapping. To fine-map the QTL, cleaved amplified polymorphic sequence (CAPS) markers and insertions and deletions (indel) markers were obtained from published sets[38] or designed manually on the basis of the *S. lycopersicum* and *S. pennellii* reference genome sequences[39]. All primers are listed in **Supplementary Table 6**.

qRT-PCR time course. We used qRT-PCR to quantify the expression of *Solyc10g005080*, *Solyc12g05660*, *Solyc03g115770* and *Solyc07g063600*, the tomato homologs of the *Arabidopsis* clock genes *LHY*, *GI*, *TOC1* and *CAB13*, respectively. We used the standard curve method and the gene encoding AP-2 complex subunit mu (*Solyc08g006960*, CAC) as an internal control to normalize transcript abundance[40]. Primers were designed to span exon-exon junctions to avoid amplification of genomic DNA and are listed in **Supplementary Table 7**. We set up 20-µl reactions in triplicates in Eppendorf twin.tec PCR plates (Eppendorf AG) using IQTM SYBR Green Master Mix (Bio-Rad Laboratories) and 500 ng cDNA, which was synthesized from 1 µg of total RNA with the SuperScriptTM II Reverse Transcriptase (Life Technologies) following the manufacturer's instructions using Oligo(dT)18 primers. PCR was performed in an Eppendorf Realplex cycler (Eppendorf AG) or Bio-Rad CFX384 cycler (Bio-Rad Laboratories GmbH) using a three-step PCR-program consisting of 50 cycles. The threshold cycle (C_t) values were calculated with the Eppendorf realplex2.2 software with the default method (Noiseband), which specifies a threshold of 10 s.d. above the noise of the baseline.

RNA-seq time course. Total RNA from leaf samples was extracted with the RNeasy Plant Mini Kit (QIAGEN). Libraries were prepared according to the Illumina TruSeq RNA protocol and sequenced on the Illumina HiSeq2000 platform (Illumina, Inc., San Diego, CA) at the Genome Center of the Max Planck Institute for Plant Breeding Research. We obtained an average of 29 million single end reads per sample with an average length of 96 bp per read (**Supplementary Table 8**).

Reads were aligned to the *S. lycopersicum* reference sequence version 2.50 using Tophat v2.0.6 (ref. 41) with the following parameters: –max-insertion-length 12–max-deletion-length 12 -g 1–read-gap-length 12–read-edit-dist 20–read-mismatches 12–no-coverage-search–read-realign-edit-dist 0–segment-mismatches 3–splice-mismatches 1–max-intron-length 8100. An average of 92.8% of all reads were uniquely aligned to the reference genome (**Supplementary Table 8**).

The number of reads per transcript was quantified on the basis of the *S. lycopersicum* ITAG annotation v2.4 using a custom R script that employs Bioconductor's Rsamtools, GenomicFeatures and GenomicAlignments packages and the function summarizeOverlaps with parameters mode = "IntersectionNotEmpty", singleEnd = T and ignore.strand = F (refs. 42–44). We divided the samples into a diurnal and a circadian data set (**Supplementary Table 8**).

doi:10.1038/ng.3447

In each experiment we calculated RPKM values for every transcript and discarded transcripts whose values were below 1 in at least one sample. This left us with 14,506 transcripts in the diurnal experiment and 14,668 transcripts in the circadian experiment (41.77% and 42.24% of the 34,725 total transcripts in ITAG2.4, respectively). Raw read counts per transcript were then normalized for each species and experiment separately with the rlog() function in Bioconductor's DESeq2 package[45].

Cycling transcripts were identified by running ARSER v2.0 for each experiment and species separately[46]. ARSER uses three statistical methods to estimate cycling parameters for each gene (yule-walker, mle and burg) and outputs the estimates for the best method on the basis of its Akaike information criterion (AIC) value. In order to obtain the necessary homogeneity to compare cycling parameters between species we modified the ARSER script to return results from the mle method only. Transcripts for which ARSER outputs more than one period or for which the mle method did not converge (marked as 'default' in the output) were discarded. Cycling genes were determined on the basis of the q values output by ARSER (fdr_BH < 0.01).

For the analysis of the circadian experiment, we averaged the two replicates per time-point and species as required by ARSER. ARSER uses a harmonic regression over a range of user-defined periods to find the best possible fit to the expression time course. Because of the large differences in period expected between *S. lycopersicum* and *S. pennellii* in circadian conditions, we used different period ranges for each species. These ranges were defined by running ARSER with sliding windows consisting of 8-h ranges starting at 4 and ending at 44 (step = 1, **Supplementary Fig. 15a**). We chose as final ranges for each species periods that included 99% of the transcripts that were called rhythmic in the window-analysis (fdr_BH < 0.05, **Supplementary Fig. 15b**). For *S. lycopersicum* we used a period window from 22 to 43 and for *S. pennellii* a window of 13 to 26.

In contrast to expression cycles obtained under circadian conditions, diurnal experiments are expected to produce expression cycles with periods of exactly 24 h owing to entrainment of the circadian clock by light and temperature. Therefore, ARSER was run with default parameters (a range of periods between 20 and 28). As the power of ARSER to estimate cycling parameters is increased when analyzing more cycles, we generated an extra day of data by randomly choosing one replicate from every time point and adding 24 h to their collection time (**Supplementary Table 8**). Cycling genes were defined using the modified ARSER script and a threshold of fdr_BH < 0.01.

Gene ontology (GO) enrichment analysis. To test the association of GO terms to genes exhibiting a delayed phase under diurnal conditions in cultivated tomato we used the Bioconductor's GSEABase and GOstats packages[47]. We first prepared a custom GO to gene mapping based on the annotation available on the sol genomics network website. We then performed the Hypergeometric test for the ontology "molecular function" with all genes that exhibited a phase delay of 1–3 h in cultivated tomato compared to its wild relative (n = 755) taking as a universe all transcripts for which a single period was reported by ARSER (n = 7,341).

Cloning of *EID1*. We cloned the *EID1* cDNA of *S. pennellii* and *S. lycopersicum* cv. M82 downstream of both native promoters (2.5 kb upstream sequence from the *EID1* start codon) using the MultiSite Gateway Pro 2.0 Kit (Life Technologies) and the destination vector pGWB1 (ref. 48). For the amplification of the four sequences we used Phusion High-Fidelity DNA polymerase (New England BioLabs GmbH). The primers used are listed in **Supplementary Table 9**. The four resulting constructs were transformed into the tomato cultivar M82 using the *Agrobacterium tumefaciens*–mediated leaf disc transformation method described previously[49]. We validated the presence of the transgene via PCR in the T2 plants that were also used for leaf movement analysis. All T2 plants carrying at least one copy of the transgene (positive in the PCR reaction) were used for determining the effect of the *EID1* promoter and cDNA on circadian rhythms. Progeny of the first two T2 populations per construct exhibiting segregation rates expected for a single insertion event (1:3) were grown in glass jars containing MS medium complemented with kanamycin (50 mg/ml) in an light room set to 16 h light/8 h dark and 25 °C to isolate homozygous lines that were used to validate complementation by the wild-type allele of the *EID1* coding sequence.

Genotyping of *EID1* in 299 re-sequenced tomato accessions. We genotyped *EID1* and the surrounding region using published re-sequencing data of 220 cultivated and 79 wild tomato accessions[26,27].

Short reads[27] were downloaded from SRA and aligned to the *S. lycopersicum* reference genome v2.50 using Bowtie2 version 2-2.0.0-b5 with default parameters[50]. Duplicated reads were removed using Picard version 1.65, and indels were realigned using GATK v2.2-8 (ref. 51). The putatively causal indel is located at position 66,893,249 on chromosome 9. Variants were called for all accessions simultaneously for chromosome 9 from 64 Mb to 69 Mb using GATK v2.2-8 using default settings. The alignment information of the indel position for each accession is presented in **Supplementary Table 4**. Three heterozygous accessions were removed (TS-124, TS-137 and TS-319). Out of the remaining 218 accessions analyzed, two *S. lycopersicum* accessions and two *S. pimpinellifolium* accessions did not exhibit the expected genotype. To clarify the origin of these accessions we constructed phylogenetic trees using 2,000 SNPs upstream and downstream of the position of the putatively causal indel (**Supplementary Fig. 16a**). For this analysis we used only biallelic SNPs that were genotyped in more than 80% of the accessions, and only those accessions for which more than 40% of the 4,000 SNPs could be genotyped by GATK. Neighbor-joining trees were calculated using the R packages ape and adegenet[52]. The trees obtained showed that the *S. lycopersicum* accessions that do not have the characteristic indel in *EID1* (TS-226 and TS-224) cluster in the *S. pimpinellifolium* group, suggesting that, in these accessions, the region was introgressed from the wild species or that the accessions were misclassified (**Supplementary Fig. 16a**). Similarly, *S. pimpinellifolium* accessions TS-222 and TS-432, that contain the indel characteristic of cultivated tomato clustered with the cultivated varieties, pointing to misclassification of these accessions (**Supplementary Fig. 16a**).

Alignment files[27] were downloaded from the Short Read Archive. These alignments are based on the *S. lycopersicum* reference genome v2.40, in which the putatively causal indel in *EID1* is located at position 62,390,868 on chromosome 9. Variants were called in all alignments simultaneously for chromosome 9 from 60 Mb to 65 Mb using GATK v2.2-8 with default parameters. The alignment information for the indel position is presented in **Supplementary Table 4**. We removed from this data set two accessions for which the indel was heterozygous (RF_037 and RF_039) and one *S. lycopersicum* x *S. cheesmaniae* hybrid (RF_054). Of the 52 cultivated varieties remaining, four lacked the deletion characteristic of the cultivated types. Using the methods described above, we constructed neighbor-joining trees with the genotypes for 2,000 SNPs on each side of the indel (**Supplementary Fig. 16b**). For this analysis we used only biallelic SNPs that were genotyped in all 81 accessions. The tree shows that all *S. lycopersicum* accessions lacking the characteristic deletion in *EID1* do not cluster within the *S. lycopersicum* group, suggesting that these accessions belong to the ancestral types or were misclassified.

Detection of selective sweeps. To identify regions of positive selection close to *EID1* we measured the level of genetic diversity (π) at the bottom of chromosome 9 employing genome sequences of 53 *S. pimpinellifolium* and 166 big-fruited *S. lycopersicum* accessions. We used a 100-kb window with a step size of 10 kb to define putative selective sweeps based on the top 5% ratio cutoff, which was 16.47 for this analysis. Further details about selective sweep detection and generation of the genome sequences used for this analysis have been described previously[27].

Phenotyping of the recombinants under different photoperiods. We measured the relative chlorophyll content of the two recombinants using the chlorophyll meter SPAD-502 (ref. 53) according to the manufacturer's instructions (Konica Minolta, Inc., Tokyo). We determined plant height from soil level to apical meristem with a ruler, rounding to the nearest half-centimeter. Chlorophyll content and plant height were measured 28 d after germination. Days to flowering are defined as days from germination until the first open flower.

Phylogenetic analyses of EID1. EID1 protein sequences for the different Solanaceae species were obtained by taking the best BLAST hit from blasting the *S. lycopersicum* EID1 protein against the according protein databases. BLAST was performed online using the BLAST tool on the sol genomics

network website. The sequence alignment was produced with MegAlign, which is part of the DNASTAR Lasergene package (DNASATR, Inc.), using the ClustalW method.

30. Chitwood, D.H. *et al.* A quantitative genetic basis for leaf morphology in a set of precisely defined tomato introgression lines. *Plant Cell* **25**, 2465–2481 (2013).
31. Monson-Miller, J. *et al.* Reference genome-independent assessment of mutation density using restriction enzyme-phased sequencing. *BMC Genomics* **13**, 72 (2012).
32. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
33. Koboldt, D.C. *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22**, 568–576 (2012).
34. Salathia, N., Lynn, J.R., Millar, A.J. & King, G.J. Detection and resolution of genetic loci affecting circadian period in *Brassica oleracea*. *Theor. Appl. Genet.* **114**, 683–692 (2007).
35. Schneider, C.A., Rasband, W.S. & Eliceiri, K.W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
36. Plautz, J.D. *et al.* Quantitative analysis of *Drosophila* period gene transcription in living animals. *J. Biol. Rhythms* **12**, 204–217 (1997).
37. Broman, K.W. & Sen, S. *A guide to QTL mapping with R/qtl*. (Springer, 2009).
38. Koenig, D. *et al.* Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *Proc. Natl. Acad. Sci. USA* **110**, E2655–E2662 (2013).
39. Bolger, A. *et al.* The genome of the stress-tolerant wild tomato species *Solanum pennellii*. *Nat. Genet.* **46**, 1034–1038 (2014).
40. Expósito-Rodríguez, M., Borges, A.A., Borges-Pérez, A. & Pérez, J.A. Selection of internal control genes for quantitative real-time RT-PCR studies during tomato development process. *BMC Plant Biol.* **8**, 131 (2008).
41. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
42. R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2011).
43. Lawrence, M. *et al.* Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9**, e1003118 (2013).
44. Morgan, M., Pagès, H., Obenchain, V. & Haydon, N. Rsamtools: Binary alignment (BAM), FASTA, variant call (BCF) and tabix file import. R package version 1.18.2.
45. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
46. Yang, R. & Su, Z. Analyzing circadian expression data by harmonic regression based on autoregressive spectral estimation. *Bioinformatics* **26**, i168–i174 (2010).
47. Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**, 257–258 (2007).
48. Nakagawa, T. *et al.* Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *J. Biosci. Bioeng.* **104**, 34–41 (2007).
49. McCormick, S. *et al.* Leaf disc transformation of cultivated tomato (*L. esculentum*) using *Agrobacterium tumefaciens*. *Plant Cell Rep.* **5**, 81–84 (1986).
50. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
51. DePristo, M.A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
52. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
53. Markwell, J., Osterman, J.C. & Mitchell, J.L. Calibration of the Minolta SPAD-502 leaf chlorophyll meter. *Photosynth. Res.* **46**, 467–472 (1995).