Variation in *Arabidopsis* flowering time associated with cis-regulatory variation in CONSTANS

Ulises Rosas1, Yu Mei2, Qiguang Xie3, Joshua A. Banta1,4, Royce W. Zhou1, Gabriela Seufferheld2, Silvia Gerard1, Lucy Chou1, Naeha Bhambhra1, Jennifer Deane Parks4, Jonathan M. Flowers1, C. Robertson McClung3, Yoshie Hanzawa2 & Michael D. Purugganan1

The onset of flowering, the change from vegetative to reproductive development, is a major life history transition in flowering plants. Recent work suggests that mutations in cis-regulatory mutations should play critical roles in the evolution of this (as well as other) important adaptive traits, but thus far there has been little evidence that directly links regulatory mutations to evolutionary change at the species level. While several genes have previously been shown to affect natural variation in flowering time in *Arabidopsis thaliana*, most either show protein-coding changes and/or are found at low frequency (<5%). Here we identify and characterize natural variation in the cis-regulatory sequence in the transcription factor CONSTANS that underlies flowering time diversity in *Arabidopsis*. Mutation in this regulatory motif evolved recently and has spread to high frequency in *Arabidopsis* natural accessions, suggesting a role for these cis-regulatory changes in adaptive variation of flowering time.
The onset of flowering, the change from vegetative to reproductive development, is a major life history transition in flowering plants. Flowering phenology is critically tied to the reproductive ecology of flowering plants, and is an important determinant of fitness in a variable environment. Flowering time in Arabidopsis thaliana, a weedy annual plant, has a wide range of variation, and molecular genetic studies have elucidated many of the key pathways that plants utilize to sense seasonal cues, providing unique opportunities for investigators to examine the molecular genetic basis of flowering time variation in an ecological and evolutionary context.

CO is a photoperiod-dependent flowering time locus that encodes a zinc-finger transcription factor downstream of photoreceptor and circadian clock genes. Sequence analysis in 25 accessions identified a 7-bp insertion/deletion (indel) polymorphism in the CO promoter (Fig. 1a). This indel is in a variable number of tandem repeat region consisting of multiple repeats of the 7-bp sequence 5′-CTTTACA-3′ (Fig. 1b). Previous work has demonstrated that the CYCLING DOF FACTOR 1 (CDF1), whose expression exhibits circadian oscillation with the highest peak during the day, represses daytime CO mRNA expression through direct binding to a cis-regulatory element (CTTT) in this 7-bp repeat in the CO promoter.

Here we show that natural variation in the copy number of this CO indel tandem repeat is associated with variation in the developmental transition of flowering time in Arabidopsis accessions. We further show that variation in flowering time can be explained by differences in expression between the alleles. Our results suggest that increased copy number of this tandem repeat evolved recently and quickly spread in Arabidopsis populations, and may play a role in life history variation in natural environments.

**Results**

**Variation and recent origin of a CO promoter polymorphism.** We observe allelic variation in a number of repeat units among the 25 A. thaliana accessions. Seven of the 25 accessions have four complete and one incomplete repeat of this 7-bp motif, resulting in four complete tandem copies of the DOF-binding site (4X). Seventeen accessions have three copies (3X) of the binding site, and one accession (Cvi-0) has two copies (2X) (Fig. 1c). We also sequenced the CO orthologue in A. lyrata, and we find that this sister species has only one complete copy as well as a truncated copy of this motif (TTACACTTTACA) (Fig. 1b). Using a haplotype network (Fig. 1c), we find that all seven alleles containing the 4X copies belong to a single monophyletic group with a single origin (haplotype group A or Hap A), while most of the other A. thaliana accessions have 3X copies (Hap B).

The only molecular difference between the haplotype group A and its ancestral haplotype in group B is the 7-bp insertion that results in an additional CDF1-binding site motif. We found that there are no other indels or single nucleotide polymorphisms that differentiate Hap A from Hap B among our 25 A. thaliana accessions (Fig. 1), suggesting a relatively recent origin of Hap A. Moreover, the alleles within the Hap A group contain no polymorphisms (nucleotide diversity $\pi = 0$) among the seven accessions, in contrast to the Hap B group that shows a moderately low level of nucleotide diversity ($\pi = 0.0025$). Despite its apparent recent origin, the Hap A group is at relatively high frequency; ~41 percent of A. thaliana accessions have Hap A CO alleles, while 59 percent possess Hap B (Supplementary Table 1). Among the 271 accessions analysed, only Cvi-0 carries two repeats of the CDF1-binding site motif. Moreover, previous work using the pairwise haplotype sharing test has shown that the region around CO has an extended...
CO promoter type is associated with flowering time variation.
To examine whether the two most common cis-regulatory CO variants (4X and 3X promoter types) lead to natural variation in flowering time in natural accessions, we undertook a structured candidate gene association analysis. There is a significant association between CO promoter type and bolting time (restricted maximum likelihood (REML) mixed-model analysis of variance (ANOVA), \( P < 0.0023 \)) and rosette leaf number (REML mixed-model ANOVA, \( P < 0.0375 \)) under long days. Previous work indicates that the FRIGIDA (FRI) gene has a key role in repressing the vegetative–reproductive developmental transition, with both wild-type (delayed flowering) and loss-of-function alleles (early flowering) prevalent in natural populations\(^2\). We asked whether a genetic background that represses flowering (that is, FRI) has an effect on the promoter type variation. We found an epistatic interaction of the genetic background defined by the FRI alleles (wild-type or deletion), and CO promoter type for flowering time (REML mixed-model ANOVA, \( P < 0.0027 \)), but not in rosette leaf number (REML mixed-model ANOVA, \( P < 0.1862 \)). In this analysis, there is a significant difference in flowering time in the FRI+: backgrounds (least square means estimate ± s.e. is 52.44 ± 4.82 for 4X versus 46.22 ± 4.83 days for 3X), but no difference in FRI− backgrounds (least square means ± s.e. is 43.35 ± 4.72 for 4X versus 42.98 ± 4.72 days for 3X). This association test suggests that delaying flowering (that is, in the presence of FRI functional alleles) is more likely to show the phenotypic outcomes of natural variation of the CO promoter type. However, these results should be interpreted cautiously, as the definition of FRI functional alleles is not predictive of activity\(^6\), and associations can arise from other epistatic interactions in different A. thaliana accessions\(^9\).

CO promoter type affects flowering time in transgenic Arabidopsis. To validate that the two most common cis-regulatory CO variants (4X and 3X promoter types) lead to functional differences in the ability to regulate flowering time, we created transgenic constructs of the Col-0 CO promoter fused to the Col-0 CO genomic region (Fig. 2a). We developed seven transgenic Arabidopsis lines of pCO (4X):CO FRI-Sf2 co-1, three lines of pCO (4X):CO fri- co-1, five lines of pCO (3X):CO FRI-Sf2 co-1 and three lines of pCO (3X):CO fri- co-1. This transgene partially complements the co-1 mutant allele, which was expected, as the ~1 kb promoter region is unlikely to be the full promoter. Approximately 50 individuals from each of the 18 lines were grown in a fully randomized block design and scored for bolting time and rosette leaf number. The experiment was terminated after 122 days when all genotypes flowered, except for several plants carrying FRI-Sf2 without vernalization (Fig. 2b,c). We analysed the data using a REML mixed model to test for CO, FRI, CO × FRI interaction, block and family effects (see Methods). In the analysis of bolting time, we find that the FRI genotype or CO transgenic promoter type effects are significant (REML mixed-model ANOVA, \( P < 0.0001 \)) for both vernalized and unvernalized plants. The analysis indicates that an increase in the number of the cis-regulatory repeat motif in the CO promoter from 3X to 4X leads to later flowering (Supplementary Table 2). For example, least squares mean estimate for bolting time is 48.46 ± 0.93 days for the CO 3X promoter type and 59.65 ± 0.93 days for the CO 4X promoter type in unvernalized genotypes (Supplementary Table 3). We also tested several CO 2X promoter constructs. Unlike the 3X and 4X transgenic lines, these showed greater variability in flowering times—early flowering in fri− backgrounds similar to that of 3X plants, but later flowering in FRI-Sf2 lines similar to 4X constructs in unvernalized conditions (Supplementary Tables 4 and 5).

Our analysis also indicates an epistatic interaction between CO and the timing of the reproductive transition, in our case driven by variation in FRI in vernalized (REML mixed-model ANOVA, \( P < 0.0001 \)) and unvernalized (REML mixed-model ANOVA, \( P = 0.0001 \)) conditions. Thus, the effect of CO promoter type seems to be dependent on how fast the developmental transition occurs as determined by the FRI background (Fig. 2, Supplementary Table 2). In a fri− background, there is a marginally significant difference in bolting time between CO promoter types in unvernalized condition (post hoc Tukey’s test, \( P = 0.02 \); Fig. 2d), and no significant differences in vernalized conditions (post hoc Tukey’s test, \( P = 0.50 \); Fig. 2e). In contrast, the bolting time of CO 4X plants carrying FRI-Sf2 was significantly later than CO 3X plants carrying FRI-Sf2 both with and without vernalization (post hoc Tukey’s test, \( P < 0.0001 \); Fig. 2e, Supplementary Table 3).

These patterns are consistent if we use rosette leaf number on bolting as a developmental surrogate for flowering time; the mean number of leaves for CO 4X plants are 44.58 ± 0.52, while CO 3X has 36.00 ± 0.53 leaves in unvernalized conditions (REML mixed-model ANOVA, \( P < 0.0001 \)). Moreover, unlike in bolting time, there is a significant difference in rosette leaf number between CO promoter types in fri− backgrounds in both unvernalized and vernalized conditions (post hoc Tukey’s test, \( P < 0.0001 \); Fig. 2f; Supplementary Table 6). The transgenic results are consistent with the structured association results, where 4X CO promoter types showed later flowering in the FRI-active background.

CO promoter type controls differences in gene expression. An obvious hypothesis is that variation in the number of the 7-bp indel motif may result in gene expression differences that lead to phenotypic variability in flowering time. We measured CO expression in 16 Arabidopsis accessions with combinations of CO and FRI allele types. A two-way ANOVA with interaction gave differences in CO expression, with twofold higher expression in the CO 4X allele (8.65 ± 1.7) than accessions with CO 3X (4.22 ± 1.1); these results, however, were marginally non-significant (ANOVA, \( P = 0.0519 \)). FRI allele type or the CO–FRI interaction gave no significant differences (Supplementary Tables 7 and 8).

FRI is an activator of the flowering time repressor FLC\(^10,11\), which is ~2 Mb distal from CO on chromosome V. Thus, we also measured FLC expression, and performed an analysis of variance ( ANCOVA) of allele type with FLC expression with interaction term, and found that FLC expression does not explain the variation in CO expression (Supplementary Table 9). Moreover, linkage disequilibrium (LD) decays in about 10 kb in Arabidopsis\(^12\), and the LD between CO promoter type and FLC haplotype\(^13\) is low \((r^2 = 0.00021)\). This suggests that neither FRI nor FLC gene products are responsible for variation in CO expression, although the CO phenotypic effect might only be seen in a genetic background with delayed flowering (that is, FRI+).

To further test this, we measured relative levels of CO mRNA from CO 3X and 4X promoters in heterozygous plants\(^14,15\) in a fri− background. The CO 4X allele had ~1.5-fold higher expression levels than CO 3X allele (Fig. 3a). To study this expression difference in greater detail, we transformed A. thaliana...
Col-0 with pCO (4X, 3X or 2X)::LUC reporter gene constructs (Fig. 3b). LUC expression under control of CO 4X, 3X and 2X promoters was continuously monitored in planta for 120 h under long-day conditions after a 96-h entrainment under short days (Fig. 3c,d). LUC expression followed a circadian pattern under all three CO promoter types.

The levels of LUC expression is always highest in pCO(4X)::LUC followed by pCO(3X)::LUC (Fig. 3c,d) in both FRI-Sf2 and fri backgrounds throughout the time course of the experiment, while expression of the pCO(2X)::LUC construct is consistently lower. Although overall CO expression increases with greater numbers of the binding site, there are nevertheless clear differences in the pattern of daytime CO expression from the different promoter types. We examined the ratio of the minimum CO expression level during the daytime to its maximum at dawn (Fig. 4). Therefore, based on our data, the simplest explanation is that the greater the number of the motif, the higher the daytime repression of the CO.

The difference between CO 4X and 3X daytime repression is significant in the FRI background (t-test, \(P<0.026\)) but not in fri (t-test, \(P<0.186\)). The expression
Figure 3 | CO promoter type drives differences in gene activity. (a) Expression ratio in a semi-hybrid (Col-0) heterozygote genetic background measured with pyrosequencing. Genomic DNA was used to quantify PCR biases \( (n = 7\) plants). (b) Luciferase gene expression in transgenic reporter lines (see Fig. 2a legend). Luciferase expression was continuously monitored in seedlings in the first 96 h under short days, followed by 120 h in long days. Experimental periods of dark are shown by the light grey boxes. Expression levels of LUC for 2X \( (n = 36\) plants), 3X \( (n = 60\) plants) and 4X \( (n = 60\) plants) CO promoter reporter constructs are shown for (c) FRI + and (d) fri − backgrounds. Error bars indicate s.e.m.

Discussion

We identified cis-regulatory variation in CO, a gene that plays a key role in the flowering time developmental transition. The sequence variation in the promoter region seems to have evolved relatively recently, and is responsible for differences in flowering time and gene expression under laboratory conditions. This CO promoter element may orchestrate a complex series of regulatory interactions. While this motif at the CO promoter binds the DoF transcription factor CDF1, there are also multiple examples in which DoF factors interact with other types of transcription factors, such as basic leucine zipper proteins and MYB proteins, and modulate their binding with DNA. It is also possible that the CO 4X/3X promoter variation affects other nearby binding sites of transcriptional activators of CO, such as the case of FHB transcription factors. Moreover, the protein GIGANTEA binds to this region of the CO promoter to facilitate CDF1 degradation and thus increase CO expression. Together, these suggest a complex interplay between activators and repressors on this cis-regulatory element, which might give subtle phenotypic outcomes in an early-flowering background, but larger phenotypic outcomes in genetic backgrounds with delayed reproductive transition (that is, FRI and FLC-active backgrounds). The molecular mechanisms that connect these regulatory interactions to flowering time phenotypes require further study, but it is clear that natural variation in this cis-regulatory motif has a significant effect on both CO expression and flowering time phenotype.

Although it has been argued that cis-regulatory mutations underlie much of phenotypic evolution, there are as yet relatively few functionally validated examples of identified cis-regulatory changes that are known to have high frequency in natural populations. In the case of FRI, the natural allelic variation results functional variation in the protein-coding region or regulatory sequences. Our work now provides evidence for the role of a high-frequency CO promoter mutation in flowering time variation in A. thaliana, which was not observed on the observed coding region mutations. It remains to be seen whether the role of the CO cis-regulatory variation we observe in controlled laboratory conditions extends to natural environments in the wild. Comprehensive analyses of gene
expression and phenotype in the field, as well as geographical allele distributions across the species range, will be crucial in further understanding the evolutionary and ecological significance of this promoter polymorphism. Nevertheless, our study further supports the importance of modulation of gene expression levels in quantitative phenotypic variation and possibly in evolutionary diversification, and demonstrates that gene regulatory mutations can play important roles in life history diversification in this model wild plant species.

Methods

Plant material and growth conditions. A. thaliana natural accessions used in this study were obtained from the Arabidopsis Biological Resource Center (Ohio, USA). A. lyrata seed was provided by O. Savolainen (Oulu, Finland). The CO-1 mutant line that is introgressed in Col-0 was provided by G. Coupland (Koln, Germany). FRI Sf2 co-1 line carrying the FRI locus that is introgressed from Sf2 into Col-0, and the FRI Sf2 co-1 line carrying the FRI locus from Sf2 and the co-1 mutation were provided by R. Amasino (Madison, USA).

Plants were grown under continuous fluorescent light, long-day (14L:10D) or short-day (10L:14D) conditions at 21 °C on Metromix 360 soil (Sungro Horticulture, USA). For screening transgenic plants, plants were grown under continuous fluorescent light on 1.0% (w/v) agar plates containing MS salts (Sigma), Sodium salts 0.05% (Gibco BRL), sucrose 0.2% and agar 0.8% (Bacto Agar BD).

DNA sequence analysis. Twenty-four accessions were used for sequencing the CO locus, including ~1 kb promoter region, 61 bp 5′ UTR, the coding region, 160 bp 3′ UTR and ~300 bp 3′ flanking region (Genbank accession numbers GQ176989-GQ177102). All sequences were assembled using Phred/Phrap (CodonCode, Dedham, MA) and aligned against the Col-0 CO sequence using BioLign2.9 (Tom Hall, Ibis Therapeutics, Carlsbad, CA). Levels of nucleotide diversity per silent site (θs) and the Watterson’s population mutation parameter θw were estimated using DNASP4.1.2. Haplotype tree was constructed using a maximum parsimony analysis in PAUP*, with a heuristic search and stepwise addition. One insertion/deletion in the promoter region was treated as a single step at 72 °C for 2 min and a final step at 72 °C for 10 min.

Candidate gene association and LD tests. Structured Association Mapping was performed using JMP Genomics version 5.0 (SAS Institute, Cary, NC) using 264 accessions that were phenotyped for flowering time with no vernalization and phenotype in the field, as well as geographical diversity in this model wild plant species.

Expression analyses in Arabidopsis accessions. A set of 16 accessions that have been previously studied for natural variation in vernalization were chosen to measure CO, FLC (At5g10140) and b-actin (At3g18780) expression. Seeds were sown on plates (14–16 seeds per plate) on MS Basal Salt Mixture (Sigma), MES Sodium salts 0.05% (Gibco BRL), sucrose 0.2% and agar 0.8% (Bacto Agar BD). Every Arabidopsis accession was grown in 3–4 plates, and the pool of plants within each plate was considered a biological replicate. Plates were kept in either long dark for 4 days and then transferred to a growth chamber (Percival Scientific). To mimic the growing conditions of the LUC assay, we grew the plants on a short-day condition (10L:14D) for 7 days and then switched them to a long-day condition (16L:8D) for 4 days. Pools of plant tissue were harvested at the start of the light period (9:30 AM) and frozen in liquid nitrogen (kept on ice). Total RNA was extracted from the Plant RNeasy kit (Qiagen), DNA was depleted with DNase I and 2 μg of RNA was used for the first-strand cDNA synthesis with the SuperScript III (Invitrogen), resulting in a single insertion/pooling of 10 T1 plants carrying single-insertion pCO (4X):CO and 10 T1 plants carrying single-insertion pCO (3X):CO were crossed with FRI- Sf2 plants in Col-0, and F1 seeds were selected for kanamycin resistance on MS agar plates. FRI genotypes were tested by PCR for the presence of a characteristic 16 bp deletion in Col-0 fri- in the F2 generation using the primers UJ26 and UJ34. To measure modulation of gene expression levels in quantitative phenotypic variation and possibly in evolutionary diversification, and demonstrates that gene regulatory mutations can play important roles in life history diversification in this model wild plant species.
Expression analysis in a Col-0 heterozygous background. A near isogenic line of Col-0 introgressed with CO 3X from the accession Ler was generated. A back-cross heterozygote for 3X/4X was self-crossed to obtain the segregating genotypes 3X/3X, 3X/4X and 4X/4X. Seeds were disinfected in ethanol–bleach–water 4:1:3 and rinses of sterile water, and plated on MS media (Sigma-Aldrich) supplemented with sucrose 0.1% (Gibco BRL) and agar 1% (Bacto Agar BD), and plates were kept at 22°C in darkness. Plants were rinsed with sterile water, and plated on MS media supplemented with sucrose 0.1% (Gibco BRL) and agar 1% (Bacto Agar BD), and plates were kept at 22°C in darkness. Plants were monitored continuously.

LUC expression analysis. Single-insertion homozygous transgenic plants that carry pCO (4X)::LUC (five lines), pCO (3X)::LUC (five lines) or pCO (2X):CO (three lines) in combination with FRISF2 or fr-1 were obtained as described above. Twelve seedlings per line were sterilized and grown on MS agar (three lines) in combination with CO (Perkin Elmer Life Sciences) for the following 7 days and expression levels of LUC were monitored continuously.

Expression analysis in a Col-0 heterozygous background. A near isogenic line of Col-0 introgressed with CO 3X from the accession Ler was generated. A back-cross heterozygote for 3X/4X was self-crossed to obtain the segregating genotypes 3X/3X, 3X/4X and 4X/4X. Seeds were disinfected in ethanol–bleach–water 4:1:3 and rinses of sterile water, and plated on MS media (Sigma-Aldrich) supplemented with sucrose 0.1% (Gibco BRL) and agar 1% (Bacto Agar BD), and plates were kept at 22°C in darkness. Plants were rinsed with sterile water, and plated on MS media supplemented with sucrose 0.1% (Gibco BRL) and agar 1% (Bacto Agar BD), and plates were kept at 22°C in darkness. Plants were monitored continuously.

CO allele complementation tests. Single-insertion homozygous transgenic co-1 plants that carry pCO (4X)::CO, pCO (3X)::CO or pCO (2X):CO in combination with FRISF2 or fr-1 were obtained as described above. Twelve seedlings per line were sterilized and grown on MS agar plates for 16 days after germination under short day (10L:14D), and transferred into individual wells of 96-well microtiter plates containing substrates for LUC. The plants were subjected to long day (16L:8D) in the TopCount luminometer (Perkin Elmer Life Sciences) for the following 7 days and expression levels of LUC were monitored continuously.

References
1. Bernier, G. The control of floral evocation and morphogenesis. Annu. Rev. Plant Physiol. Plant Mol. Biol. 39, 175–219 (1988).
2. Elzinga, J. A. et al. Time after time: flowering phenotype and biotic interactions. Trends Ecol. Evol. 22, 432–439 (2007).
3. Putterill, J., Robson, F., Lee, K., Simon, R. & Coupland, G. The CONSTANS gene of Arabidopsis promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. Cell 80, 847–857 (1995).
4. Flowers, J. M., Hanszawa, Y., Hall, M. C., Moore, R. C. & Purugganan, M. D. Population genomics of the Arabidopsis thaliana flowering time gene network. Mol. Biol. Evol. 26, 2475–2486 (2009).
5. Imaiuzumi, T., Schultz, T. F., Harmon, F. G., Ho, L. A. & Kay, S. A. FKF1 F-Box protein mediates cyclical repression of a repressor of CONSTANS in Arabidopsis. Science 309, 293–297 (2005).
6. Tajima, F. Evolutionary relationship of DNA sequences in finite populations. Genetics 97, 417–434 (1980).
7. Johanson, U. et al. Molecular analysis of FRI, a major determinant of natural variation in Arabidopsis flowering time. Science 290, 344–347 (2000).
8. Shindo, C. et al. Role of FRI and FLOWERING LOCUS C in determining variation in flowering time of Arabidopsis. Plant Physiol. 138, 1163–1173 (2005).
9. Werner, J., Bonevitz, J., Ulenhaut, N. & Weigel, D. FRI/FRIGIDA-independent variation in flowering time of natural Arabidopsis thaliana accessions. Genetics 170, 1197–1207 (2005).
10. Michaels, S. D. & Amasino, R. M. FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. Plant Cell 11, 1999–2014 (1999).
11. Sheldon, C. C. et al. The FLF MADs box gene: a repressor of flowering in Arabidopsis regulated by vernalization and methylation. Plant Cell 11, 445–458 (1999).
12. Kim, S. et al. Recombination and linkage disequilibrium in Arabidopsis thaliana. Nat. Genet. 39, 1151–1155 (2007).
13. Caicedo, A. L., Stinchcombe, J. R., Olsen, K. M., Schmitt, J. & Schmut, M. D. Epistatic interaction between Arabidopsis FRI and FLC flowering time genes generates a latitudinalcline in a life history trait. Proc. Natl Acad. Sci. USA 101, 15670–15675 (2004).
14. Wittkopp, P. J., Haerum, B. K. & Clark, A. G. Evolutionary changes in cis and trans gene regulation. Nature 430, 85–88 (2004).
15. Rossa, U., Barton, N. H., Copsey, L., Barbier de Reuille, P. & Coen, E. Cryptic variation between species and the basis of hybrid performance. PLoS Biol. 8, e1000429 (2010).
16. Fornera, F. et al. Arabidopsis DOF transcription factors act redundantly to reduce CONSTANS expression and are essential for a photoperiodic flowering response. Dev. Cell 17, 163–172 (2009).
17. Johanson, U. et al. Natural variation in flowering time of natural Arabidopsis thaliana accessions. Genetics 170, 1197–1207 (2005).
18.Bowler, S. et al. GIGANTEA: a circadian clock-controlled gene that regulates photoperiodic flowering in Arabidopsis and encodes a protein with several possible membrane-spanning domains. EMBO J. 18, 4679–4688 (1999).
19. Park, D. H. et al. Control of circadian rhythms and photoperiodic flowering by the Arabidopsis GIGANTEA gene. Science 285, 1579–1582 (1999).
20. Sawa, M., Nusinow, D. A., Kay, S. A. & Imaiuzumi, T. FKF1 and GIGANTEA complex formation is required for day-length measurement in Arabidopsis. Science 318, 261–265 (2007).
21. Song, Y. H., Smith, R. W., To, B. J., Millar, A. J. & Imaizumi, T. FKF1 conveys timing information for CONSTANS stabilization in photoperiodic flowering. Science 336, 1045–1049 (2012).
22. Hoecker, H. E. & Couye, J. A. The locus of evolution: evo devo and the genetics of adaptation. Evolution 61, 995–1016 (2007).
23. Pridhomme, B., Gompel, N. & Carroll, S. B. Emerging principles of regulatory evolution. Proc. Natl Acad. Sci. USA 104, 8605–8612 (2007).
24. Richards, C. L., Rosas, B. A., Stankova, L., Bhambra, N. & Purugganan, M. D. Genome-wide patterns of Arabidopsis gene expression in nature. PloS Genet. 8, e1002662 (2012).
25. Wilczek, A. M. et al. Effects of genetic perturbation on seasonal life history plasticity. Science 323, 930–934 (2009).
26. Banta, J. A. et al. Climate envelope modelling reveals intraspecific relationships among flowering phenology, niche breadth and potential range size in Arabidopsis thaliana. Ecol. Lett. 15, 769–777 (2012).
27. Watterson, G. A. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* 7, 256–276 (1975).
28. Rozas, J., Sánchez-DelBarrio, J. C., Meseguer, X., & Rozas, R. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19, 2496–2497 (2003).
29. Rozas, J., Sánchez-DelBarrio, J. C., Messeguer, X. & Rozas, R. *DnaSP, DNA polymorphism analyses by the coalescent and other methods.* Bioinformatics 19, 2496–2497 (2003).
30. Yu, J. et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38, 203–208 (2006).
31. Olsen, K. M. et al. Linkage disequilibrium mapping of *Arabidopsis CRY2* flowering time alleles. *Genetics* 167, 1361–1369 (2004).
32. Ehrenreich, I. M. et al. Candidate gene association mapping of *Arabidopsis* flowering time. *Genetics* 183, 325–335 (2009).
33. Desrousseaux, D., Sandron, F., Siberchicot, A., Cierco-Ayrolles, C. & Mangin, B. *LDcorSV,* Linkage disequilibrium corrected by the structure and the relatedness. Version 1.3.1, http://cran.r-project.org/web/packages/LDcorSV/index.html (2013).
34. Team, R. C. *R: a Language and Environment for Statistical Computing* (R Foundation, 2013).
35. Morris, K. et al. DAY NEUTRAL FLOWERING represses CONSTANS to prevent *Arabidopsis* flowering early in short days. *Plant Cell* 22, 1118–1128 (2010).
36. Mockler, T. C. et al. Regulation of flowering time in *Arabidopsis* by K homology domain proteins. *Proc. Natl Acad. Sci. USA* 101, 12759–12764 (2004).
37. Salome, P. A., Xie, Q. & McClung, C. R. Circadian timekeeping during early *Arabidopsis* development. *Plant Physiol.* 147, 1110–1125 (2008).

**Acknowledgements**

We thank Guillermo Perez-Perez for providing access to the pyrosequencer. We thank Zunping Luo for assistance on genotyping, and Joanna F. Shaw for flowering time data gathering. This work was supported by a Human Frontier Science Program Postdoctoral Fellowship to U.R., and the NSF grants MCB-0929338 to M.D.P. and IOS-0923752 and IOS-1025965 to C.R.M.

**Author contributions**

Y.H. and M.D.P. conceived the research; Y.H. and J.M.F. performed the sequence diversity analyses; J.A.B. and J.D.P. performed the association analysis; Y.H., U.R., Y.M., R.W.Z., J.A.B., S.G., L.C., N.B. and G.S. generated the transgenic lines; Q.X. and C.R.M. performed the LUC activity assays; U.R. and R.W.Z. performed the qPCR and pyrosequencing assays; U.R. and R.W.Z. obtained the phenotypic complementation data; Y.H. and M.D.P. directed the research; and U.R., Y.H. and M.D.P. wrote the manuscript.

**Additional information**

**Accession codes:** CO sequences of 24 *Arabidopsis* accessions: GQ176989-GQ177012.

**Supplementary Information** accompanies this paper at http://www.nature.com/nature communications

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at http://npg.nature.com/reprintsandpermissions/

**How to cite this article:** Rosas, U. et al. Variation in *Arabidopsis* flowering time associated with cis-regulatory variation in CONSTANS. *Nat. Commun.* 5:3651 doi: 10.1038/ncomms4651 (2014).