Rajapaksha Patirange Dilan Sarange (Orcid ID: 0000-0001-6308-1838)

Emrani Nazgol (Orcid ID: 0000-0001-5673-3957)

Haplotype variations of major flowering time genes in quinoa unveil their role in the adaptation to different environmental conditions

Dilan S.R. Patiranage[1], Edward Asare[1], Nathaly Maldonado-Taipe[1], Elodie Rey[2], Nazgol Emrani[1], Mark Tester[2] and Christian Jung[1,*]

[1]Plant Breeding Institute, Christian-Albrechts-University of Kiel, Olshausenstr. 40, D-24098 Kiel, Germany

[2]King Abdullah University of Science and Technology (KAUST), Biological and Environmental Sciences & Engineering Division (BESE), Thuwal, 23955-6900, Saudi Arabia.

* Corresponding authors' email address: c.jung@plantbreeding.uni-kiel.de

Correspondence to: Prof. Dr. Christian Jung

Plant Breeding Institute

Christian-Albrechts-University of Kiel

Am Botanischen Garten 1-9

D-24118 Kiel

Germany

Tel.: +49 431 880-7364

Fax: +49 431 880-2566

Running Head: Haplotype variation of major FTi genes in quinoa

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/pce.14071

Abstract

Response to photoperiod is of major importance in crop production. It defines the adaptation of plants to local environments. Quinoa is a short day plant which had been domesticated in the Andeans regions. We wanted to understand the adaptation to long-day conditions by studying orthologs of two major flowering time regulators of Arabidopsis, *FLOWERING LOCUS T* (*FT*) and *CONSTANS* (*CO*) in quinoa accessions with contrasting photoperiod response. By searching the quinoa reference genome sequence, we identified 24 *FT* and six *CO* homologs. *CqFT* genes displayed remarkably different expression patterns between long-day and short-day conditions, whereas the influence of the photoperiod on *CqCOL* expressions was moderate. Cultivation of 276 quinoa accessions under short-day and long-day conditions revealed great differences in photoperiod sensitivity. After sequencing their genomes, we identified large sequence variations in 12 flowering time genes. We found non-random distribution of haplotypes across accessions from different geographical origins, highlighting the role of *CqFT* and *CqCOL* genes in the adaptation to different day-length conditions. We identified five haplotypes causing early flowering under long days. This study provides assets for quinoa breeding because superior haplotypes can be assembled in a predictive breeding approach to produce well-adapted early flowering lines under long-day photoperiods.

Keywords

Chenopodium quinoa, floral transition, expression analysis, photoperiod, day-length, plant breeding

Rajapaksha Patirange Dilan Sarange (Orcid ID: 0000-0001-6308-1838)

Emrani Nazgol (Orcid ID: 0000-0001-5673-3957)

Introduction

Quinoa (*Chenopodium quinoa* Willd.) is a traditional Andean crop that was domesticated around 5000-7000 years ago. The production and consumption of quinoa have rapidly expanded in recent years due to its broad adaptability and nutritional value (Murphy & Matanguihan, 2015). The crop is widely adaptable to different agroecological conditions due to its ecological plasticity and diversity (González, Eisa, Hussin, & Prado, 2015). Quinoa germplasm possesses a broad genetic diversity, which can benefit its adaptation and cultivation in Northern Europe. Understanding the underlying genetics of flowering is essential for the fast improvement of the crop.

The onset of flowering is preceded by a major transition of the shoot apical meristem, resulting in floral organ primordia formation. This transition results from signal cascades involving transcription factors and epigenetic modification, which is well understood in the model plant *Arabidopsis thaliana*. *FLOWERING LOCUS T* (*FT*) is the major floral activator integrating signals from upstream regulators, indirectly responding to photoperiod, temperature, plant age, and hormones (Blümel, Dally, & Jung, 2015). *FT* is expressed in leaves, and the protein encoded is transferred to the shoot apical meristem to initiate flowering in combination with other genes. In Arabidopsis, under long-day (LD) conditions *FT* expression peaks in the morning and around dusk (Song et al., 2018). Photoperiod is an essential factor that affects flowering in plants, and most of the genes involved are highly conserved among crops (Eshed & Lippman, 2019). *CONSTANS* (*CO*) is a key factor in photoperiodic sensing, and it directly governs the expression of *FT* under LD conditions. *CO* is controlled by circadian clock genes, and expression peaks at the end of the day. *CO* encodes for two adjacent B-BOX (BBX) and CCT (CO, CO-like, TOC) domains. The CO protein is stable during the light, and it is degraded during the dark. *CO* is not involved in flowering under short-day (SD) (Andres & Coupland, 2012). A recent study showed that flowering was promoted under SD conditions in *co* mutants, indicating its role in delaying flowering under SDs (Luccioni et al., 2019).

Quinoa belongs to the plant family Amaranthaceae together with sugar beet (*Beta vulgaris*), spinach, amaranth, *Chenopodium album,* and *C. rubrum*. Sugar beet is a LD species but needs vernalization followed by LDs for flowering. It has two *FT* homologs, *BvFT1* and *BvFT2,* that act antagonistically.

Before vernalization, *BvFT1* is active and suppresses floral transition. After exposure to a long period of cold temperatures, *BvFT1* expression decreases, and *BvFT2* becomes transcriptionally active to induce flowering (Pierre A Pin et al., 2010). Sugar beet is lacking a true *CO* ortholog, which encodes both CCT and BBX domains. Two *CONSTANS-LIKE* (*COL*) genes have been investigated for their involvement in floral induction in sugar beet. *BvCOL1* complemented flowering in an Arabidopsis *co* mutant, but it did not show the typical *CO* expression pattern (Chia, Müller, Jung, & Mutasa-Göttgens, 2008), nor was there any genetic evidence for its function as a flowering time gene. However, two other genes have been proven to be upstream regulators of *BvFT1* and *BvFT2. BOLTING TIME CONTROL 1* (*BTC1*) is a *PSEUDO RESPONSE REGULATOR* (*PRR*) gene with a CCT domain but lacking the BBX domains (Pierre A. Pin et al., 2012), whereas *BvBBX19* carries the BBX but is lacking the CCT domain (Dally, Xiao, Holtgräwe, & Jung, 2014). Both proteins form a dimer, thus acquiring a CO function to regulate flowering (Dally, Eckel, Batschauer, Höft, & Jung, 2018).

Flowering time regulation has also been studied in the SD species *C. rubrum*. This wild relative of beet and quinoa has two *FT* homologs; *CrFTL1* and *CrFTL2* (Cháb, Kolář, Olson, & Štorchová, 2008), but only *CrFTL1* was found to be essential to initiate flowering (Drabešová, Cháb, Kolař, Haškovcová, & Štorchová, 2014). In the same study, *CrCOL1* and *CrCOL2* genes were investigated. Their expression patterns were similar to beet *COL* genes, but they did not complement Arabidopsis *co* mutants. A recent transcriptome study in two other close relatives, *C. ficifolium* and *C*. *suecicum,* showed that the expression of *FT*-like (*CfFTL1 and CsFTL1*) genes is correlated with flowering since they showed characteristic diurnal *FT* expression patterns. However, under LDs, their expression was negligible (Štorchová et al., 2019), indicating *C. ficifolium* flowers under LD without upregulation of *FT*-like genes.

Quinoa is a short-day species owing to its origin. Nevertheless, it has been shown that quinoa can flower under varying photoperiods depending on the genotype. Therefore, quinoa has been regarded as a facultative SD crop (H. Bertero, King, & Hall, 1999). The day-length response varies vastly between accessions from different geographical regions (H. D. Bertero, 2003). In a study with quinoa germplasm from Bolivia, the latest accession flowered 85 days after the earliest accession (Rojas, 2003). Strong photoperiod sensitivity is the major hurdle for quinoa cultivation under LD conditions in temperate regions such as Northern Europe and Southern Chile. So far, flowering time research in

quinoa has been in its infancy. Christiansen, Jacobsen, and Jørgensen (2010) investigated the phenological development of two quinoa accessions with contrasting days to maturity under different photoperiod regimes (SD, LD, and reciprocal transfer). They speculated that day-length sensitivity in quinoa is governed by a few genes only, but the mode of inheritance and position of genes in the genome could not be identified. Recently, a reference genome sequence has been published, which laid the foundation to investigate flowering time genes on a genome-wide scale. Five *FT* homologs were identified as putative floral inducers. Sequence similarity to *BvFT* suggested their putative function as floral activators/repressors (Jarvis et al., 2017). A study was performed to uncover the evolutionary history of *FT*-like genes in quinoa and identified 14 *FTL* genes belonging to seven *FTL* families. This study proposed a unified gene nomenclature of *FT* genes in the Amaranthaceae family species to avoid confusion when comparisons are made between species. Interestingly members of the *FTL3* gene family specific for Amaranthaceae species, were also identified in quinoa, and their potential role in the adaptation to different environments was hypothesized (Štorchová, 2020). Another genome-wide survey uncovered 611 sequences with some homology to Arabidopsis flowering time genes and 459 quinoa-specific flowering genes. This study was based on a bioinformatics approach that employed BLAST search, orthologue, and collinearity identification (Golicz, 2019). However, genetic or functional evidence surmounting their role as flowering time regulators is still lacking.

This study aimed to understand the molecular reasons of day-length sensitivity in quinoa. The general hypothesis was that *FT* and *COL* homologs play a similar role in quinoa as in sugar beet. However, we expected divergent functions of at least one *FT* ortholog because beet has a strong requirement for vernalization. Selected quinoa genotypes with contrasting day-length requirements were grown under controlled conditions, and the transcriptional activities of six *FT* and six *COL* homologs were analyzed at different stages of development. We found strong day-length responses of *CqFT* genes, whereas the influence of the photoperiod on *CqCOL* expressions was moderate. We further analyzed haplotypes of 12 genes based on SNPs identified from a diversity panel of 303 quinoa accessions. For some genes, haplotypes were non randomly distributed among accession with different geographical origins. This indicates the function of these genes as major factors for the adaptation to varying photoperiods. We identified accessions carrying specific haplotypes, which do not respond to photoperiod, whereas others respond to photoperiod. We then compared the geographical distribution of haplotypes. We found that individual haplotypes are more frequent in the southern latitude

compared to other haplotypes. Our results shed new light on the regulation of photoperiod response and vernalization requirement in related species with contrasting life cycle regimes. The data will be important for quinoa breeding because haplotypes can be assembled in a predictive breeding approach to produce early flowering lines well adapted to northern latitudes.

Materials and methods

Plant materials and growth conditions

We worked with a diversity panel of 303 quinoa accessions representing different geographical regions of the world and different breeding intensities, from landraces without breeding history to inbred lines (Patiranage et al., 2020). Seeds from 276 accessions (Table S1) were sown in 35x multi-tray pots, and 10 plants/accession were grown in the climate chamber under SD (8h light and 16h dark) and LD (16h light and 8h dark cycles) photoperiodic conditions at 20°C. Halogen lamps with an intensity of 300 µmol m^{-2} s^{-1} were used in both chambers. We randomized the multi-tray pots weekly to minimize spatial effects. We recorded flowering when the first flower opened, and days to flowering (DTF) were measured from the sowing date (experiment 1, E1).

To analyze gene expression (experiment 2, E2), we selected two early- (Titicaca, seed code 171230 and PI-614886, seed code 170867) and two late- (PI-587173, seed code 171605 and CHEN-109, seed code 170876) flowering accessions from different geographical regions. Seeds were sown in 9 cm pots, and 15 plants/accession were grown in climate chambers under SD and LD. DTF and days to bolting (DTB) were recorded. We collected samples for RNA isolations and expression analysis for two different experiments: diurnal expression and temporal expression. For each sampling stage/point, three biological replicates were used, and the sampling was not destructive. We performed diurnal sampling at growth stage 7 (bolting; pyramid shape of the inflorescence) (Jacobsen and Stølen (1993). To measure temporal expression, we took leaf samples from plants at nine different time points: two weeks after sowing (vegetative), two weeks before bolting (2WBB), one week before bolting (1WBB), bolting (B), one week after bolting (1WAB), one week before flowering (1WBF), flowering (F), one week after flowering (1WAF) and two weeks after flowering (2WAF). All samples were immediately frozen in liquid nitrogen.

Rajapaksha Patirange Dilan Sarange (Orcid ID: 0000-0001-6308-1838)

Emrani Nazgol (Orcid ID: 0000-0001-5673-3957)

In-silico identification of quinoa *FT*-like genes and *CONSTANS* like genes

We used the BvFT2 (Pierre A Pin et al., 2010) protein sequence as a query for a BLAST search against an improved version of the quinoa reference genome (Jarvis et al., 2017). This reference genome was assembled using a day-neutral coastal ecotype (PI-614886). The improved genome version is stored in the CoGe database (https://genomevolution.org/coge/) under the genome id: CoGe id60716. Moreover, we checked the Conserved Domain Database (CDD) in NCBI to identify conserved domains present in the proteins. For *CONSTANS*-like gene identification, we used the AtCO protein sequence as a query, and then we followed the same procedure as explained above.

Phylogenetic tree construction

We used the protein sequence obtained from candidate genes for multiple sequence alignment (Pierre A Pin et al., 2010). As outliers for phylogenetic tree construction of *FT*-like genes, we used AtFT, CrFTL1, CrFTL2, BvFT1, BvFT2, Hd3a, and RFT protein sequences. For CONSTANS-like protein alignment, we used AtCOL, BvCOL, CrCOL, and HD1 protein sequences. Then, we generated a phylogenetic tree using the CLC-main workbench 8.1.3 based on the Neighbor-Joining algorithm with 1000 bootstraps. We obtained a protein pairwise identity percentage matrix from the Clustal Omega (Madeira et al., 2019).

RNA isolation and gene expression analysis

Total RNA was isolated using the NucleoSpin RNA isolation kit (Macherey-Nagel) and treated with DNaseI (Life Technologies GmbH, ThermoFisher Scientific Darmstadt, Germany). One microgram (1µg) of total RNA was taken for cDNA synthesis using the First Strand cDNA Synthesis Kit (Life Technologies GmbH, ThermoFisher Scientific Darmstadt, Germany). We diluted the cDNA by 20-fold (1:20) and used 2µl of diluted cDNA as a template for RT-qPCR with three biological replicates per sampling point and three technical replicates. We performed RT-qPCR with SYBR qPCR Supermix w/ROX (Life Technologies GmbH, ThermoFisher Scientific Darmstadt, Germany), using a CFX96 Real-Time System (Bio-Rad Laboratories GmbH, Germany). We used paralog-specific primers to conduct the RT-qPCR (Table S4). We analyzed amplification curves using Bio-Rad CFX Manager 3.1

and calculated the relative expression of each gene using the ΔCt method. As reference genes, we used the geometric mean of *CqIDH-A* and *CqPTB* genes. These genes turned out to be best suited as reference genes because they were stably expressed during different growth stages and different time points of the day (Maldonado-Taipe, Sarange, Schmöckel, Jung, & Emrani, 2020).

DNA sequencing and haplotype analysis

We used whole-genome re-sequences from 303 quinoa accessions from the diversity panel described by Patiranage et al. (2020). We mapped Illumina short-read sequences against an improved version of the quinoa reference genome (Jarvis et al., 2017). SNP variants were identified on a population scale using the "—emitRefConfidence" function of the HaplotypeCaller algorithm on each sample, followed by joint genotyping on all samples using GenotypeGVCFs from the Genome Analysis Toolkit (McKenna et al., 2010). The resulting VCF file was further filtered based on minor allele frequency (MAF) ≥ 0.05, minimum mean depth (min-meanDP) ≥ 5.0, and maximum-missing (max-missing) ≤ 0.5 using VCFtools (Danecek et al., 2011) and obtained a high confident SNP data set (HC-SNPs). Then, for each gene, we created a bed file containing chromosome name, start position, and end position. We added -/+ 20kb regions to the start position and end positions to obtain the 20kb flanking regions of genes. For each gene, we created a VCF file using VCFtools (Danecek et al., 2011). Then, we converted VCF files into Flapjack format. After that, we used the Flapjack software (Milne et al., 2010) to sort accessions according to similarity. Subsequently, we manually checked the sequence of each accession to assemble them into haplotypes. We excluded accessions with missing and heterozygous SNPs. For haplotype calling, we first analyzed the transcribed regions. If SNPs were absent in the transcribed region, we further analyzed the regions 5kb upstream of the transcription start. We inferred haplotype networks using haplotNet function from the R package "pegas" (Paradis, 2010).

Statistical analysis

We compared DTF within haplotypes using "Welch's" two-sample t-test ($P<0.05$). Based on the *P*-values obtained from the t-test, we classified haplotypes into three groups (1) insensitive ($P>0.05$), (2) moderately sensitive ($0.001<P<0.05$), and (3) highly sensitive ($P<0.001$). We performed an analysis of variance to compare the DTF of haplotype groups within a photoperiod treatment (SD or LD). The population criteria (Highland and Lowland) obtained from Patiranage et al. (2020) were included in the

model as a cofactor to control confounding effects from the origin accessions. Haplotypes were then grouped based on a post hoc comparison with the Tukey test at $P<0.05$ level. All statistical analyses were performed in R software version 3.6.1 (RCore 2019). We used the recently proposed estimation graphics (J. Ho, Tumkaya, Aryal, Choi, & Claridge-Chang, 2019) in combination with Student's t-test to compare SD and LD flowering time data for each accession. We classified an accession as 'short day' or 'long day' if flowering dates were significantly different between both light regimes. 'day neutral' accessions did not show any significant difference between LDs and SDs.

Results

Flowering time variation under different photoperiodic conditions

Among 303 accessions from our quinoa diversity panel, 276 accessions were selected for the growth experiment (E1) under LD and SD conditions. We observed a broader variation for DTF under LDs (CV=19.66%) compared to SDs (CV=12.4%). Under LD conditions, DTF ranged from 38 to 123 days (mean: 60.02 ± 10.4), whereas under SDs, it ranged from 35 to 72 days (mean: 50 ± 5.2 days). We identified all three types of photoperiodic response ('short day', 'long day', and 'day neutral') among the 276 accessions. We identified 67.39% of the accessions to be 'short day' accessions, which flowered significantly earlier under SD conditions. We also found that 16.67% and 7.6% of the accessions are 'day neutral' and 'long day' accessions, respectively. Then, we compared DTF under long- and short- days with the geographical origin of the accessions. Notably, we found that southern Chile accessions are photoperiod insensitive, which indicates that day-length sensitivity was lost when cultivation was extended towards lower latitudes (Table 1 and Figure S1).

Expression analysis of putative *FT* and *COL* orthologs

We reasoned that putative *COL* and *FT* orthologs function in quinoa in the same way as in their close relatives. Searching the quinoa genome for *FT* homologs, we identified 24 *PEBP* (*phosphatidyl ethanolamine-binding proteins*) family genes, five of which had been identified before as putative *FT* orthologs (Jarvis et al., 2017). Protein identity between 24 *PEBP* genes and BvFT2 ranged from 40% to 92% (Table S2 and Figure S2). We generated phylogenetic trees to analyze the phylogenetic relationship of the candidate genes. We identified 5, 2, 4, and 13 genes belonging to *TFL1*- like, *BFT*-like, *MFT*-like, and *FT*-like clades. Like in *C. rubrum*, the *FT*-like clade could be split into three

subclades, which were belonging to *FTL1*, *FTL2,* and *FTL3.* Seventeen *PEBP* genes were encoding the complete protein with the PEBP domain (Figure S5 A). We observed discrepancies between reference annotation genome version 1 and improved version 2, resulting from differences between genome assemblies. Unfortunately, genes have been given different names by Jarvis et al. (2017) and Štorchová (2020). Therefore, we provide a side-by-side comparison of all genes from both publications in Table S2. Furthermore, we performed an *in-silico* characterization of previously identified *cis*-regulatory elements (CORE1, CORE2, P1, and P2) in six *CqFT* genes, which were later used in the expression analysis (Adrian et al., 2010; Song et al.,2012; Tiwari et al., 2010). All these elements were found in all genes within 6kb upstream of the ATG. However, their positions relative to the translation start site was varying between genes (Figure S3). We identified six *COL* genes closely resembling *AtCOL2*, *AtCOL4*, and *AtCOL5*. Quinoa COL proteins showed 40% to 89% protein identity to BvCOL1 and 37% to 55% identity to CO (Table S3 and Figure S4). All genes encode two B-boxes and CCT domains. A phylogenetic tree was constructed with sequences from CO and COL genes of quinoa and four other species. *CqCOL* genes were distributed across three clades, demonstrating their resemblance to beet and *Chenopodium* spp. *COL* genes (Figure S5 B).

Then, we chose six *FTL* genes and six *COL* genes for expression analysis, which showed the highest protein similarity to BvFT2 and BvCOL, respectively (Table S2 and Table S3). We questioned whether their transcriptional activities would respond to different photoperiods and growth stages. We also expected different expressions between ‘early’ and ‘late’ genotypes. Therefore, we grew four quinoa accessions with strikingly different flowering times under LDs and SDs (E2). We observed more pronounced vegetative growth under LDs than under SDs. Moreover, we observed that red color pigmentation in the stem is more prominent under LDs than under SDs (Figure S6).

We also phenotyped those four accessions for DTB and DTF (E2). The ‘early’ accessions Titicaca (seed code: 171230) and PI-614886 (seed code: 170867) started bolting 23 and 27 DAS under SDs and LDs, respectively, indicating that both accessions do not respond to photoperiod. In contrast, the ‘late’ accessions were bolting differently. The accession PI-587173 (seed code:171605) bolted nine days earlier under SDs, whereas accession CHEN-109 (seed code:170876) bolted five days earlier under LDs. As expected, accessions Titicaca and PI-614886 flowered earlier than PI-587173 and CHEN-109 under both conditions. Accessions Titicaca and CHEN-109 did not respond to different

day-length, whereas PI-614886 and PI-587173 flowered earlier under SDs (3 and 19 days, respectively) (Figure S7). We decided to use the 'day-neutral' accession Titicaca (early) and the 'short-day' accession PI-587173 (late) for expression profiling (E2), due to their contrasting flowering time under different photoperiodic regimes, and due to their contrasting responses to photoperiod.

In Arabidopsis and sugar beet, *FT* and *COL* genes are diurnally expressed. We reasoned that their putative orthologs from quinoa are also diurnally regulated under different photoperiods. We collected leaves from two accessions (E2) at the bolting stage. We observed significantly different expression patterns of all six *CqFT-like* genes investigated between LDs and SDs (Figure 1) in both accessions. The transcriptional activities of all *CqFT* genes were negligible under LDs except for *CqFT1B-2,* which showed stable expression under LDs (accession PI-587173). All *CqFT* genes except *CqFT1B-1* were diurnally regulated under SDs, although at varying levels of expression. *CqFT2A*, *CqFT1A*, and *CqRFT* showed similar expression patterns under SDs; those genes showed the highest expression in the morning and gradually reduced expression during the day. The lowest expression was observed at 16ZT. In general, expression levels of *CqFT* genes in the early accession Titicaca were relatively low compared to the late accession under both LDs and SDs. Diurnal expression of *CqFT2B* under SDs in accession PI-587173 showed late evening peak similar to floral activators (Andres & Coupland, 2012; Hayama, Agashe, Luley, King, & Coupland, 2007; Pierre A Pin et al., 2010), but the expression level was lower than *CqFT2A.*

Then, we analyzed the expression pattern of six *CqCOL* genes. *CqCOL4A-2* (LD and SD) and *CqCOL2B* (SD) showed very low expression, while the other four genes showed diurnal expression under both conditions (Figure 1). Generally, the relative expression of *CqCOL* genes was higher under SDs. The expression levels in the early accession (Titicaca) were lower than the late accession (PI-587173). In Titicaca, all *CqCOL* genes showed a peak after 4h of light under LDs. The *CqCOL5* gene showed different expression patterns between SDs and LDs, but expression patterns were not different between accessions. Taken together, SD expression patterns of *CqCOL* genes were similar to what has been reported for sugar beet *BvCOL1* (Chia et al., 2008) and Arabidopsis *COL1* and *COL2* (Ledger, Strayer, Ashton, Kay, & Putterill, 2001).

As a next step, we analyzed the temporal expression of the candidate genes taking leaf samples at ZT4 over 11 weeks, starting from the second week after sowing (E2) (supplementary Table S5). We

found strikingly different expression profiles between genes in response to day-length. Under LDs, we could not detect any *CqFT2A* transcripts in the late accession; however, in the early accession, both paralogs were expressed at the flowering stage and one week after flowering (Figure 2). In contrast, under SDs, the expression of *CqFT2B* was higher at the vegetative stage and decreased until bolting. *CqFT2A* expression could be detected only at the flowering stage and at later developmental stages under SDs in the late accession. It was expressed one week before bolting in the early accession. In the early accession, both *CqFT2* genes showed similar expression patterns under SDs. Also, *CqFT1A* expression was similar to *CqFT2A* expression under SDs. Moreover, we did not detect *CqFT1B-1* expression under SD and LD in the late accession, whereas we observed a weak expression in the early accession. *CqFT1B-2* expression showed similar expression patterns under both conditions in both accessions, indicating that it is not responding to photoperiod. *CqRFT* was, on the other hand, influenced by the photoperiod, and under SDs, its expression gradually dropped with the developmental stage. In contrast, the *CqRFT* expression of the early accession was very low or undetectable under both photoperiods (Figure 2).

CqCOL expression patterns were drastically different between the accessions. In the early accession, *CqCOL* genes were higher expressed at the early stages of development and gradually decreased during later developmental stages, showing similar expression patterns like the sugar beet floral repressor *BvFT1* (Pin et al. 2010). *CqCOL2A, CqCOL2B,* and *CqCOL4A-1* showed similar expression patterns in the late accession, while *CqCOL4A-2* was expressed at a very low rate. In the late accession, the *CqCOL5A* and *CqCOL5B* genes showed opposite expression patterns than the *CqFT2B* gene under SDs (Figure 2). This may indicate that the *CqCOL5A* and *CqCOL5B* are upstream repressors of *CqFT2B*.

Sequence variation in *FT* and *COL* orthologs correspond with the geographical origin and flowering time

We analyzed the haplotypes of *CqFT* genes and *CqCOL* genes based on high confidence SNPs (HC-SNPs) identified from a diversity panel of 303 quinoa accessions. We excluded individuals with heterozygous SNPs and missing SNPs in the corresponding gene during haplotype identification. We identified 2 to 6 haplotypes across all genes investigated (49 haplotypes in total) (Table 2).

Then, we studied the association between sequence variation in 12 putative flowering time regulators and flowering time under SDs and LDs. Moreover, we analyzed whether sequence variations altered the function of the predicted proteins. *CqFT1B-1* displayed the highest sequence variation (5 haplotypes) among *CqFT* genes, whereas *CqCOL2A* showed the highest sequence variation (6 haplotypes) among *CqCOL* genes. In contrast, seven genes did not display any SNPs causing amino acid (aa) substitutions (*CqFT1B-2*, *CqFT2A*, *CqFT2B*, *CqRFT*, *CqCOL4A-1*, *CqCOL5A, CqCOL5B*), indicating high protein conservation. The highest number of aa substitutions was identified in the *CqCOL2A* gene. Both *CqCOL5* genes were highly conserved, and SNPs were only found in the promoter regions (Figure 3 and Figure S8).

We then analyzed haplotype variation related to day length response. Seven haplotypes (*CqFT1B-1*$_b$, *CqFT1B-1*$_e$, *CqFT2A*$_a$, *CqRFT*$_d$, *CqCOL4A-2*$_c$, *CqCOL5A*$_a$, and *CqCOL5A*$_c$) resulted in earlier flowering under SDs, whereas accessions with five haplotypes (*CqFT1A*$_a$, *CqFT2B*$_c$, *CqCOL2A*$_c$, *CqCOL2B*$_b$, and *CqCOL5B*$_c$) flowered earlier under LDs. We identified 12 haplotypes of six genes (*CqFT1A*, *CqFT1B-2, CqFT2A*, *CqFT2B*, *CqCOL2A*, *CqCOL4A-1,* and *CqCOL5A*) that are not significantly associated with photoperiod response. Besides, we identified 14 haplotypes with moderate photoperiod sensitivity, whereas the remaining haplotypes were highly sensitive to photoperiod (Table 2, Figure 3, and Figure S8).

We hypothesize that the adaptation to LD conditions at Southern (e.g., Chile) and Northern latitudes was due to mutations within *CqFT* and *CqCOL* genes. Therefore, we grouped the haplotypes according to their abundance on geographical distribution. We found six genes (*CqFT1A*, *CqFT2B,* CqCOL2A, *CqCOL2B*, *CqCOL5A,* and *CqCOL5B)* with strong associations between geographical origin and haplotype (Figure 3 and Figure S8). For example, Southern Chilean accessions predominantly carry haplotype *CqFT1A*$_a$, whereas most Peruvian and Highland accessions have haplotype *CqFT1A*$_b$ (Figure 3). Also, haplotype *CqFT2B*$_c$ and *CqCOL5B*$_c$ (Figure 3) were mostly present in Southern Chilean accessions. These alleles should be preferentially used in European and North American breeding programs.

Haplotypes of six genes (*CqFT1B-1*, *CqFT1B-2*, *CqFT2A*, *CqRFT*, *CqCOL4A-1,* and *CqCOL4A-2*) were equally dispersed across all geographical regions (Figure S8). We did not compare the

geographical origin and haplotypes of the accession obtained from USDA gene banks and accessions grown in Europe because the exact origin of those accessions was not clear.

Discussion

This study provides the first comprehensive analysis of major flowering time regulators in quinoa. The photoperiod effect on flowering time in quinoa has been reported as early as 1949 (Fuller, 1949). But in the meantime, neither genetic nor molecular mechanism of flowering time regulation in quinoa have been discovered. We hypothesized that sequences with high similarity to Arabidopsis *FT*- and *COL*-genes are major flowering time regulators in quinoa and that haplotype variation is associated with flowering time variation and geographical distribution. We showed that under LD conditions, both *CqFT1* and *CqFT2* genes were expressed at an extremely low level. In contrast, the *CqCOL* genes were expressed under LD and SD conditions. However, under LD conditions, expression levels were lower compared to SDs. We could uncover sequence variation in those flowering time genes and their association with flowering under different photoperiodic regimes and geographical distribution. Associations were found between haplotype variation, day-length responsiveness, and geographical distribution, adding to the understanding of adaptation to varying environmental conditions.

Regulation of flowering is controlled by a complex network that involves multiple pathways with hundreds of genes (Blümel et al., 2015; Bouché, Lobet, Tocquin, & Périlleux, 2015). The *FT* gene plays a major role in flowering regulation in plant species (Wickland & Hanzawa, 2015). TFL is a closely related protein of FT with only 39 nonconservative amino acid differences. In contrast to *FT*, *TFL* strongly suppresses flowering (W. W. H. Ho & Weigel, 2014). During domestication, flowering time and plant architecture were modified due to mutations in *FT*/*TFL* (florigen/anti-florigen) (Eshed & Lippman, 2019). Quinoa, as a polyploid species with a short evolutionary history, carries multiple copies of *FT* homologs. Therefore, we focused on genes that are highly related to proven *FT* orthologs from the Amaranthaceae family (Drabešová et al., 2014; Pierre A Pin et al., 2010; Štorchová et al., 2019). We selected five *FT* homologs that had already been found in the reference genome sequence (Jarvis et al., 2017). Also, we chose an *RFT* homolog gene because, in the short day crop rice, *RFT* acts as a floral activator under LDs (Komiya, Yokoi, & Shimamoto, 2009). We expected a regulatory pathway as in other related species where *FT* is under the transcriptional control of *CO* or *COL* genes.

The CO protein activates *FT* transcription by binding to a unique *cis*-element in the *FT* promoter (Tiwari et al., 2010). Therefore, we decided to analyze *CO* in our study as it is a crucial factor in understanding the photoperiodic flowering pathway. Two B-box domains and a conserved C-terminal (CCT) domain are characteristic of Arabidopsis CO and related pseudo-response regulator (*PRR*) genes (Griffiths, Dunford, Coupland, & Laurie, 2003). All *CqCOL* genes we analyzed in this study were encoding two B-boxes and CCT domains.

Due to gene redundancy in polyploids, it is challenging to identify functional orthologues. During evolution, gene function changes due to the neo- or sub-functionalization of paralogs. Moreover, gene silencing and gene loss are reducing the complexity of paralogs over time. To get a clue about the function of our candidate genes, we measured their transcriptional activities. In allopolyploid species, subgenome-specific gene expression has been described (Bird, VanBuren, Puzey, & Edger, 2018). For instance, in shoots of tetraploid rapeseed (*Brassica napus*), the transcriptional activity of C genome genes was higher as compared to their A genome paralogs. Contrasting expression patterns were found in the roots of the same species (Chalhoub et al., 2014). Likewise, we observed expression pattern bias toward one subgenome (e.g., *CqFT2A* and *CqFT2B*). All the genes in our study were expressed at least at one developmental stage, which may indicate that they are not pseudogenes. Nevertheless, two genes, *CqFT1B-1*, *CqFT1B-2* are unlikely to function as flowering time regulators because their transcriptional profiles are in contrast to their Arabidopsis orthologs.

To which extent are *FT* gene functions conserved across species of the plant family Amaranthaceae? Since quinoa and beet belong to the same family, we hypothesized that *CqFT* genes are functional orthologs of the *B. vulgaris FT* genes (Pierre A Pin et al., 2010). Under SDs, three *CqFT* genes were highly expressed before the onset of flowering, similar to the floral repressor *BvFT1*. But only *CqFT2B* showed a diurnal expression pattern typical for *FT* orthologs. Moreover, this gene was also highly active before bolting. Therefore, we reason that *CqFT2B* is a floral activator under SDs. Our results suggest that this gene is the only functional ortholog of sugar beet florigen gene *BvFT2*. The discrepancy between beet and quinoa *FT* gene expression can be explained by the different life-history regimes of both species. In contrast to quinoa, sugar beet needs a long period of vernalization for the transition from the vegetative to the generative phase, which starts with shoot elongation (‘bolting’). In quinoa, the *FT* gene's function as a floral repressor would be highly unlikely because

quinoa has no vernalization requirement to flower. Therefore, it is likely that quinoa has other floral repressors such as *TFL* homologs because the balance between floral repressors and activators is important for the adaptation of crops to different environments (Eshed & Lippman, 2019). *CqFT2A* may also have alternative functions apart from flowering because it is highly expressed at very early and late developmental stages. It has been reported from other plants that FT proteins are involved in many developmental processes such as fruit setting, vegetative growth, tuberization, stomatal development, and spikelet development (Bi et al., 2019; Blümel et al., 2015). Evidently, *FT* genes in quinoa underwent neo-functionalization and gene silencing during evolution, typical for polyploids (Jiao et al., 2011).

Downregulation of *CqFT* genes under LD condition may be the reason for late flowering under LD like in soybean. Soybean is a SD plant where flowering is controlled by two *FT* homologs (*GmFT2a* and *GmFT5a*). Under LDs, both *GmFT2a* and *GmFT5a* genes are downregulated, and as a result, flowering is significantly delayed (Kong et al., 2010). Interestingly, under LDs, quinoa produces flowers without any *CqFT2* expression. Similarly, in *C. ficifolium* under LDs, the expression of *FT* homologs was negligible, although plants flowered regularly (Štorchová et al., 2019). It is tempting to speculate that other regulatory pathways control flowering in quinoa under LDs. The Arabidopsis *ft/tsf* double mutant can flower without the florigen (Yamaguchi, Kobayashi, Goto, Abe, & Araki, 2005). Flowering of these double mutants was found to be age-dependent, controlled by the microRNA156 and *SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE* (miRNA156/*SPL*) module (Hyun et al., 2019). Depending on day-length, rice produces two different florigens. In LDs, flowering is initiated by the activation of *RFT1*, a homolog of *Hd3a* (Tsuji, Taoka, & Shimamoto, 2011). However, quinoa *CqRFT* is unlikely to be a functional ortholog of *BvFT2* because it is expressed only in late accessions under SDs.

We questioned whether there is a *CO* ortholog in quinoa. Arabidopsis *CO* shows an evening peak in its diurnal expression (Andres & Coupland, 2012), which contrasts with all quinoa *COL* genes. *CqCOL* expression under SDs was higher than LDs, similar to sugar beet *COL1* (Chia et al., 2008). Also, *CqCOL* expressions in SDs are more similar to Arabidopsis *COL1* and *COL2* expression patterns. Therefore, these genes cannot be considered *CO* orthologues. Although Arabidopsis *COL1* and *COL2* are not involved in flowering regulation (Ledger et al., 2001), *COL* genes have been identified as floral

activators in other species. In legumes, *COL* genes show similar expression patterns to quinoa *COL* genes. In soybean (*Glycine max*) under LDs, *GmCOL1a* and *GmCOL1b* are repressors of florigen encoding genes *GmFT2a* and *GmFT5a,* whereas under SDs they are involved in the upregulation of florigen genes (Cao et al., 2015). In rice, the *CO* homolog *Hd1* acts as the SD floral activator, whereas under LDs *Hd1* turns into a repressor (Lee & An, 2015). Interestingly, *CqCOL* and *Hd1* display highly similar expression patterns under SDs. In conclusion, we reason that like in sugar beet, quinoa is lacking a true *CO* orthologue (Dally et al. 2014).

We have found substantial haplotype variation among flowering time genes in quinoa. We found varying degrees of photoperiod sensitivity depending on *FT-* and *COL* haplotypes. Allelic diversity in major flowering time genes is tightly linked with ecotype differentiation in crops, such as *Bna.A10.FLC* (Schiessl, Huettel, Kuehn, Reinhardt, & Snowdon, 2017) and *BnaA3.FRI* (Yi et al., 2018) in rapeseed and *Ppd-D1* (Guo, Song, Zhou, Ren, & Jia, 2010) in wheat. Overall, haplotype variations of *CqFT* genes are higher than *BvFT* genes (Höft, Dally, Hasler, & Jung, 2018), which is expected due to the short breeding history of quinoa. Remarkably, we did not identify non-functional haplotypes indicating the functional significance of *FT*-like and *COL* genes in quinoa adaptation. Interestingly, four genes (*CqFT1A*, *CqCOL2B*, *CqCOL4A-1*, and *CqCOL5B*) were associated with flowering time variation under LDs, but not under SDs, indicating a putative effect of these genes on flowering time. This might also be related to their different expression patterns under different photoperiods, which supports the assumption that quinoa uses different regulatory pathways to regulate flowering time under LDs and SDs. Notably, five alleles (*CqFT1A*$_a$, *CqFT2B*$_c$, *CqCOL2A*$_c$, *CqCOL2B*$_b$, and *CqCOL5B*$_c$) promoted early flowering under LD conditions. Accessions carrying those haplotypes are abundant in southern Chile with LDs during summer. However, differences in flowering time under LD and SD conditions of low abundant haplotypes should be taken with care due to small sample sizes (e.g., *CqFT2B*$_f$). Nevertheless, the early flowering alleles found in this study are important for quinoa breeding in northern latitudes.

How can haplotype variation uncovered in this study be used for breeding locally adapted quinoa varieties? Whenever crops have been moved from their center of origin to other geographical regions, photoperiod sensitivity has been essential for local adaptation (Meyer & Purugganan, 2013). Genetic regulation of photoperiodic sensitivity has been described in several crops. In soybean, allele

combinations of three major maturity loci *E1*, *E3*, and *E4* were important for local adaptation (Lu et al., 2017). Likewise, *Vgt1* and *ZmCCT* haplotype variation in maize (Huang et al., 2017) and *StCDF1* haplotype variation in potatoes (Gutaker et al., 2019) were major reasons for the worldwide distribution of these important crops. Consequently, breeders can fine-tune photoperiodic sensitivity and thus adjust flowering time by combining different alleles. Selecting quinoa genotypes that are early flowering under LDs is essential for cultivation in North America and Europe. Our study identified haplotypes that are causing photoperiod insensitivity associated with early flowering under LDs. We propose a haplotype-based breeding strategy by pyramiding favorable alleles (e.g., *CqFT1A*$_a$, *CqFT2B*$_c$, *CqCOL2A*$_c$, *CqCOL2B*$_b$, and *CqCOL5B*$_c$) to breed quinoa varieties which are early maturing under LDs and are thus suitable for cultivation in Northern latitudes.

Acknowledgments

We thank Prof. David Jarvis for providing access to the improved version of the quinoa reference genome and Prof. George Coupland for patiently read through the manuscript and provide valuable suggestions to improve it. Also, we thank Monika Bruisch, Brigitte Neidhardt-Olf, and Verena Kowalewski, for technical assistance and the Institute of Clinical Molecular Biology (IKMB, Kiel University) for Sanger sequencing. This work was financially supported by the Competitive Research Grant (Grant No. OSR-2016-CRG5-2966-02) of the King Abdullah University of Science and Technology, Saudi Arabia.

Author Contributions

D.S.R.P, E.A, and N.M planned the experiments, produced, and analyzed the data, and involved in manuscript writing. E.R, N.E, M.T, and C.J examined the data and were involved in manuscript writing.

Conflict of interest

The authors declare no conflict of interest.

Data Availability

The data generated in this study are available upon request to the corresponding author.

References

Andres, F., & Coupland, G. (2012). The genetic basis of flowering responses to seasonal cues. *Nature Reviews Genetics, 13*(9), 627-639. doi:10.1038/nrg3291

Bertero, H., King, R., & Hall, A. (1999). Photoperiod-sensitive development phases in quinoa (*Chenopodium quinoa* Willd.). *Field Crops Research, 60*(3), 231-243.

Bertero, H. D. (2003). Response of developmental processes to temperature and photoperiod in quinoa (*Chenopodium quinoa* Willd.). *Food Reviews International, 19*, 87-97. doi:10.1081/FRI-120018870

Bi, X., van Esse, G. W., Mulki, M. A., Kirschner, G., Zhong, J., Simon, R., & von Korff, M. (2019). CENTRORADIALIS interacts with *FLOWERING LOCUS T*-like genes to control spikelet initiation, floret development and grain number. *Plant physiology*, pp.01454.02018. doi:10.1104/pp.18.01454

Bird, K. A., VanBuren, R., Puzey, J. R., & Edger, P. P. (2018). The causes and consequences of subgenome dominance in hybrids and recent polyploids. *New Phytologist, 220*, 87-93. doi:10.1111/nph.15256

Blümel, M., Dally, N., & Jung, C. (2015). Flowering time regulation in crops-what did we learn from Arabidopsis? *Current Opinion in Biotechnology, 32*, 121-129. doi:10.1016/j.copbio.2014.11.023

Bouché, F., Lobet, G., Tocquin, P., & Périlleux, C. (2015). FLOR-ID: an interactive database of flowering-time gene networks in *Arabidopsis thaliana*. *Nucleic Acids Research, 44*(D1), D1167-D1171. doi:10.1093/nar/gkv1054

Cao, D., Li, Y., Lu, S., Wang, J., Nan, H., Li, X., . . . Yuan, X. (2015). *GmCOL1a* and *GmCOL1b* function as flowering repressors in soybean under long-day conditions. *Plant and Cell Physiology, 56*(12), 2409-2422.

Cháb, D., Kolář, J., Olson, M. S., & Štorchová, H. (2008). Two *FLOWERING LOCUS T* (*FT*) homologs in *Chenopodium rubrum* differ in expression patterns. *Planta, 228*, 929-940. doi:10.1007/s00425-008-0792-3

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. a. P., Tang, H., Wang, X., . . . Wincker, P. (2014). Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science, 345*, 950-953. doi:10.1126/science.1253435

Chia, T. Y. P., Müller, a., Jung, C., & Mutasa-Göttgens, E. S. (2008). Sugar beet contains a large *CONSTANS-LIKE* gene family including a *CO* homologue that is independent of the early-bolting (*B*) gene locus. *Journal of experimental botany, 59*, 2735-2748. doi:10.1093/jxb/ern129

Christiansen, J. L., Jacobsen, S.-E., & Jørgensen, S. T. (2010). Photoperiodic effect on flowering and seed development in quinoa (*Chenopodium quinoa* Willd.). *Acta Agriculturae Scandinavica, Section B-Soil & Plant Science, 60*(6), 539-544.

Dally, N., Eckel, M., Batschauer, A., Höft, N., & Jung, C. (2018). Two *CONSTANS-LIKE* genes jointly control flowering time in beet. *Scientific Reports, 8*, 16120. doi:10.1038/s41598-018-34328-4

Dally, N., Xiao, K., Holtgräwe, D., & Jung, C. (2014). The *B2* flowering time locus of beet encodes a zinc finger transcription factor. *Proceedings of the National Academy of Sciences, 111*(28), 10365-10370.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., . . . Group, G. P. A. (2011). The variant call format and VCFtools. *Bioinformatics, 27*(15), 2156-2158. doi:10.1093/bioinformatics/btr330

Drabešová, J., Cháb, D., Kolař, J., Haškovcová, K., & Štorchová, H. (2014). A dark-light transition triggers expression of the floral promoter *CrFTL1* and downregulates *CONSTANS-like* genes in a short-day plant *Chenopodium rubrum*. *J Exp Bot, 65*(8), 2137-2146. doi:10.1093/jxb/eru073

Eshed, Y., & Lippman, Z. B. (2019). Revolutions in agriculture chart a course for targeted breeding of old and new crops. *Science, 366*(6466).

Fuller, H. J. (1949). Photoperiodic responses of *Chenopodium quinoa* Willd. and *Amaranthus caudatus* L. *American Journal of Botany*, 175-180.

Golicz, A. A. (2019). Analysis of the quinoa genome reveals conservation and divergence of the flowering pathways.

González, J. A., Eisa, S., Hussin, S., & Prado, F. E. (2015). Quinoa: an Incan crop to face global changes in agriculture. *Quinoa: Improvement and sustainable production*, 1-18.

Griffiths, S., Dunford, R. P., Coupland, G., & Laurie, D. A. (2003). The evolution of *CONSTANS-like* gene families in barley, rice, and Arabidopsis. *Plant physiology, 131*(4), 1855-1867.

Guo, Z., Song, Y., Zhou, R., Ren, Z., & Jia, J. (2010). Discovery, evaluation and distribution of haplotypes of the wheat *Ppd-D1* gene. *New Phytologist, 185*(3), 841-851.

Gutaker, R. M., Weiß, C. L., Ellis, D., Anglin, N. L., Knapp, S., Fernández-Alonso, J. L., . . . Burbano, H. A. (2019). The origins and adaptation of European potatoes reconstructed from historical genomes. *Nature Ecology & Evolution, 3*(7), 1093-1101.

Hayama, R., Agashe, B., Luley, E., King, R., & Coupland, G. (2007). A Circadian Rhythm Set by Dusk Determines the Expression of *FT* Homologs and the Short-Day Photoperiodic Flowering Response in Pharbitis. *The Plant Cell, 19*, 2988-3000. doi:10.1105/tpc.107.052480

Ho, J., Tumkaya, T., Aryal, S., Choi, H., & Claridge-Chang, A. (2019). Moving beyond P values: data analysis with estimation graphics. *Nature Methods, 16*(7), 565-566. doi:10.1038/s41592-019-0470-3

Ho, W. W. H., & Weigel, D. (2014). Structural features determining flower-promoting activity of Arabidopsis *FLOWERING LOCUS T*. *The Plant Cell, 26*, 552-564. doi:10.1105/tpc.113.115220

Höft, N., Dally, N., Hasler, M., & Jung, C. (2018). Haplotype Variation of Flowering Time Genes of Sugar Beet and Its Wild Relatives and the Impact on Life Cycle Regimes. *Frontiers in Plant Science, 8*. doi:10.3389/fpls.2017.02211

Huang, C., Sun, H., Xu, D., Chen, Q., Liang, Y., Wang, X., . . . Tian, F. (2017). *ZmCCT9* enhances maize adaptation to higher latitudes *Proceedings of the National Academy of Sciences, 115*, E334-E341. doi:10.1073/pnas.1718058115

Hyun, Y., Vincent, C., Tilmes, V., Bergonzi, S., Kiefer, C., Richter, R., . . . Coupland, G. (2019). A regulatory circuit conferring varied flowering response to cold in annual and perennial plants. *Science, 363*(6425), 409-412.

Jacobsen, S.-E., & Stølen, O. (1993). Quinoa - Morphology, phenology and prospects for its production as a new crop in Europe. *European Journal of Agronomy, 2*, 19-29. doi:10.1016/S1161-0301(14)80148-2

Jarvis, D. E., Ho, Y. S., Lightfoot, D. J., Schmöckel, S. M., Li, B., Borm, T. J., . . . Saber, N. (2017). The genome of *Chenopodium quinoa*. *Nature, 542*(7641), 307.

Jiao, Y., Wickett, N. J., Ayyampalayam, S., Chanderbali, A. S., Landherr, L., Ralph, P. E., . . . Soltis, P. S. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature, 473*(7345), 97-100.

Komiya, R., Yokoi, S., & Shimamoto, K. (2009). A gene network for long-day flowering activates *RFT1* encoding a mobile flowering signal in rice. *Development, 136*, 3443-3450. doi:10.1242/dev.040170

Kong, F., Liu, B., Xia, Z., Sato, S., Kim, B. M., Watanabe, S., . . . Harada, K. (2010). Two coordinately regulated homologs of *FLOWERING LOCUS T* are involved in the control of photoperiodic flowering in soybean. *Plant physiology, 154*(3), 1220-1231.

Ledger, S., Strayer, C., Ashton, F., Kay, S. A., & Putterill, J. (2001). Analysis of the function of two circadian-regulated *CONSTANS-LIKE* genes. *The Plant Journal, 26*(1), 15-22.

Lee, Y., & An, G. (2015). Complex regulatory networks of flowering time in rice. *J. Rice Res, 3*(2).

Lu, S., Zhao, X., Hu, Y., Liu, S., Nan, H., Li, X., . . . Kong, L. (2017). Natural variation at the soybean *J* locus improves adaptation to the tropics and enhances yield. *Nature genetics, 49*(5), 773-779.

Luccioni, L., Krzymuski, M., Sánchez-Lamas, M., Karayekov, E., Cerdán, P. D., & Casal, J. J. (2019). CONSTANS delays Arabidopsis flowering under short days. *The Plant Journal, 97*(5), 923-932. doi:10.1111/tpj.14171

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., . . . Finn, R. D. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research, 47*(W1), W636-W641.

Maldonado-Taipe, N., Sarange, D., Schmöckel, S., Jung, C., & Emrani, N. (2020). Validation of suitable genes for normalization of diurnal gene expression studies in *Chenopodium quinoa*. *bioRxiv*.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., . . . Daly, M. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research, 20*(9), 1297-1303.

Meyer, R. S., & Purugganan, M. D. (2013). Evolution of crop species: genetics of domestication and diversification. *Nature reviews. Genetics, 14*, 840-852. doi:10.1038/nrg3605

Milne, I., Shaw, P., Stephen, G., Bayer, M., Cardle, L., Thomas, W. T. B., . . . Marshall, D. (2010). Flapjack--graphical genotype visualization. *Bioinformatics (Oxford, England), 26*(24), 3133-3134. doi:10.1093/bioinformatics/btq580

Murphy, K. S., & Matanguihan, J. (2015). *Quinoa: Improvement and sustainable production*: John Wiley & Sons.

Paradis, E. (2010). pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics, 26*(3), 419-420.

Patiranage, D. S. R., Rey, E., Emrani, N., Wellman, G., Schmid, K., Schmöckel, S. M., . . . Jung, C. (2020). Genome-wide association study in the pseudocereal quinoa reveals selection pattern typical for crops with a short breeding history. *bioRxiv*, 2020.2012.2003.410050. doi:10.1101/2020.12.03.410050

Pin, P. A., Benlloch, R., Bonnet, D., Wremerth-Weich, E., Kraft, T., Gielen, J. J., & Nilsson, O. (2010). An antagonistic pair of *FT* homologs mediates the control of flowering time in sugar beet. *Science, 330*(6009), 1397-1400.

Pin, P. A., Zhang, W., Vogt, S. H. H., Dally, N., Büttner, B., Schulze-Buxloh, G., . . . Müller, A. E. E. (2012). The role of a pseudo-response regulator gene in life cycle adaptation and domestication of beet. *Current biology : CB, 22*, 1095-1101. doi:10.1016/j.cub.2012.04.007

Rojas, W. (2003). Multivariate Analysis of Genetic Diversity of Bolivian Quinoa Germplasm. *Food Reviews International, 19*(1-2), 9-23. doi:10.1081/FRI-120018864

Schiessl, S. V., Huettel, B., Kuehn, D., Reinhardt, R., & Snowdon, R. J. (2017). Flowering Time Gene Variation in *Brassica* Species Shows Evolutionary Principles. *Frontiers in Plant Science, 8*, 1-13. doi:10.3389/fpls.2017.01742

Song, Y. H., Kubota, A., Kwon, M. S., Covington, M. F., Lee, N., Taagen, E. R., . . . Imaizumi, T. (2018). Molecular basis of flowering under natural long-day conditions in *Arabidopsis*. *Nature Plants, 4*, 824-835. doi:10.1038/s41477-018-0253-3

Štorchová, H. (2020). The Evolution of the *FLOWERING LOCUS T-Like* (*FTL*) Genes in the Goosefoot Subfamily Chenopodioideae. In P. Pontarotti (Ed.), *Evolutionary Biology—A Transdisciplinary Approach* (pp. 325-335). Cham: Springer International Publishing.

Štorchová, H., Hubáčková, H., Abeyawardana, O. A., Walterová, J., Vondráková, Z., Eliášová, K., & Mandák, B. (2019). *Chenopodium ficifolium* flowers under long days without upregulation of *FLOWERING LOCUS T* (*FT*) homologs. *Planta, 250*(6), 2111-2125.

Tiwari, S. B., Shen, Y., Chang, H.-C., Hou, Y., Harris, A., Ma, S. F., . . . Ratcliffe, O. J. (2010). The flowering time regulator *CONSTANS* is recruited to the *FLOWERING LOCUS T* promoter via a unique cis-element. *New Phytologist, 187*(1), 57-66. doi:10.1111/j.1469-8137.2010.03251.x

Tsuji, H., Taoka, K. I., & Shimamoto, K. (2011). Regulation of flowering in rice: Two florigen genes, a complex gene network, and natural variation. *Current Opinion in Plant Biology, 14*, 45-52. doi:10.1016/j.pbi.2010.08.016

Wickland, D. P., & Hanzawa, Y. (2015). The *FLOWERING LOCUS T*/*TERMINAL FLOWER 1* Gene Family: Functional Evolution and Molecular Mechanisms. *Molecular Plant, 8*, 983-997. doi:10.1016/j.molp.2015.01.007

Yamaguchi, A., Kobayashi, Y., Goto, K., Abe, M., & Araki, T. (2005). *TWIN SISTER OF FT* (*TSF*) acts as a floral pathway integrator redundantly with *FT*. *Plant and Cell Physiology, 46*(8), 1175-1189.

Yi, L., Chen, C., Yin, S., Li, H., Li, Z., Wang, B., . . . Liu, K. (2018). Sequence variation and functional analysis of a *FRIGIDA* orthologue (*BnaA3.FRI*) in *Brassica napus*. *BMC Plant Biology, 18*, 1-13. doi:10.1186/s12870-018-1253-1

Figures and Tables

Table 1: Summary statistics of photoperiod response of 276 quinoa accessions. Ten plants from each accession were grown in a growth chamber with 16h and 8h light, and DTF was determined. Estimation graphics with 5000 bootstraps at 99% confidence interval was used to compare DTF of each accession under SDs and LDs**.**

Photoperiodic sensitivity	**No. of accessions**	**Percentage %**
Short day responsive[a]	186	67.39
Day neutral	46	16.67
Long day responsive[b]	21	7.61
Not assigned	23	8.33

[a] flowering earlier under SD

[b] flowering earlier under LD

Rajapaksha Patirange Dilan Sarange (Orcid ID: 0000-0001-6308-1838)

Emrani Nazgol (Orcid ID: 0000-0001-5673-3957)

Table 2: Classification of haplotypes based on their photoperiod sensitivity (*P*-values obtained from LD and SD mean comparison). For haplotype identification, high-quality SNP data were used from 303 accessions of a worldwide core collection. Haplotypes were identified based on HC-SNP. Accessions with missing or heterozygous SNPs within the respective gene were excluded from the analysis.

Gene	No. of Accessions	Haplotype		
		Insensitive	**Moderately sensitive**	**Highly sensitive**
		(*P*>0.05)	**(0.001<*P*<0.05)**	**(*P*<0.001)**
CqFT1A	197	*CqFT1A$_c$*	*CqFT1A$_a$*	*CqFT1A$_b$*
		CqFT1A$_d$		
CqFT1B-1	168		*CqFT1B-1$_b$*	*CqFT1B-1$_a$*
			CqFT1B-1$_c$	
			CqFT1B-1$_d$	
			CqFT1B-1$_e$	
CqFT1B-2	216	*CqFT1B-2$_b$*	*CqFT1B-2$_a$*	*CqFT1B-2$_c$*
				CqFT1B-2$_d$
CqFT2A	183	*CqFT2A$_e$*	*CqFT2A$_b$*	*CqFT2A$_a$*
		CqFT2A$_d$		*CqFT2A$_c$*
CqFT2B	185	*CqFT2B$_e$*		*CqFT2B$_a$*
		CqFT2B$_f$		*CqFT2B$_b$*
				CqFT2B$_c$
				CqFT2B$_d$
CqRFT	213		*CqRFT$_b$*	*CqRFT$_a$*
				CqRFT$_c$
				CqRFT$_d$
CqCOL2A	142	*CqCOL2A$_e$*	*CqCOL2A$_c$*	*CqCOL2A$_a$*
		CqCOL2A$_f$	*CqCOL2A$_d$*	*CqCOL2A$_b$*
CqCOL2B	145		*CqCOL2B$_b$*	*CqCOL2B$_a$*
				CqCOL2B$_c$
CqCOL4A-1	223	*CqCOL4A-1$_b$*		*CqCOL4A-1$_a$*
CqCOL4A-2	222		*CqCOL4A-2$_b$*	*CqCOL4A-2$_a$*
				CqCOL4A-2$_c$
CqCOL5A	176	*CqCOL5A$_b$*	*CqCOL5A$_c$*	*CqCOL5A$_a$*
		CqCOL5A$_d$		
CqCOL5B	212		*CqCOL5B$_c$*	*CqCOL5B$_a$*
				CqCOL5B$_b$

Rajapaksha Patirange Dilan Sarange (Orcid ID: 0000-0001-6308-1838)

Emrani Nazgol (Orcid ID: 0000-0001-5673-3957)

Figure legends

Figure 1: Expression profiles of 12 *FT*-like and *COL* genes in an early (Titicaca, seed code:171230) and a late quinoa accession (PI-587173, seed code: 171605) grown under LDs (A, C, E, G, I, K) and SDs (B, D, F, H, J, L). Legends used within LD graphs are also valid for the corresponding SD graph. RNA was isolated at the bolting stage. ZT, zeitgeber time. Continuous lines: PI-587173; Broken lines: Titicaca; Error bars, ± SEM. Data are normalized against the geometric mean of *CqPTB* and *CqIDH-A*.

Figure 2: Expression profiles at different stages of development of six *FT*-like and six *COL* genes in the late accession PI-587173 (seed code: 171605) (A-L) and in the early accession Titicaca (seed code: 171230) (M-X). Plants were grown in the climate chamber under LDs (open boxes) and SDs (filled boxes). Samples were collected at ZT4. Error bars: ± SEM, Data were normalized against the geometric mean of *CqPTB* and *CqIDH-A*. (Veg.: Vegetative, WBB: Week/s before bolting, Bolt.: Bolting, WAB: Week/s after bolting, WBF: Week/s before flowering, Flow.: Flowering, WAF: Week/s after flowering).

Figure 3: Phenotypic effects of haplotype variations in four different flowering time genes (*CqFT1A*, *CqFT2B*, *CqCOL2B,* and *CqCOL5B*). The number of accessions analyzed is given in Table 2. (A) Gene structure and nucleotide polymorphisms. Different colors indicate polymorphic nucleotides. SNP positions are given relative to the start codon. The asterisks depict missense variants. (B) Haplotype networks. Circle size illustrates the frequency of the corresponding haplotype. The colors of the pie chart depict different classes of photoperiod response. SD: early flowering under short days, LD: early flowering under long days, DN: day-neutral, no difference between short and long days. Crosslines between two circles are representing the number of nucleotide polymorphisms between haplotypes. (C) Haplotype depending flowering time (days to flowering, DTF) as determined in the climate chamber under SD and LD growth conditions. The *P*-values of mean comparisons between SD and LD within haplotypes are written above the boxplots. An ANOVA was performed to determine significance among haplotypes and grouping of pairwise multiple comparison was obtained using Tukey's post hoc multiple comparison test. Student's t-test was performed to compare DTF means within haplotypes. The black and blue letters indicate significant DTF differences among haplotypes under LD and SD photoperiod, respectively. 'n' indicates the number of accessions. (D) Geographical

origin of quinoa accessions with their respective haplotypes. Only accessions are shown where passport data are available.

Supporting Information

Figure S1: Flowering time variation of 276 accessions grown under short (8h) (B) and long day (16h) (A) conditions. C and D, geographic coordinates of 123 quinoa accessions. Color gradient corresponds to flowering time, which was determined under long days (C) and under short days (D).

Figure S2: Protein alignment of PEBP proteins of quinoa, sugar beet, *C. rubrum*, rice, and Arabidopsis.

Figure S3: The CORE1 (TGTGA), CORE2 (TGTGG), P1 (CCACA) and P2 (TGTGA) *cis*-regulatory elements in 6kb region of six *CqFT* gene promoters.

Figure S4: Protein alignment of COL and CO proteins of quinoa, sugar beet, *C. rubrum*, rice, and Arabidopsis.

Figure S5: Phylogenetic relationships between PEBP family genes (A) and COL genes (B) in quinoa.

Figure S6: Comparison between growth under short-day (SD) and long-day conditions (LD). Plants were grown in 9 cm pots in controlled growth chamber conditions. A: Three weeks after sowing B: Seven weeks after sowing. C: Different pigmentation under LDs (left) and SDs (right) in two accessions sensitive to photoperiod; CHEN-109 and PI-587173.

Figure S7: Phenological development and yield components of four accessions grown in the growth chamber under short-day (SD) and long-day (LD) conditions. Seed code 171230: Titicaca; seed code 170867: PI-614886; seed code 171605: PI-587173 and seed code 170876: CHEN-109). Error bars: SEM

Figure S8: Phenotypic effects of haplotype variations in eight different flowering time genes. The number of accessions analyzed is given in Table 2. (A) Gene structure and nucleotide polymorphisms. Different colors indicate polymorphic nucleotides. SNP positions are given relative to the start codon. The asterisks depict missense variants. (B) Haplotype networks. Circle size illustrates the frequency of the corresponding haplotype. The colors of the pie chart represent different classes of photoperiod

response. SD: early flowering under short days, LD: early flowering under long days, DN: no difference between short- and long day flowering time. Crosslines between two circles are representing the number of nucleotide polymorphisms between haplotypes. (C) Haplotype depending flowering time (days to flowering, DTF) as determined in the climate chamber under SD and LD growth conditions. The black and blue letters indicate significant DTF differences among haplotypes under LD and SD photoperiod, respectively. The *P*-values of mean comparisons between SD and LD within haplotypes are written above the boxplots. An ANOVA was performed to determine significance among haplotypes and grouping of pairwise multiple comparison was obtained using Tukey's post hoc multiple comparison test. Student's t-test was performed to compare DTF means within haplotypes. 'n' indicates the number of accessions. (D) Geographical origin of quinoa accessions with their respective haplotypes. Only accessions are shown for which passport data are available.

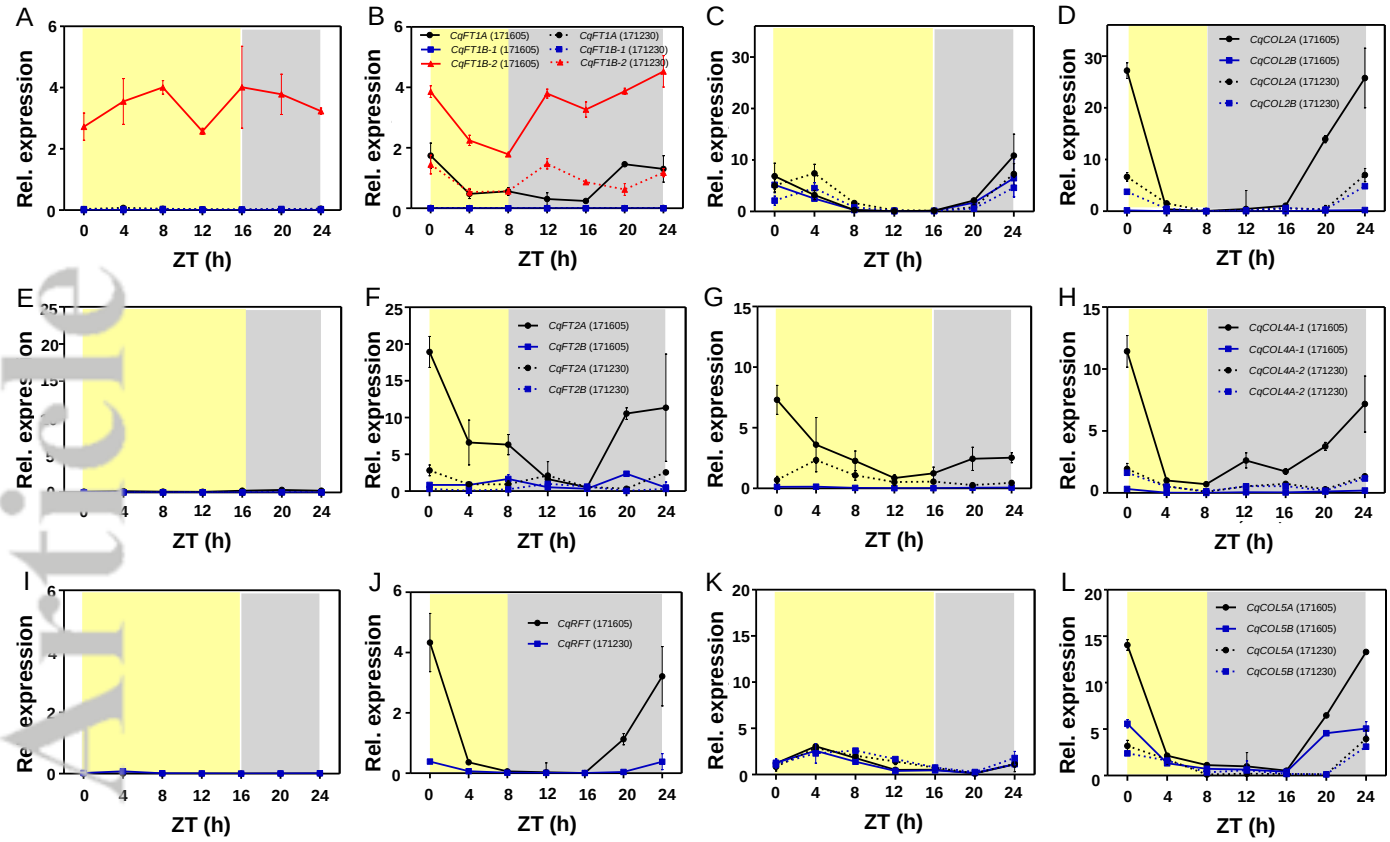
Table S1: Plant material used in this study and flowering time under short and long day photoperiods.

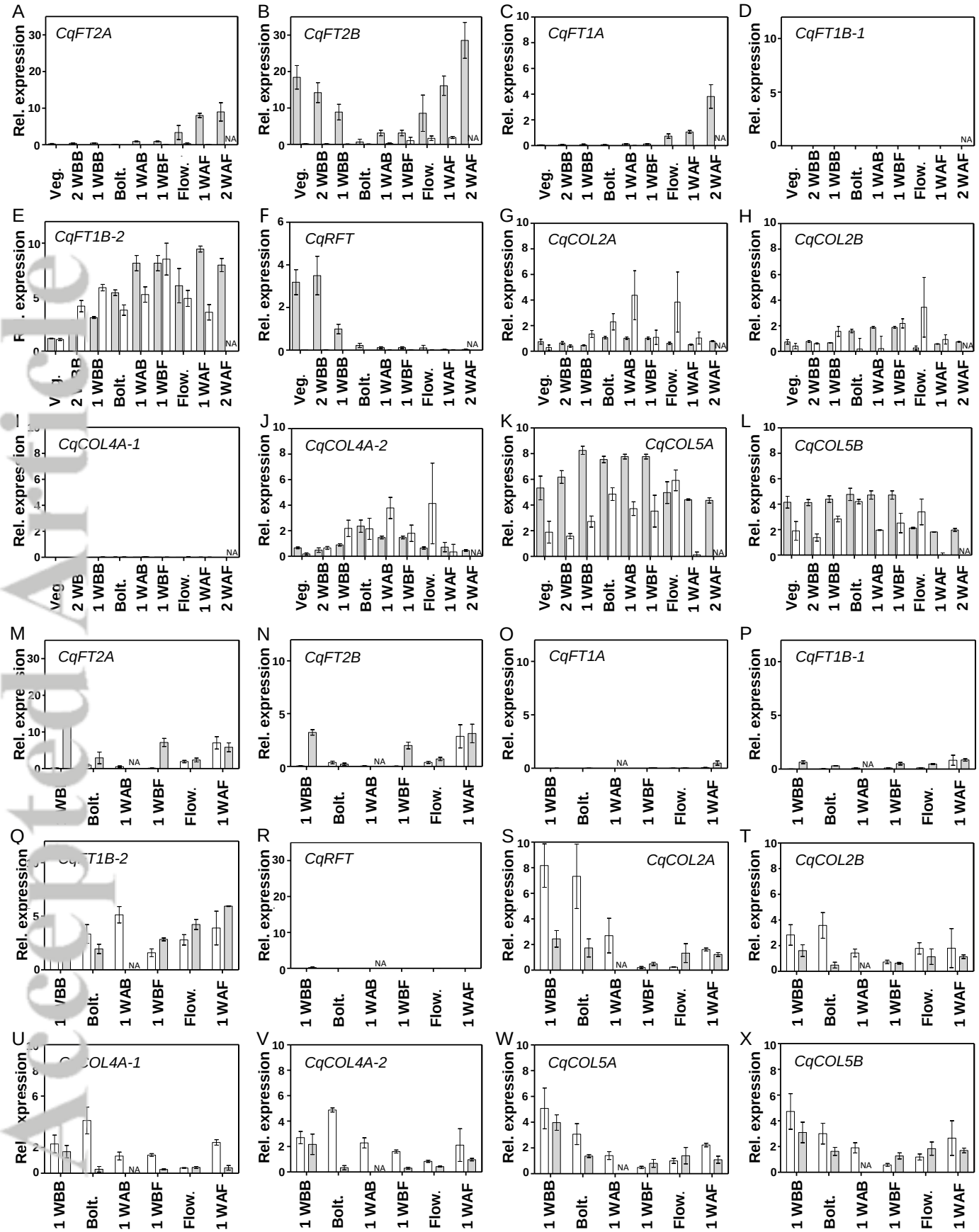
Table S2: List of PEBP family genes in quinoa. [a] according to Jarvis et al. (2017) and reference genome V2 annotation. [b] according to Štorchová 2020

Table S3: *CONSTANS-like* genes in quinoa analyzed in this study.

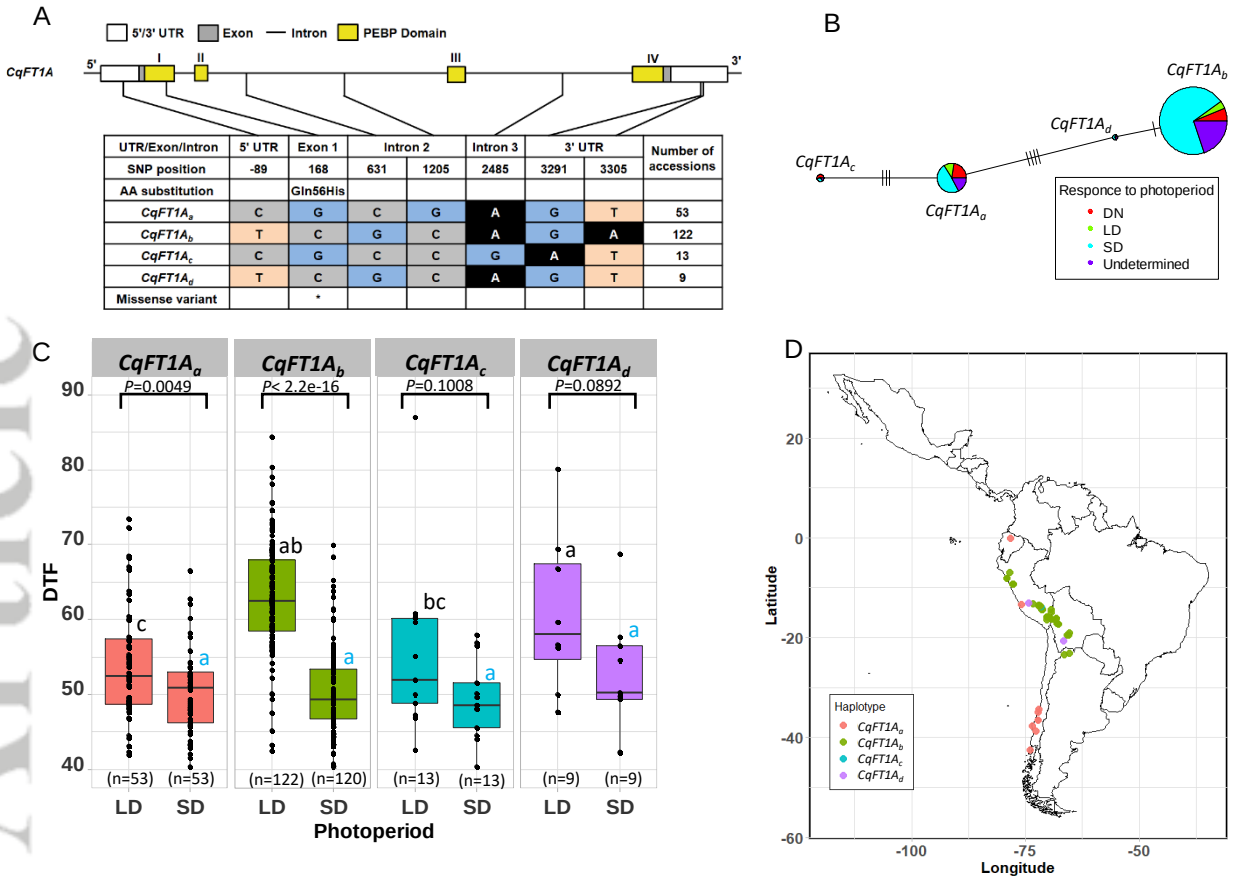
Table S4: Primer combinations used for RT-qPCR.

Table S5: Sampling stages for temporal expression analysis. The age is given in the brackets as weeks after sowing, and '✓' denotes the sample availability. In accession PI-587173 (seed code: 171605), WAB=WBF.



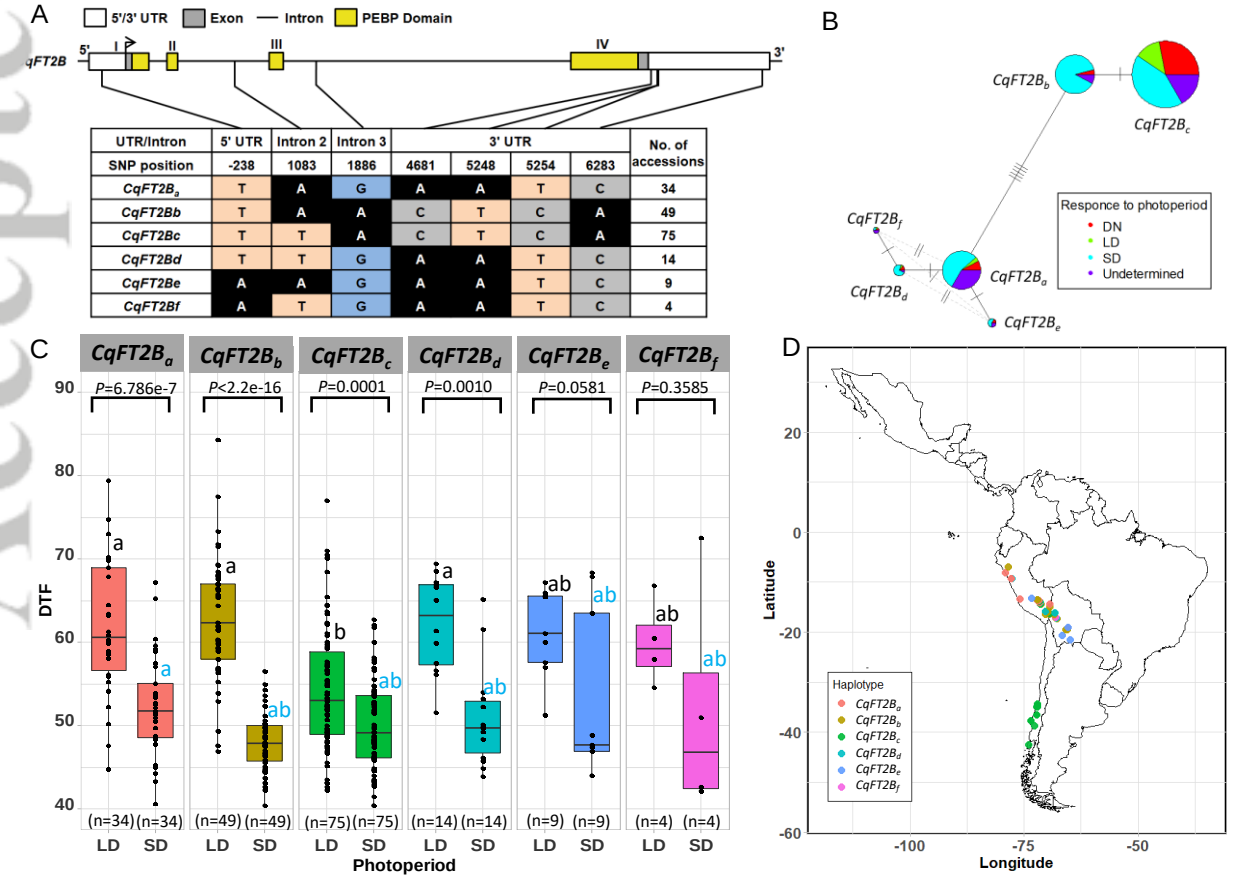


CqFT1A



UTR/Exon/Intron	5' UTR	Exon 1	Intron 2		Intron 3	3' UTR		Number of accessions
SNP position	-89	168	631	1205	2485	3291	3305	
AA substitution		Gln56His						
CqFT1A_a	C	G	C	G	A	G	T	53
CqFT1A_b	T	C	G	C	A	G	A	122
CqFT1A_c	C	G	C	C	G	A	T	13
CqFT1A_d	T	C	G	C	A	G	T	9
Missense variant		*						

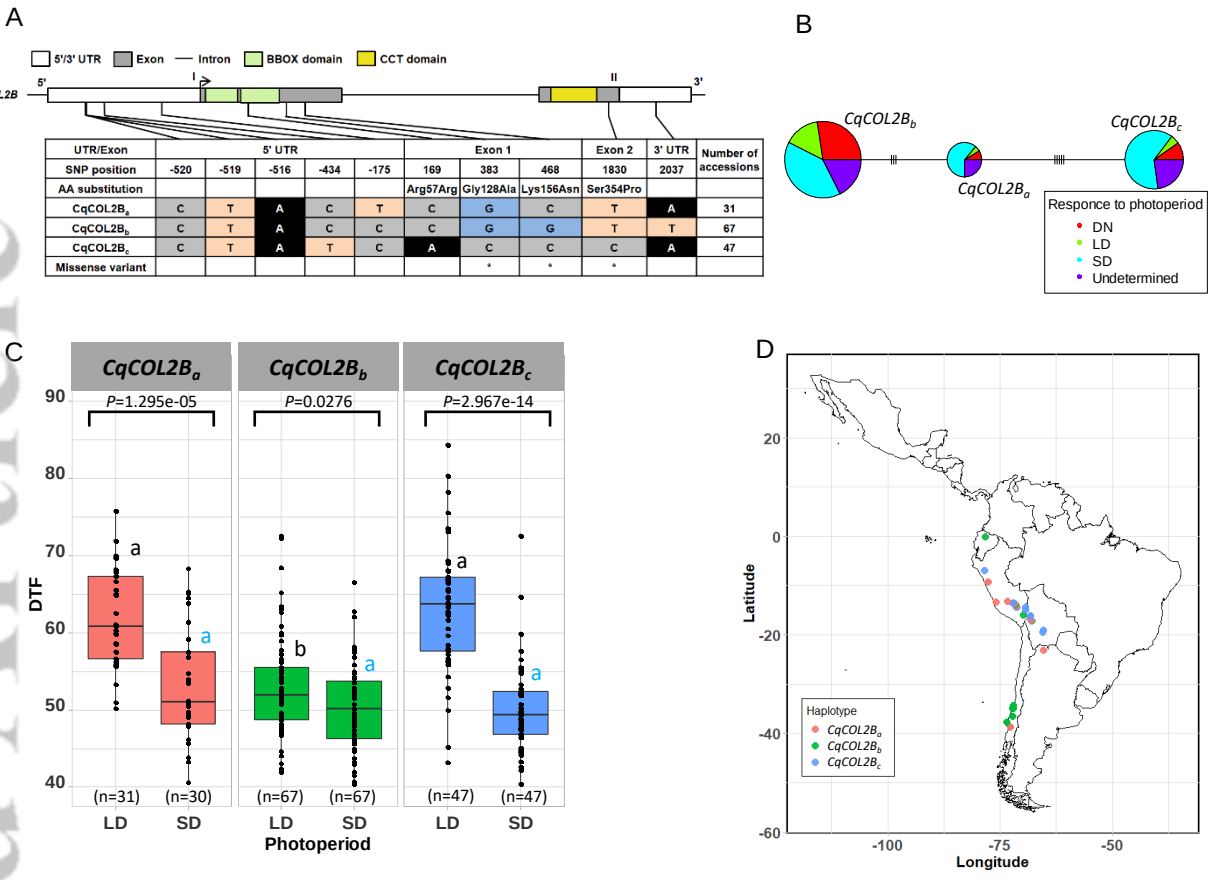
CqFT2B



UTR/Intron	5' UTR	Intron 2	Intron 3	3' UTR				No. of accessions
SNP position	-238	1083	1886	4681	5248	5254	6283	
CqFT2B_a	T	A	G	A	A	T	C	34
CqFT2Bb	T	A	A	C	T	C	A	49
CqFT2Bc	T	T	A	C	T	C	A	75
CqFT2Bd	T	T	G	A	A	T	C	14
CqFT2Be	A	A	G	A	A	T	C	9
CqFT2Bf	A	T	G	A	A	T	C	4

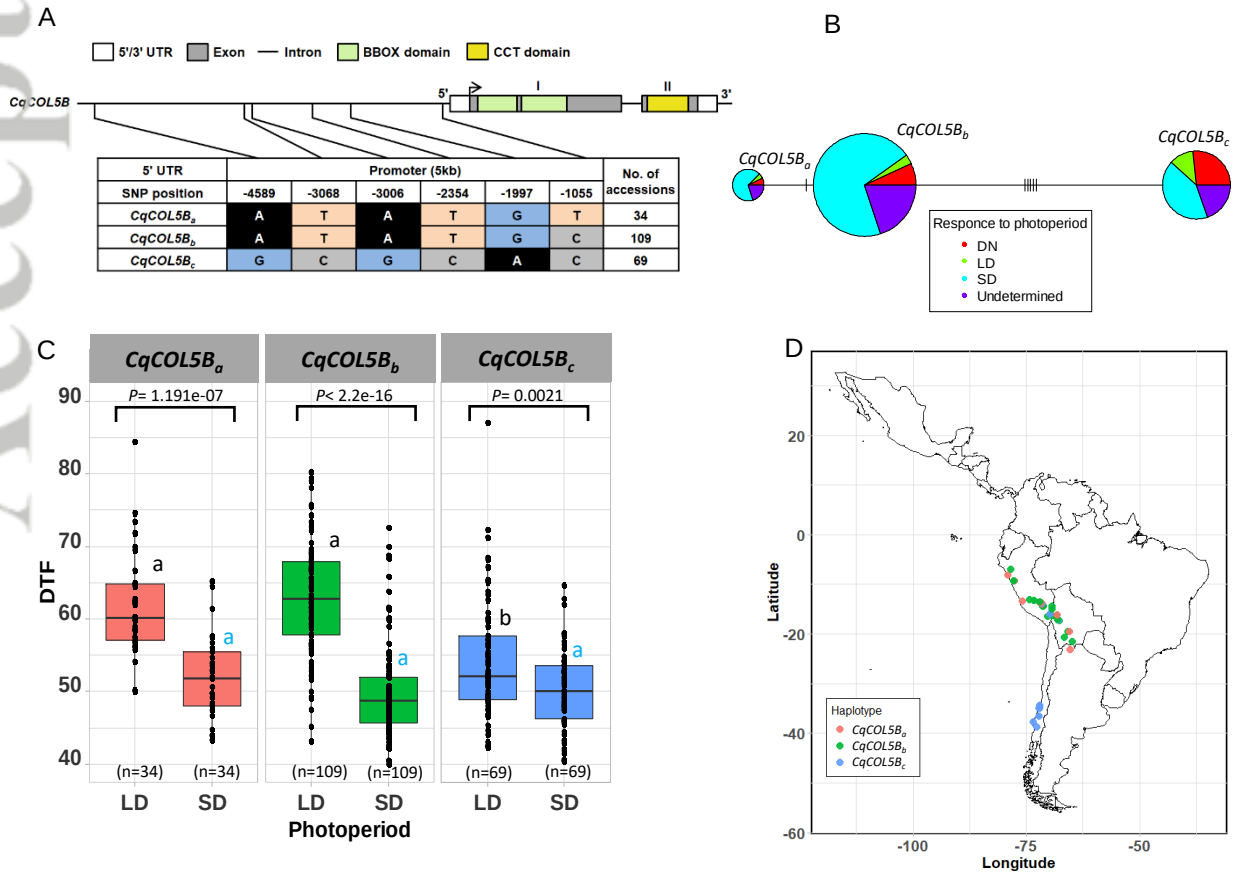
Figure 3: *continued on next page, legend follows*

CqCOL2B



UTR/Exon	5' UTR					Exon 1			Exon 2	3' UTR	Number of accessions
SNP position	-520	-519	-516	-434	-175	169	383	468	1830	2037	
AA substitution						Arg57Arg	Gly128Ala	Lys156Asn	Ser354Pro		
CqCOL2B$_a$	C	T	A	C	T	C	G	C	T	A	31
CqCOL2B$_b$	C	T	A	C	C	C	G	G	T	T	67
CqCOL2B$_c$	C	T	A	T	C	A	C	C	C	A	47
Missense variant							*	*	*		

CqCOL5B



5' UTR	Promoter (5kb)						No. of accessions
SNP position	-4589	-3068	-3006	-2354	-1997	-1055	
CqCOL5B$_a$	A	T	A	T	G	T	34
CqCOL5B$_b$	A	T	A	T	G	C	109
CqCOL5B$_c$	G	C	G	C	A	C	69

Summary statement

We studied structural variations and expression profiles of major flowering time regulators of Quinoa. Haplotype variation was correlated with phenological development in response to photoperiod sensitivity. We propose haplotype building to breed cultivars adapted to northern latitudes.