Flax latitudinal adaptation at LuTFL1 altered architecture and promoted fiber production

Rafal M. Gutaker1,2, Maricris Zaidem2, Yong-Bi Fu3, Axel Diederichsen3, Oliver Smith4,5, Roselyn Ware1 & Robin G. Allaby1

After domestication in the Near East around 10,000 years ago several founder crops, flax included, spread to European latitudes. On reaching northerly latitudes the architecture of domesticated flax became more suitable to fiber production over oil, with longer stems, smaller seeds and fewer axillary branches. Latitudinal adaptations in crops typically result in changes in flowering time, often involving the PEBP family of genes that also have the potential to influence plant architecture. Two PEBP family genes in the flax genome, LuTFL1 and LuTFL2, vary in wild and cultivated flax over latitudinal range with cultivated flax receiving LuTFL1 alleles from northerly wild flax populations. Compared to a background of population structure of flaxes over latitude, the LuTFL1 alleles display a level of differentiation that is consistent with selection for an allele III in the north. We demonstrate through heterologous expression in Arabidopsis thaliana that LuTFL1 is a functional homolog of TFL1 in A. thaliana capable of changing both flowering time and plant architecture. We conclude that specialized fiber flax types could have formed as a consequence of a natural adaptation of cultivated flax to higher latitudes.
later flowering, suggesting the potential importance of this family in latitudinal adaptations of crops during the spread of agriculture\(^2\). Furthermore, \(FT\) and \(TFL1\) can control plant stature through promotion or delaying floral initiation\(^3\) and are highly conserved in eudicots\(^4\). The orthologs of \(FT\) and \(TFL1\) in tomato have a strong effect on tomato fruit weight and number\(^5\), plant architecture and height, and flowering time\(^6,7\). Such differences in fruit weight and plant stature underlie the distinction between fiber and oil specialized varieties in flax, raising the possibility that the PEBP gene family in flax could be associated with both architectural changes and adaptation to northerly latitudes through modifying flowering time.

In this study we screened PEBP gene orthologs in flax for signatures of selection along increasing latitudes in European wild and cultivated flax. We additionally genotyped restriction site-associated regions of flax genome to understand the underlying population structure in flax. Finally, we validated the effect of a gene candidate for selection, \(LuTFL1\) on plant architecture and flowering time through heterologous expression in Arabidopsis thaliana.

**Results**

**PEBP orthologs in flax and the diversity at \(LuTFL1\) locus.** In order to investigate the role of PEBP genes in latitudinal adaptation and architecture control in cultivated flax, we surveyed orthologs of this gene family. A scan for homologs of \(A.\ thaliana\) PEBP genes in the flax assembly\(^8\) revealed eight loci of interest with architectures similar to \(TFL1\) gene (Supplementary Fig. S1; Supplementary Table S1). Out of those, seven loci were amplified and sequenced from flax genomic DNA (Supplementary Fig. S2), four could be easily re-sequenced without cloning and showed conserved four-exon structure (Supplementary Fig. S1). Phylogenetic analyses comparing the flax, \(A.\ thaliana\) and Lombardy poplar orthologs revealed that three putative PEBP members in flax are closely related to floral suppressors TFL1 and ATC, while one has similarity to their antagonist, FT (Supplementary Fig. S3). Resequencing of those genes in six accessions of flax indicated high levels of polymorphism in pale flax at the \(LuTFL1\) locus, suggestive of broad standing variation.

We re-sequenced the \(LuTFL1\) locus in 32 wild and 113 cultivated flax accessions (Supplementary Fig. S4A, Supplementary Table S2). For comparison, we also re-sequenced the \(LuTFL2\) locus in the same population of flaxes. We discovered wide molecular diversity at both loci, with 16 and 11 alleles in \(LuTFL1\) and \(LuTFL2\) respectively (Supplementary Table S3). A phylogenetic network of \(LuTFL1\) shows that allelic diversity is greater in pale flax (top plane, Fig. 1a) when compared to cultivated flax (bottom plane, Fig. 1a). The majority (100 out of 103) of cultivated flaxes were associated with just two of the groups (I and III) also found in the wild. There were several groups apomorphic to groups I and III that appeared only in our cultivated samples that likely represent phasmism in pale flax at the \(LuTFL2\) locus, suggestive of broad standing variation.

There are multiple alleles in both \(LuTFL1\) and \(LuTFL2\) that are shared by pale and cultivated flaxes. Such a pattern could be a result of multiple alleles being included during domestication or through post-domestication gene flow from wild to domestic species. The phylogeographic distribution of \(LuTFL1\) alleles in pale flax (Fig. 1c), supports the notion that allele I was passed from pale flax to cultivated flax during the initial domestication process in the Near East. Conversely, in pale flax allele III is associated with the Bosphorus strait and the Balkans beyond the domestication area but not in Eastern or Southern Turkey. We conclude that \(LuTFL1\) III likely entered the cultivated flax gene pool through gene flow from pale flax during the agricultural expansion into Europe, and reached higher frequencies in more northerly latitudes (Fig. 1b). The high diversity of \(LuTFL2\) alleles present in the cultivated gene pool but absent from the wild gene pool suggests that they may have originated from wild populations outside our sample set, or have become extinct in the wild. Despite not being closely located in the genome, allele X of \(LuTFL2\) occurs in a strong disequilibrium with allele III of \(LuTFL1\) (Supplementary Table S4), shows a similar latitudinal gradient (Supplementary Fig. S5B), and is absent from cultivated flaxes carrying the \(LuTFL1\) allele I.

The significant latitudinal clines of \(LuTFL1\) alleles I more frequent in the south and III more frequent in the north (Fig. 1b) are more evident in the historic landraces (Supplementary Fig. S6). A notable exception to this trend is allele XII from the group I associated alleles which occurs at higher frequencies at higher latitudes, but is absent from the historic samples. We considered populations as ‘northern’ or ‘southern’ relative to 40th circle of latitude of northern hemisphere for calculation of fixation index statistics \(F_{ST} = 0.26\), which indicated high population differentiation at \(LuTFL1\) locus. Enrichment of \(LuTFL1\) III in the north could be a result of drift combined with isolation by distance (IBD), or alternatively it could be an effect of selection acting on adaptive alleles.

We tested if \(LuTFL1\) and \(LuTFL2\) alleles are distributed in the frequencies that match the expectation under neutral evolution using Tajima’s D, R2, Fu and Li’s D and F methods (Supplementary Table S5). Based on \(LuTFL2\) data we could not reject the neutrality hypothesis, however, significantly negative scores in \(LuTFL1\) tests indicated an excess of rare alleles in cultivated flax. Such a pattern of molecular diversity could be explained either by population structure, demographic effects such as expansion, or selection. To distinguish between these possibilities, we investigated the population structure and allele frequencies at neutral loci through a restriction site-associated (RAD) sequencing approach\(^3,5\).

**Population structure from genome-wide SNPs in wild and pale flax.** We surveyed the genomes of 90 accessions of pale (28) and cultivated (62) flax (Supplementary Table S2) by RAD sequencing\(^3\), from which we found 993 polymorphic RAD tags encompassing the total of 1686 SNPs present in at least 80% accessions. Based on those, we calculated genetic distances, which were used for multidimensional scaling analysis. Three main clusters that correspond to three populations were revealed: cultivated flax, dehiscent flax and pale flax (Fig. 2A). Using an ADMIXTURE\(^5\) approach we identified four distinct ancestral components (lowest 5-fold cross validation error for \(K = 4\)). The pale flax of Eastern...
Anatolia represents all the admixture components that are present in cultivated flaxes, including dehiscent varieties (Fig. 2B), which supports an eastern population involvement with the original domestication process. Two pale flaxes located in eastern Greece outside the domestication center show sizeable admixture with dehiscent varieties. This is congruent with previous reports that identify dehiscent varieties as genetically distinct from other domesticated flax varieties, possibly reflecting an earlier independent domestication37,38. A latitudinal gradient is apparent in the cultivated flax types, evident from a tendency for southern populations to be affiliated to an orange grouping, and a cyan grouping in northern populations in Fig. 2B.

Randomly sampled RADtags across the genome are a good proxy for neutrally evolving parts of the genome. We used allele frequencies of RAD SNPs in southern and northern flax populations as a null distribution for changes over latitudes due to population structure and demographic processes. We then tested allele frequencies of three $LuTFL1$ SNPs, which distinguish allele I and III, against RAD distribution. Firstly, we compared fixation index statistics, Fig. 3A. We found that the $F_{ST}$ values for three SNPs associated with the differences between $LuTFL1$ alleles I and III are significantly higher than the background range of values based on RAD loci (q-value 0.045). Secondly, we placed the position of the $LuTFL1$ alleles on the two dimensional allele frequency spectrum of RAD loci (frequencies of northern population on x-axis; southern population on y-axis), and considered the relative deviation of allele frequency from the diagonal distribution, Fig. 3B. We found that the three SNPs defining allele III lie significantly outside the background range (Supplementary Table S6), but close to the edges of the null distribution, whereas SNPs associated with allele XII are within the general range. Together, these data suggest that allele III was selected for as cultivated flax moved northwards.

Given the proximity of the general differentiation of allele III between northern and southern populations relative to the null distribution evident in the allele frequency spectrum, we speculated that weak selection was probably involved. To investigate this further we applied a method to determine the most likely selection coefficient associated with allele frequency changes over time39, using archaeological dates of the arrival of agriculture at different latitudes to date the latitudinal allele frequencies, Fig. 4. Our estimates confirmed weak selection with

---

**Figure 1.** Haplotype network of $LuTFL1$ locus in cultivated and pale flax: (a) Splits Tree network of pale (top) and cultivated (bottom) flax, size of nodes is proportional to number of samples with the same haplotype, continuous branches denote molecular distance between haplotypes, vertical dotted lines link different species within same haplotype. (b) Histogram showing latitudinal gradient of $LuTFL1$ alleles in cultivated flax, in green, frequency of northern haplotype cluster ($LuTFL1$.III), in red, frequency of southern haplotype cluster ($LuTFL1$.I) with fitted logistic regression curve (p-value of 0.00144) reflecting occurrence probability of northern haplotype in latitude gradient. (c) Map of Europe marked with wild distribution of pale flax (black line) and pale flax sampling locations (colours correspond to haplotypes in splits network).
values of $s$ generally between 0.001 and 0.007 (mean 0.003), but interestingly we found an increase in selection strength over time and latitude, indicating that allele III was progressively more strongly selected for as flax moved to higher latitudes.

Finally, we investigated whether the RAD loci data harboured evidence of movement of alleles between wild and cultivated flax populations beyond the area of domestication, which we implied from geographic distribution of LuTFL1 alleles in wild and cultivated flaxes (Fig. 1b,c). We observed no clear signal using $f_3$ statistics (cultivatedN; cultivatedS, wildN), however, varying rates of population differentiation through founder effects and multiple complex processes of gene flow can mask signals of gene movement between populations. We therefore considered portions of the data where evidence of gene movement may be clearer. We identified subsets of alleles that were discriminatingly higher at a range of frequencies in either northern or southern populations of wild flax. We reasoned that if cultivated flax had spread northwards without contact with wild populations, then the differentiation between northern and southern populations at these subsets of loci should not be perturbed relative to the background distribution of loci differentiation. We found that there was no significant difference between the cultivated populations at loci that were higher in southern wild populations, but that there was an elevated level of differentiation between cultivated populations at loci that were differentially high in northern but not southern wild populations (Supplementary Table S7). This indicates a bias against the neutral expectations of uncorrelated drift between the cultivated and wild population pairs that could represent gene flow from northern wild populations into northern cultivated populations. Alternatively, this signature could be an effect of a parallel selection for traits associated with high frequency RAD loci alleles in the wild population. We deemed the latter unlikely since 78% of loci that had elevated frequencies in northern wild populations also had elevated frequencies in northern cultivated populations, which would require an explanation of selection being responsible more frequently than drift for high frequency alleles in the north.

**Functional homology of LuTFL1.** We explored the function of LuTFL1 and LuTFL2, first by investigating expression of these genes in the leaf and shoot apex tissue of pale flax plants from six populations over the course of the plant development to flowering using a semi-quantitative PCR approach. LuTFL1 was expressed continuously from the first point of measurement (40 days) to flowering (Supplementary Fig. S8) consistent with TFL1 expression in A. thaliana as opposed to ATC, which is expressed in the hypocotyl in A. thaliana. We detected no expression of LuTFL2 in shoot tissue. Based on phylogenetic evidence, LuTFL1 is more similar to...
Figure 3. Outliers to neutral allele frequencies between northern and southern cultivated flax populations: (A) $F_{ST}$ values for RAD neutral SNPs plotted as a function of heterozygosity. Outliers with elevated $F_{ST}$ and heterozygosity are marked with red circles. Outlier SNPs distinguishing alleles I and III in $LuTFL1$ locus are marked with black. (B) Two dimensional allele frequency spectrum of RAD loci alleles between northern and southern cultivated flaxes. Frequency of $LuTFL1$.III SNPs are marked as circles. To the right: probability heat maps of sampling frequency combinations given the allele frequency spectrum based on distance from the diagonal.

Figure 4. Strength of selection on $LuTFL1$ III in modern and historic flaxes. Estimates of selection coefficient over latitude and time, where latitude correlates to time of arrival of agriculture.
ATC, however characterization of its expression pattern led us to conclude that LuTFL1 is most likely functionally orthologous to TFL1.

We tested the hypothesis that LuTFL1 functions as TFL1 by complementing TFL1 knock-out mutants of A. thaliana with a LuTFL1 construct under a 35 S promoter. We obtained a stable line after transforming a tfl1–2 mutant in Landsberg erecta (Ler) genomic background using an Agrobacterium system. Over 160 plants of 35 S::LuTFL1 inbred progeny (T2) were grown in long day conditions. For all the plants we measured days to bolting, days to open flowers, days to end of flowering and stem height at the end of flowering. Flowering time phenotypes segregated in T2 with a Mendelian 15 to 1 ratio (Supplementary Fig. S9), which suggests that two copies of transgene were integrated in tfl1–2 mutant. Transgene presence/absence was validated using primers specific to the transgene in 21 plants with extreme phenotypes, Supplementary Fig. S9. Plants that contained the 35 S::LuTFL1 transgene were compared with tfl1–2 mutants and Ler-0 wild type. The start and end of flowering was significantly lower in tfl1–2 mutants than in Ler-0 and 35 S::LuTFL1 plants (Fig. 5; Supplementary Fig. S9), suggesting that two copies of transgene were integrated in tfl1–2 mutant. Transgene presence/absence was validated using primers specific to the transgene in 21 plants with extreme phenotypes, Supplementary Fig. S9. Plants that contained the 35 S::LuTFL1 transgene were compared with tfl1–2 mutants and Ler-0 wild type. The start and end of flowering was significantly lower in tfl1–2 mutants than in Ler-0 and 35 S::LuTFL1 plants (Fig. 5; Supplementary Fig. S9), suggesting that the flax LuTFL1 I allele can rescue non-functional TFL1 in A. thaliana. Constitutive expression of this allele under 35 S promoter further delays flowering, when compared to Ler-0. Subsequently, the amount of expressed TFL1/LuTFL1 has a positive impact on plant height at the end of flowering. We conclude that, similar to TFL1 in A. thaliana, LuTFL1 functions in delaying flowering, promoting indeterminate growth and finally resulting with increased plant height at senescence. The experimental data from transgenic 35 S::LuTFL1 A. thaliana plants support the notion that the correlation between allele type and phenotype could have a causative relationship.

Although LuTFL1 is a functional homolog of TFL1, we still do not fully understand the effect of allelic diversity in flax at this locus. We detected no non-synonymous substitutions between groups I and III that could account for a potential functional basis for the difference in selection histories between the alleles. However, a transcription factor-binding site was identified to be present in the promoter regions of LuTFL1 III and VIII, but absent from LuTFL1 I (Supplementary Fig. S4B). We hypothesized that such a difference between the allele groups could result in a change of expression pattern of LuTFL1 and in consequence, altered flowering time, which in turn could be under selection in the northerly latitudes. To help understand how an adaptation to latitude through the TFL/FT system could produce the observed flowering and architectural changes in flax we applied a plant development model. The model was used to explore how selection at the TFL1 locus and

Figure 5. Flax LuTFL1 gene complements TFL1 mutant in Arabidopsis thaliana: (A) Images of plants in 2x magnification at the median time to first open flower in tfl1–2 mutant (20 days). Both Ler-0 wild type and tfl1–2 mutant complemented with 35 S::LuTFL1 are not yet flowering at that time. (B) Plants at their median time to end of flowering in respective cohorts. Mutant tfl1–2 ended flowering after 23 days and reached average height of 69 mm; by contrast Ler-0 wild type and 35 S::LuTFL1 plants flowered until 36th and 40th day reaching 134 and 211 mm height respectively.
latitudinal movement could affect stem height, which correlates with increased fiber content. The model predicts that increased TFL1 expression is expected to produce longer stemmed plants, but also that at higher latitudes such plants would be more fit through a greater rate of flower production relative to long stemmed forms in the south (Fig. 6). This model outcome predicts that adaptation to variable climate in northerly latitudes by indeterminate flowering through selection at a TFL1 locus allows increase in stem height, which is associated with an improvement in fiber content.

Discussion
Crop plants were domesticated independently around the world at the ‘Domestication Centers’ in prehistoric times. Early farmers spread out from those centers and brought with them major crops as they expanded into new geographic zones. There are numerous examples of crop plants that adapted to new environments through changes in flowering time, particularly during northwards expansions. Maize, domesticated in Mesoamerica, would only be successfully cultivated in the temperate North America after adapting through reduction of flowering time. Similarly, after domestication of rice in Eastern China, its successful cultivation in Korea and Japan required changes in flowering behaviour. In many plants, TFL1 family genes that influence flowering time by acting antagonistically with FT to delay flowering have been identified as targets for selection in adaptation to northern latitudes and have also been shown to influence plant architecture and fruit yield. TFL1 homologs have played an important role in adaptation to temperate climate in barley and soybean, where it has been noted that adaptive variant at this locus is strongly enriched in the northern end of crop distribution.

Cultivated flax was domesticated in the Near East alongside wheat and barley and likewise adapted to temperate climate of prehistoric Europe. In this paper, we show that flax genome encodes multiple PEBP orthologs. Of those, we show that LuTFL1 has latitudinally structured allelic diversity and molecular signatures consistent with selection. This has been validated against null distribution of neutrally evolving loci, which we generated using RADseq approach. To further reinforce evidence of selection on LuTFL1.III in the northern Europe, we anticipate the availability of full genome information for flax populations. Similarly to other crop plants, adaptation to Northerly latitudes in flax was achieved through modifications in flowering behaviour. Through heterologous overexpression of LuTFL1, we show that this gene is in fact a functional homolog of TFL1 and is capable of changing flowering time. Similarly, LuTFL1 is shown to be a decisive factor in establishing plant architecture and height. Based on that we conclude that the increase in frequency of LuTFL1.III in the north and signatures consistent with natural selection are results of flax adaptation in Europe through changes in flowering time. Interestingly, this change had an impact on flax architecture and could have predisposed flax to fibre production. This evolutionary transition may have been key to the rise of the textile revolution in central Europe, which is evidenced in the reduction of flax seed size and improved tools for fiber production in Neolithic contexts. Our conclusions are consistent with phenotypic associations in cultivated flax, in which fibre content is correlated with plant height and flowering time at the same time.

Flax is shown to have adapted to Northerly latitudes in a similar fashion to other crop plants. However, this study additionally provides an unusual example on how the influence of natural adaptations within a crop may ultimately influence the use of that crop. Normally, the evolution of domesticated plants is associated with an initial rise of a suite of traits collectively termed the domestication syndrome, later followed by diversification or improvement traits. Such traits are considered for the most part to be agronomically advantageous. Mounting evidence shows that plant adaptations within the human environment are not restricted to crops and within crop plants not all adaptations are of obvious benefit to agronomic productivity and yield, but related to the wider
ecology\textsuperscript{58}. Examples include a tendency to increase rates of water loss in domesticates\textsuperscript{59}. Such adaptations highlight a tension between the effective exploitation of plants as a resource by endowment of useful traits and ongoing natural adaptations that occur in plants that may compromise productivity. This study shows that the adaptation of flax to the European variable climate during Neolithic likely involved the \textit{LuTFL1} locus and resulted in a change of cultivation purpose through substantial modification of the stem height, which is correlated with the technical fiber quality.

Natural adaptations in crops are commonly acquired from wild relatives’ standing variation. One case of such transfer is the Mexicana teosinte introgression into highland maize landraces\textsuperscript{60}. It is notable that flax natural adaptation appears to have occurred through a transmission from the wild to the cultivated gene pool despite the highly selfing nature of both cultivated and pale flax, at only 3–5% outcrossing. Furthermore, wild flax stands are typically of low density. It is possible in this case that entomophilous pollination could have played an important role either by natural pollinators or even the action of domesticated bees of the early farmers\textsuperscript{61}. Further research is required to understand the interplay of these evolutionary forces and how they contribute to the tensions between artificial and natural selection.

Materials and Methods

Plant materials. Samples from 16 populations were collected from natural habitats in the summer of 2011 from Croatia, Montenegro, Albania, Greece and Bulgaria. These were combined with a further 16 populations of pale flax supplied by the Plant Genetic Resources of Canada, Agriculture and Agri-Food Canada, Saskatoon Research Centre. Seed material for 28 modern cultivars of flax and 18 landraces were obtained from the Plant Genetic Resources of Canada, Agriculture and Agri-Food Canada, Saskatoon Research Centre. A further 28 historic landrace accessions were obtained from Plant Genetic Resources of the N. I. Vavilov Research Institute of Plant Industry, St. Petersburg, Russia. Samples ranging in age up to 194 years were sampled from Herbaria variously from the Natural History Museum, University of Oxford and the University of Warsaw.

Identification and resequencing of \textit{LuTFL1} locus. Based on multiple alignments that contain sequences from different eudicots, degenerate primers were designed in exonic regions to cover between 400 bp of \textit{TFL1} region. The putative sequences of \textit{TFL1} in flax were subject to BLAST searches against the cultivated flax genome scaffolds v1.0 and CDs v1.0 databases\textsuperscript{62} in order to identify full sequences of genes of interest and their flanking regions. Phylogenetic trees with \textit{TFL1} homologs from \textit{A. thaliana} and \textit{P. nigra} were estimated using Bayesian inference in MRBAYES v3.2\textsuperscript{63}.

A total of 148 samples were used in this study, including 58 cultivars, 18 landraces, 38 historic landraces of cultivated flax, 32 pale flax samples and two \textit{L. decumbens} accessions. The total DNA from 20 mg of seeds was isolated using modified DNEasy Plant Mini Kit (Qiagen) protocol. In case of herbarium specimens – mixed plant tissue was used for extraction using CTAB protocol\textsuperscript{64}. A 1367 bp target region was amplified from \textit{LuTFL1} spanning exons 1–4 using primers (5'-TTACACTCCAAAGGAATGTC, and 5'-TGTCCTGCCGTGAGCATT). Haplotype networks was made using uncorrected_P character transformation and Rooted Equal Angle splits transformation in SplitsTree\textsuperscript{45}. The size of the network nodes was made proportional to the number of samples that share a particular haplotype. The relationship between latitude and \textit{LuTFL1} haplogroups I or III was modeled using logistic regression using \textit{glm} function with \textit{family = “binomial”} setting in \textit{R} programme and plotted using \textit{logi.hist.plot} function from \textit{popbio} package.

Selection signatures at \textit{LuTFL1} locus. In order to test if \textit{LuTFL1} and \textit{LuTFL2} were inherited neutrally or were under selection, neutrality tests with variety of estimators were carried out. Tajima's D\textsuperscript{66}, Fu and Li's D, Fu and Li's F\textsuperscript{67} statistics were calculated in Intrapop and the R2 statistic\textsuperscript{68} was calculated in using the Pegas package. The extent of linkage disequilibrium was examined between the alleles of the \textit{LuTFL1} locus. Expected frequencies of genotypes were calculated from the observed individual allele frequencies. The correlation coefficient of \textit{D}, \textit{r}, was calculated from observed genotyping data where:

\[
r = \frac{[p_1 p_2] \times [q_1 q_2] - [p_1 q_2] \times [q_1 p_2]}{\sqrt{p_1 p_2 q_1 q_2}}
\]

Square brackets indicate the observed proportion of the sample of genotype combinations, \(p_1, p_2, q_1\) and \(q_2\) refer to the allele frequencies of \(p\) and \(q\) type alleles at the first and second locus respectively, and the numerator equates to the statistic \(D\).

Each latitude in Europe is associated with a different time of arrival of agriculture based on archaeological data. The program \textit{SELECTION_TIME.pl}\textsuperscript{39} estimates selection coefficients from dated frequencies. The model takes as input a series of dated frequencies, mating strategy of the organism, whether the data is phenotype or family\textsuperscript{69} setting in \textit{R} programme and plotted using \textit{logi.hist.plot} function from \textit{popbio} package.
RAD sequencing and processing. From the total collection, 90 plants were chosen for the RADseq experiment (Supplementary Table S1). They represent 28 pale flax and 62 cultivated flax accessions. Plants were grown in glasshouse conditions and harvested after seedlings reached 10 cm of height. Snap-frozen material was ground with 3 glass beads (3 mm diameter) in TissueLyser machine (Qiagen) and then used for DNA isolation with DNeasy Plant Mini Kit (Qiagen) following the manufacturer’s manual. DNA was digested with 2U of SbfI HF restriction enzyme (New England Biolabs) for one hour and then the reaction was heat inactivated. For RADseq library preparation, 40 μl of genomic DNA at concentration of 25 ng/μl was prepared following previous methods. Barcoded libraries of 10 different DNA isolates were pooled together into the total of 2 μg of DNA in 300 μl of solution. DNA was sheared ten times on ice in Bandelin Sonoplus HD 2070 sonicator. Sheared DNA was purified using AMPure XP beads (Agencourt) in proportion 1 to 1 with buffer. Approximately 20 ng of library was amplified with Phusion HF Master Mix (NEB) in total of 14 cycles. The first 50 libraries were sequenced on HiSeq. 2000 Illumina platform at Oxford Genomics Centre using the TrueSeq reagents. Another 40 samples were submitted for sequencing on Genome Analyzer II Illumina platform in Genomic Centre at University of Warwick.

Raw reads from Illumina sequence runs were assembled into FastQ format and assessed in FASTQC software v0.10.1 to check for standard quality measures: quality scores, overrepresented sequences, GC content and N content. We generated SNPs from de novo assembled RADtags. Scripts from the STACKS pipeline v1.05 was employed to de-multiplex and discover RAD sequence markers. Low quality sequences and those characterized by erroneous barcodes were discarded. Remaining sequences were sorted according to their barcodes. SNPs were called with the following settings: 5 identical, raw reads required to create marker stack, 5 mismatches were allowed between alleles in single individual and calling haplotypes from secondary reads was disabled.

Population structure and differentiation. Distances calculated from identity by sequence between all individuals were measured using plink (v1.07). This distance matrix was used for multi-dimensional scaling (MDS) approach in R with cmdscale function. Population ancestry of each individual was modelled using ADMIXTURE (v1.23). We assumed existence of two to seven ancestral populations and chose the model, which was characterised with the lowest five-fold cross-validation error.

We utilized the ancestral information content concept to employ a high variant test to explore possible gene movement between populations when there may be complex confounding movements masking a general signal. We identified three subsets of loci in which the difference in frequency (Δf) between northern and southern wild populations was more than 0.5, 0.4 or 0.3, with the higher allele in the northern population. We examined the ancestry information content of the Δf loci subsets using the Ia (I_{a}) statistic relative to the information content of all loci (I_{p}). We then compared the Δf subset mean FST with the total null background of markers 100000 randomly sampled subsets, for details see Supplementary Methods.

Cultivated flax individuals were segregated into two groups based on their latitude; individuals from above N40° were included in the northern subpopulation while individuals from below N40° in southern. We calculated size-corrected and uncorrected FST values between northern and southern cultivated flax populations for each RADtag using functions in OutFLANK R package. We applied trimming approach to infer the distribution of FST for neutral markers and used likelihood estimation for outliers. Allele frequency spectrums were generated from in house scripts from STRUCTURE format files acquired from de novo assembly RAD tags. P value heat maps were generated from the allele frequency spectrums using in house scripts that calculated probability density functions of Gaussian distributions from mean and standard deviations of frequency distributions perpendicular to the diagonal.

Expression and functional homology of LuTFL1. Plants of six populations (W042, W043, W067, W069, W077 and W094) were sown in 100 replicates each. Seeds were stratified for 3 days in the dark at 4 °C and grown at room temperature with a 16-hour photoperiod for 10 days followed by vernalization in cold room at 4 °C for a period of 40 days. Subsequently plants were moved to growth chambers with 16 h daylight at 24°C until they flowered. Samples of three plants were taken at 0, 15, 17, 19 and 21 days. Samples were snap-frozen and the total RNA extraction was carried out from 20 mg of ground tissue using mirVana miRNA isolation kit (Invitrogen) following the standard protocol. DNA contaminants were digested in reaction with 2 units of DNase I (Invitrogen) and 2 μg of total RNA in DEPC-treated water. cDNA was synthesized from 1 μg of DNase-treated RNA with use of SuperScript II reverse transcriptase and unspecific Oligo dT primer following the manufacturer instructions. Specific primers were designed in exonic regions of LuTFL1, LuTFL2 and LuFT to cover approximately 150 bp long fragment. The GAPDH housekeeping gene was chosen as an expression control. Semi-quantitative PCR was carried out for each sample with 30 cycles.

We have used Landsberg erecta (Ler-0) ecotype with functional TFL1 gene as a reference and its tfl1 mutant, tfl1–2 for complementation experiment with LuTFL1.III coding sequence. In these experiments plants of Arabidopsis thaliana were grown on soil in long days (16 h light/8 hours dark) under a mixture of cool and warm white fluorescent light at 23°C and 65% humidity.

Coding sequence of LuTFL1.III allele was amplified and introduced into a Sma1-digested Gateway entry vector, pLBlue[rev], as a blunt-end fragment. The resulting entry vector, propagated in Escherichia coli DH5α strain, was then used in a recombination reaction with a modified Gateway recombination cassette pGREEN-HIS2. This recombination reaction effectively placed the LuTFL1.III allele behind the constitutive CaMV 35 S promoter in the pGREEN vector conferring resistance to BASTA. Sequence of PCR products and subsequent plasmid vectors were checked by Sanger sequencing and compared to the expected input sequence from the cDNA and vector backbone. The expression construct was introduced into A. thaliana (tfl1–2 mutants by Agrobacterium tumefaciens-mediated transformation). Transformed seeds were stratified and selected in soil with BASTA herbicide. Seeds from a single line with resistance to BASTA were collected.
Second generation of 35 S::LuTFL1 III tfl1–2 transformants was sown together with wild type Ler-0 and tfl1–2 mutant plants for phenotyping. We have measured flowering time as days to emergence of floral meristem and days to opening of first flower. Additionally, we measured days to end of flowering and plant height at the end of flowering. We have selected random 32 individuals for transgene genotyping. The difference in mean phenotypes for 35 S::LuTFL1 III, Ler-0 and tfl1–2 was tested using t-test in R.

Growth models under different LuTFL1 expression levels simulated using in-house R script pgrowth, for details see Supplementary Methods.

References
1. Purugganan, M. D. & Fuller, D. Q. The nature of selection during plant domestication. Nature 457, 843–848 (2009).
2. Zohary, D., Hopf, M. & Weiss, E. Domestication of Plants in the Old World: The Origin and Spread of Domesticated Plants in South-West Asia, Europe, and the Mediterranean Basin, 4th Edition (2012).
3. Colledge, S., Conolly, J. & Shennan, S. The Evolution of Neolithic Farming from SW Asian Origins to NW European Limits. *European Journal of Archaeology* https://doi.org/10.1177/14619577105060697 (2010).
4. Shennan, S. et al. Regional population collapse followed initial agriculture booms in mid-Holocene Europe. Nat. Commun. 4, 2486 (2013).
5. Timpson, A. et al. Reconstructing regional population fluctuations in the European Neolithic using radiocarbon dates: a new case-study using an improved method. J. Archaeol. Sci. 52, 549–557 (2014).
6. Allaby, R. G., Kitchen, J. L. & Fuller, D. Q. Surprisingly Low Limits of Selection in Plant Domestication. *Evol. Bioinform. Online* 11, 41–51 (2015).
7. Worland, A. J. et al. The influence of photoperiod genes on the adaptability of European winter wheats. *Euphytica* 100, 385–394 (1998).
8. Turner, A., Beales, J., Faure, S., Dunford, R. P. & Laurie, D. A. The pseudo-response regulator Ppd-H1 provides adaptation to photoperiod in barley. *Science* 310, 1031–1034 (2005).
9. Jones, H. et al. Population-based resequencing reveals that the flowering time adaptation of cultivated barley originated east of the fertile crescent. *Mol. Biol. Evol.* 25, 2211–2219 (2008).
10. Comadran, J. et al. Natural variation in a homolog of Antirrhinum CENTRORADIALIS contributed to spring growth and environmental adaptation in cultivated barley. *Nat. Genet.* 44, 1388–1392 (2012).
11. Diaz, A., Zuikhali, M., Turner, A. S., Isaac, P. & Laurie, D. A. Cytotype number variation affecting the photoperiod-B1 and vernalization-A1 genes is associated with altered flowering time in wheat (Triticum aestivum). *PLoS One* 7 (2012).
12. Takenaka, S. & Kawahara, T. Evolution and dispersal of emmer wheat (Triticum sp.) from novel haplotypes of Ppd-1 (photoperiod response) genes and their surrounding DNA sequences. *Theor. Appl. Genet.* 125, 999–1014 (2012).
13. Darapuneni, M. K., Morgan, G. D., Ibrahim, A. M. H. & Duncan, R. W. Effect of vernalization and photoperiod on flax flowering time. *Euphytica* 195, 279–285 (2014).
14. Uysal, H., Kurt, O., Fu, Y.-B., Diederichsen, A. & Kusters, P. Variation in phenotypic characters of pale flax (Linum bienne Mill.) from Turkey. *Genet. Resour. Crop Evol.* 59, 19–30 (2012).
15. Uysal, H. et al. Genetic diversity of cultivated flax (Linum usitatissimum L.) and its wild progenitor pale flax (Linum bienne Mill.) as revealed by ISSR markers. *Genet. Resour. Crop Evol.* 57, 1109–1119 (2010).
16. Dillman, A. C. Natural crossing in flax. *Journal of the American Society of Agronomy* 30 (1938).
17. Diederichsen, A. & Hammer, K. Variation of cultivated flax (Linum usitatissimum L. subsp. usitatissimum) and its wild progenitor pale flax (subsp. angustifolium (Huds) Thell). *Genet. Resour. Crop Evol.* 42, 263–272 (1995).
18. Fu, Y.-B. & Allaby, R. G. Phylogenetic network of Linum species as revealed by non-coding chloroplast DNA sequences. *Genet. Resour. Crop Evol.* 57, 667–677 (2010).
19. Herbig, C. & Maier, U. Flax for oil or fibre? Morphometric analysis of flax seeds and new aspects of flax cultivation in Late Neolithic wetland settlements in southwest Germany. *Veg. Hist. Archaeobot.* 20, 527–533 (2011).
20. Leuzinger, U. & Rast-Eicher, A. Flax processing in the Neolithic and Bronze Age pile-dwelling settlements of eastern Switzerland. *Veg. Hist. Archaeobot.* 20, 535–542 (2011).
21. Maier, U. & Schlichterhe, H. Flax cultivation and textile production in Neolithic wetland settlements on Lake Constance and in Upper Swabia (south-west Germany). *Veg. Hist. Archaeobot.* 20, 567–578 (2011).
22. Kelapa, W. & Danert, S. Zur systematik von Linum usitatissimum L. *Kulturpflanze* 3, 341–388 (1962).
23. Diederichsen, A. & Raney, J. P. Seed colour, seed weight and seed oil content in Linum usitatissimum accessions held by Plant Gene Resources of Canada. *Plant Breed.* 125, 372–377 (2006).
24. Diederichsen, A. & Fu, Y. B. Phenotypic and molecular (RAPD) differentiation of four infraspecific groups of cultivated flax (Linum usitatissimum L. subsp. usitatissimum). *Genet. Resour. Crop Evol.* 53, 77–90 (2006).
25. Soto-Cerda, B. J., Diederichsen, A., Raguinay, R. & Cloutier, S. Genetic characterization of a core collection of flax (Linum usitatissimum L.) suitable for association mapping studies and evidence of divergent selection between fiber and linsed types. *BMC Plant Biol.* 13, 78 (2013).
26. Prusinkiewicz, P., Erasmus, Y., Lane, B., Harder, L. D. & Coen, E. Evolution and development of inflorescence architectures. *Science* 316, 1452–1456 (2007).
27. Karligren, A. et al. Evolution of the PEPP Gene Family in Plants: Functional Diversification in Seed Plant Evolution. *Plant Physiol.* 156, 1967–1977 (2011).
28. Igasaki, T., Watanabe, Y., Nishiguchi, M. & Kotsoda, N. The flowering locus t/terminal flower 1 family in Lombardy poplar. *Plant Cell Physiol.* 49, 291–300 (2008).
29. Krieger, U., Lippman, Z. B. & Zamir, D. The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato. *Nat. Genet.* 42, 459–463 (2010).
30. Shalit, A. et al. The flowering hormone florigen functions as a general systemic regulator of growth and termination. *Proc. Natl. Acad. Sci. USA* 106, 8392–8397 (2009).
31. Lifschitz, E. et al. The tomato FT ortholog triggers systemic signals that regulate growth and flowering and substitute for diverse environmental stimuli. *Proc. Natl. Acad. Sci. USA* 103, 6838–6843 (2006).
32. Wang, X. et al. The genome of flax (Linum usitatissimum) assembled de novo from short shotgun sequence reads. *Plant J.* 72, 461–473 (2012).
33. Baird, N. A. et al. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3 (2008).
34. Ettler, P. D., Basham, S., Hohenlohe, P. L. A., Johnson, E. A. & Cresko, W. A. SNP discovery and genotyping for evolutionary genetics using RAD sequencing, in Molecular Methods for Evolutionary Genetics (ed. Rockman, M. V.) 157–178 (Humana Press 2011).
35. Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W. & Postlethwait, J. H. Stacks: building and genotyping loci de novo from short-read sequences. *G3-Genes Genomes Genetics* 1, 171–182 (2011).
36. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664 (2009).
37. Fu, Y.-B., Diederichsen, A. & Allaby, R. G. Locus-specific view of flax domestication history. *Ecol. Evol.* 2, 139–152 (2012).
