Adaptation to European latitudes through assimilation of wild diversity at the *LuTFL1* locus altered architecture and promoted fiber production in flax.

Rafal M Gutaker[1,2], Maricris Zaidem[2], Yong-Bi Fu[3] , Axel Diederichsen[3], Oliver Smith[1,4] , Roselyn Ware[1], Robin G Allaby[1]*

1. School of Life Sciences, University of Warwick, United Kingdom CV4 7AL

2. Max Planck Institute for Developmental Biology, Tübingen, Germany

3. Plant Gene Resources of Canada, Saskatoon Research and Development Centre, SK S7N 0X2, Canada

4. Centre for Geogenetics, Copenhagen University of Copenhagen, 1307 Copenhagen K, Denmark

Corresponding author r.g.allaby@warwick.ac.uk

Abstract

After domestication in the Near East around 10,000 years ago several founder crops spread and adapted to European latitudes. On reaching northerly latitudes the architecture of domesticated flax changed becoming 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. We identify PEBP family genes in the flax genome, describe molecular diversity of two loci, *LuTFL1* and *LuTFL2* in wild and cultivated flax that vary over latitudinal range, and show that cultivated flax received *LuTFL1* alleles from more northerly wild flax populations. Compared to a genomic background of population structure of flaxes over latitude determined using restriction site-associated loci, the *LuTFL1* alleles display a level of differentiation that is consistent with selection both for a northern associated allele (III) and against a southern associated allele (I) in the north. We also demonstrate that *LuTFL1* alleles are associated with differing flowering time, and through heterologous expression in *Arabidopsis thaliana* that *LuTFL1* is a functional homolog of *TFL1* in *Arabidopsis thaliana* and is capable of changing both flowering time and plant architecture. We conclude that specialized fiber flax types formed as a consequence of a natural adaptation of cultivated flax to higher latitudes.

Significance statement

It is usually assumed that domesticated plants acquire traits that are useful to humans alone. The unusual twist in this study is that the domesticated plant involved, flax, has an adaptation under domestication that serves both the cultivators and the plant , resulting in smaller seed size. In this case, the adaptation could have opened the way for the ensuing textile revolution as flax became predisposed for fiber production.

Introduction

Cultivated flax (*Linum usitatissimum* L.) was domesticated from pale flax (*L. bienne* Mill.) (1) around 12,000 years ago in the Near East (2, 3). Flax was part of a 'Neolithic Package' (4) that spread out of the Near East into Europe (5). Originally domesticated for its oily seeds (6), flax is currently grown as either an oil or fiber crop. Specialized fiber varieties likely emerged during the Neolithic 'Textile Revolution' (7–9), which started in Central Europe about 6,000 years Before Present (BP). The two flax forms are different in their seed size and distinct plant architecture (10). In particular, specialized fiber varieties are characterized with lower seed weight, fewer axillary branches and higher technical stem length, relative to oil varieties (11, 12).

The environmental conditions associated with the higher latitudes of Europe presented the plants of Near Eastern origin with major challenges that have been associated with repeated agricultural collapse (13, 14), possibly due to the limited rates at which adaptation could take place (15). Long-day plants would have reached threshold day lengths earlier in the year as photoperiods extend with latitude, which combined with a later onset of mild conditions in the Spring would have resulted in a shortening of the growing season for crops. Typically, wheat and barley adapted to these conditions by reducing sensitivity to environmental cues. Such spring varieties utilize more of the northern growing season through photoperiod insensitivity which affects their flowering time (16-21). Some northern flax types also show similar loss of photoperiodical sensitivity phenotypes (22). However, in contrast to the Neolithic cereals, flax had a potential advantage in that it could have assimilated adaptations from its wild progenitor species, which extends in range to northwestern Europe and shows patterns of extensive local adaptation (23, 24).

While change in flowering time is an important factor in plant adaptation to increasing latitudes in general, some flowering time genes can incidentally also affect architecture (25). In *Arabidopsis thaliana*, both flowering time and plant architecture can be simultaneously modified by changes in regulation of the *FT* and *TFL1* genes (25), which belong to the phosphatidyl ethanolamine-binding protein (PEBP) family (26) and control plant development through regulation of floral meristem initiation. An allelic variant of a PEBP gene in barley, *HvCET* has been involved in

adapting this crop to the European climate through later flowering, suggesting the potential importance of this family in latitudinal adaptations of crops during the spread of agriculture (17). Furthermore, *FT* and *TFL1* can control plant stature through promotion or delaying floral initiation (25) and are highly conserved in eudicots (27). The orthologs of *FT* and *TFL1* in tomato have a strong effect on tomato fruit weight and number (28), plant architecture and height, and flowering time (29, 30). 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 could be associated with both architectural and flowering time changes in flax.

In this study we genotyped restriction site-associated regions of flax genome to examine population structure in European wild and cultivated flax and address the possibility that cultivated flax may have interacted with pale flax as it spread through Europe. To see if contact between two species might have been beneficial to cultivated flax adaptation we investigated PEBP family genes that are associated with flowering time and architectural changes. These efforts revealed that flax adapted to the European variable climate through assimilation of allele diversity at the *LuFTL1* locus to alter plant architecture and promote fiber production in flax.

Results

Population structure of wild and pale flax. We investigated the population structure of wild and cultivated flaxes through a restriction site-associated (RAD) sequencing approach (31, 32). Analyses were restricted to 2210 regions that flank SbfI cut sites in the reference genome (33). We surveyed 90 accessions of pale (28) and cultivated (62) flax (Table S1), and found 1695 polymorphic RAD tags present in at least 80% accessions, which encompass the total of 18594 depth filtered SNPs. We found that the pale flax is more diverse than cultivated flax, while landraces have intermediate levels of nucleotide diversity (Table S2). Specialized varieties have a lower diversity within both modern and historic cultivars, which suggests a reduction during specialization, particularly in fiber varieties. In a multidimensional scaling analysis cultivated flax and dehiscent flax appear to cluster closely, while pale flax shows a wide spread, (Figure S1).

We identified four distinct ancestral components using an ADMIXTURE (34) approach (Figure S2). The pale flax of Eastern Anatolia include all the admixture components that are present in cultivated flaxes, including dehiscent varieties (Figure 1), which supports an eastern population involvement with the original domestication process. Two pale flaxes located in eastern Greece outside the domestication center show admixture only with fully indehiscent varieties. This is congruent with previous reports that identify dehiscent varieties as genetically distinct from other domesticated flax varieties, possibly reflecting an earlier independent domestication (35, 36). A latitudinal gradient is apparent in the cultivated fiber flax types, landraces and intermediate forms, evident from a tendency for southern populations to be affiliated to a brown grouping, and a yellow grouping in northern populations in Figure 1. Interestingly, we see no such latitudinal structure in oil flax types.

Northern and southern flax population differentiation. Given the latitudinal cline apparent in the fiber varieties, we considered populations as 'northern' or 'southern' relative to 40th circle of latitude of northern hemisphere. This distinction is based on the origins of accessions either side of the Bosphorus strait, in order to gain insight into how northern and southern cultivated flax populations may have differentiated. An allele frequency spectrum based on all variant call SNPs to the mapping genome indicates an asymmetric distribution of allele frequencies between northern and southern populations, Figure S3. Southern populations have an excess of low to intermediate frequencies higher than the northern populations (36% of all SNPs), while northern populations are significantly overrepresented by higher frequency alleles when the mapping reference state is also used.(55% of all SNPs). This could be due to the relative similarity between the mapping genome and the northern populations. To overcome such imperfections of the mapping approach we generated 1684 SNPs based on *de novo* assembled RAD loci, which provide an unbiased allele frequency spectrum giving a null distribution of population differentiation, Figure 2. We did not find any change to the estimated population structure with this unbiased data set.

We investigated whether the RAD loci data harboured evidence of movement of alleles between wild and cultivated flax populations beyond the area of domestication. We observed no clear

signal using f_3 statistics (cultivated; cultivatedS, wildN) (37), however, varying rates of population differentiation through founder effects and multiple complex processes of gene flow can mask signals of gene movement between populations (37). 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 wild populations, Figure S4, S5, S6 and Table S3. These findings are verified by an increased ancestral information content (I_a) (38) at loci that were differentially high in northern but not southern wild populations, Table S3. 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 had elevated frequencies relative to the background distribution, which would require an explanation of selection being responsible more frequently than drift for high frequency alleles in the north.

PEBP orthologs in flax. We investigated whether potential contact between wild and cultivated flax populations might involve functionally significant loci by surveying PEBP genes, which have previously been shown to be involved in latitudinal adaptation in other crops (17). A scan for homologs of *A. thaliana* PEBP genes in the flax assembly (33) revealed eight loci of interest with architectures similar to *TFL1* gene (Figure S7; Table S4). Out of those, seven loci were amplified and sequenced from flax genomic DNA (Figure S8), four could be easily re-sequenced without cloning and showed conserved four-exon structure (Figure S7). Phylogenetic analyses comparing the flax, *A. thaliana* and Lombardy poplar orthologs (Table S2) 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 (Figure S9). Preliminary resequencing of those genes in six accessions of flax indicated high levels of polymorphism in pale flax at the *LuTFL1* locus (Table S5), suggestive of broad standing variation.

Diversity in *LuTFL1* and *LuTFL2*. We re-sequenced the *LuTFL1* locus in 32 wild and 113 cultivated flax accessions (Table S1). 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 (Table S6). A phylogenetic network of *LuTFL1* shows that allelic diversity is greater in pale flax (top plane, Figure 3A) when compared to cultivated flax (bottom plane, Figure 3A). 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 post-domestication diversification (alleles XI, XII, XIII, XIV, XV and XVI). Interestingly, the *LuTFL2* network shows greater diversity in cultivated flax allele compared to pale flax (Figure S11A, Table S6).

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 (Figure 3C, Table S7), 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 (Figure 3B). 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* (Tables S8), shows a similar latitudinal gradient (Figure S11B), 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 (Figure 3B) are more evident in the historic landraces (Figure S12B). 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. Enrichment of alleles *LuTFL1* III, *LuTFL1* XII and *LuTFL2* X 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.

Signature of selection in *LuTFL*. 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_2 and F methods (Table S9). 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 demographic effects such as expansion, or selection. To distinguish between these possibilities, we compared the differentiation of the *LuTFL1* alleles in northern and southern cultivated flax populations to the null distribution of differentiation obtained from RAD loci in two ways. Firstly, we compared fixation index statistics, Figure S13. We found that the Fst values associated with the differences between *LuTFL1* alleles I and III are significantly higher than the background range of values based on RAD loci. Secondly, we placed the position of the *LuTFL1* alleles on the allele frequency spectrum of RAD loci, and considered the relative deviation of allele frequency from the diagonal (general), horizontal (with respect to northern populations) and vertical (with respect to southern populations), Figure 2. We found that both alleles I and III lie significantly outside the background range in general (p_d Figure 2B, Table S10), but close to the edges of the null distribution, whereas allele XII is within the general range. In the case of allele III, the frequency is weakly lower than expected in the southern populations given the northern frequency (p_v, Figure 2C, Table S10), but strongly higher than expected in the northern population given the southern frequency (p_h, Figure 2D). Allele I is very significantly higher in the southern population given the northern frequency (p_v), and significantly lower than expected in the northern population given the southern frequency (p_h). Together, these data indicate that allele I was selected against and 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 time (39), using archaeological dates of the arrival of agriculture at different latitudes to date the latitudinal allele frequencies, Figure S12A. Our estimates confirmed weak selection with 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.

Functional homology of *LuTFL1*. 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 (Figure S14B). 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.

Many *TFL* family genes that influence flowering time by acting antagonistically with *FT* to delay flowering (40-42) have been identified as targets for selection in adaptation to northern latitudes (17, 43) and have also been shown to influence plant architecture (25, 29, 30) and fruit yield (28). 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 (Figure S15; Table S11) consistent with *TFL1* expression in *A. thaliana* (44) as opposed to *ATC*, which is expressed in the hypocotyl in *A. thaliana* (45). We detected no expression of *LuTFL2* in leaf tissue. Based on phylogenetic evidence, *LuTFL1* is more similar to *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 35S promoter. We obtained a stable line after transforming a *tfl1-2* mutant in Landsberg *erecta* (Ler) genomic background (46) using an *Agrobacterium* system. Over 160 plants of 35S::*LuTFL1* inbred progeny (T2) were grown in long day conditions and phenotyped. Flowering time phenotypes segregated in T2 with a Mendelian 15 to 1 ratio (Figure S16), 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, Figure S16. Plants that contained the 35S::*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 35S::*LuTFL1* plants (Figure 4; Figure S17; Table S12), suggesting that the flax *LuTFL1* I allele can rescue non-functional *TFL1* in *A. thaliana*. Constitutive expression of this allele under 35S 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 (Figure 4; Figure S17). 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.

To investigate the effect of *LuTFL1* allelic variation in flax we measured the time to flowering in our pale flax populations and compared this to phenotypic data we had previously generated for cultivated flaxes (47) (Figure S18). In cultivated flaxes we found a clear correlation between flowering time and allele type, with *LuTFL1* allele I flowering earlier (mean 190 days) and allele III flowering later (mean 194 days; t-test p-value = 0.0016). Similarly, allele I plants are associated with shorter stem height (mean 60 cm) compared to carriers of allele III (mean 78 cm; t-test p-value = 0.0012). Finally, *LuTFL1*.III carriers are less branched, satisfying the expected features of fiber varieties (t-test p-value = 0.0014). Pale flaxes reflect our findings with allelic association of flowering time, with alleles I and III associated with earlier and later flowering respectively (t-test p-value = $2.2*10^{-16}$). However, if allele II is included in comparison due to its close relation to allele I, the flowering time (mean 98 days) is not significantly different from plants with allele III (mean 102

days; t-test p-value = 0.073). We also observed marginally significant correlation between *LuTFL2* allele type and flowering (mean for VI = 190 days, mean for X =193 days; t-test p-value = 0.047; Figure S19), which might be due to linkage disequilibrium between allele *LuTFL1*.III and *LuTFL2*.X (Table S8). The co-occurrence of *LuTFL1* allele III, late flowering and tall plant stature in a pale and cultivated flax genome background together with experimental data from transgenic 35S::*LuTFL1 A. thaliana* plants support the notion that the correlation between allele type and phenotype could have a causative relationship.

A model of latitudinal adaptation. In general, in cultivated flaxes flowering time is correlated positively with traits of taller plants, reduction of branch number and enhanced fiber content, and negatively with seed mass (Table S13; Figure S20). Within our samples we additionally observe that flowering time, plant height and branching pattern are associated with allelic diversity at the *LuTFL1* locus. Finally, we provide evidence for functional homology of *LuTFL1* to *A. thaliana TFL1*, which is known for its effect on phenology and architecture in multiple plant species (25, 30, 48). 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 latitudinal movement could affect stem height, which correlates with increased fiber content (47). 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 (Figures S21, S22). 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. In line with this prediction, *LuTFL1* group III alleles are associated with both later flowering time and improved fiber content through a heightened stem length (Figure S18).

Discussion

This study 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 (49), later followed by diversification or improvement traits (50). 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 (51-53) and within crop plants not all adaptations are of obvious benefit to agronomic productivity and yield, but related to the wider ecology (54). Examples include a tendency to increase rates of water loss in domesticate forms (55). 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. Previous to reaching Central European latitudes flax had already been used for both oil and fiber. The spread of agriculture stalled at the latitudes around the Balkan regions during a period of adaptation either to community structures or the physical environment (5, 56). This study shows that the adaptation of flax to the European variable climate during Neolithic likely involved the *LuFTL1* locus and resulted in a substantial modification of the stem height, which is correlated with the technical fiber quality. This evolutionary transition may have been key to the rise of the textile revolution in central Europe (9), which is evidenced in the reduction of flax seed size (7) and improved tools for fiber production (8) in Neolithic contexts.

It is notable that such a 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 (57). 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

All methods and materials used are reported in the supplementary information.

Acknowledgments

We would like to acknowledge Sabine Karg, for her help in clearing the archaeological background for this study and help in obtaining historic samples from Vavilov Institute, Sankt Petersburg. We would like to thank Nina Brutch (Vavilov Institute) and Dallas Kessler (PGRC) for dispatching seeds from seed banks. Toni Nikolic and Arne Strid helped in locating pale flax populations in the Balkans. Mariusz Czarnocki-Cieciura, Agnieszka Cakala and Ewa Samorzewska helped during pale flax collection. Finally, we would like to thank Hernan Burbano and Detlef Weigel for support during revisions and functional analyses.

RMG was supported by University of Warwick Chancellor's Scholarship scheme, OS is supported by NERC (NE/L006847/ 1) and RW is supported by NERC (NE/F000391/1).

Sequence data has been entered into GenBank: LuTFL1 (KU240116 - KU240259) LuTFL2 (KU240260 - KU240389), RADseq (PRJNA304385).

Figure Legends

Figure 1 Population structure of pale flax and varieties of cultivated flax based on RADseq data. Samples of flax were arranged in major groups: Wild – pale flax, D – dehiscent varieties of cultivated flax, Oil - oil varieties, Fiber - fiber varieties, Int. - intermediate varieties and Land – landraces. Pale flax samples are arranged from east (Eastern Anatolia) to northwest (Central Europe), while cultivated samples are arranged by increasing latitude within groups.

Figure 2. Allele frequency spectrum of RAD loci alleles between northern and southern cultivated flaxes. Frequencies of *LuTFL1* alleles are marked as circles, I (purple), III (black) and XII (grey). A. Allele frequency spectrum. Panels B-D: probability heat maps of sampling frequency combinations given the allele frequency spectrum. B. Probabilities based on distance from the diagonal. C. Probabilities based on row frequency distribution. D. Probabilities based on column frequency

distribution.

Figure 3 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).

Figure 4 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 35S::*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 35S::*LuTFL1* plants flowered until 36th and 40th day reaching 134 and 211 mm height respectively.

References

1. Diederichsen A, Hammer K (1995) Variation of cultivated flax (Linum usitatissimum L. subsp. usitatissimum) and its wild progenitor pale flax (subsp. angustifolium (Huds) Thell). *Genet Resour Crop Evol* 42(3):263–272.

2. Hillman GC (1975) The plant remains from Tell Abu Hureyra: a preliminary report. *Proceedings of the Prehistoric Society* 41:70–73.

3. Van Zeist W, Bakker-Heeres JAH (1975) Evidence for linseed cultivation before 6000 BC.

Journal of Archaeological Sciences 2:215–219.

4. Zohary D, Hopf M, Weiss E (2012) *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*.

5. Colledge S, Conolly J, Shennan S (2016) The Evolution of Neolithic Farming from SW Asian Origins to NW European Limits. *European Journal of Archaeology*. doi:10.1177/1461957105066937.

6. Fu YB, Allaby RG (2010) Phylogenetic network of Linum species as revealed by non-coding chloroplast DNA sequences. *Genet Resour Crop Evol* 57(5):667–677.

7. Herbig C, Maier U (2011) 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(6):527–533.

8. Maier U, Schlichtherle H (2011) Flax cultivation and textile production in Neolithic wetland settlements on Lake Constance and in Upper Swabia (south-west Germany). *Veg Hist Archaeobot* 20(6):567–578.

9. Leuzinger U, Rast-Eicher A (2011) Flax processing in the Neolithic and Bronze Age pile-dwelling settlements of eastern Switzerland. *Veg Hist Archaeobot* 20(6):535–542.

10. Kulpa W, Danert S (1962) Zur systematik von Linum usitatissimum L. *Kulturpflanze* 3:341–388.

11. Diederichsen A, Fu YB (2006) Phenotypic and molecular (RAPD) differentiation of four infraspecific groups of cultivated flax (Linum usitatissimum L. subsp. usitatissimum). *Genet Resour Crop Evol* 53(1):77–90.

12. Diederichsen A, Raney JP (2006) Seed colour, seed weight and seed oil content in Linum usitatissimum accessions held by Plant Gene Resources of Canada. *Plant Breed* 125(4):372–

377.

13. Shennan, S., Downey, S.S., Timpson, A., Edinborough, K., Kerig, T., Manning, K., Thomas, M.G. (2013) Regional population collapse followed initial agricultural booms in mid –Holocene Europe. *Nat. Comms.* 4:2486.

14. Timpson A. *et al.* (2014) 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.

15. Allaby, R.G., Kitchen, J.L., Fuller, D.Q. (2015) Surprisingly low limits of selection in plant domestication. *Evolutionary Bioninformatics* 11 (S2), 41-51.

16. Jones H, et al. (2008) Population-based resequencing reveals that the flowering time adaptation of cultivated barley originated east of the fertile crescent. *Mol Biol Evol* 25(10):2211–2219.

17 Comadran J, et al. (2012) Natural variation in a homolog of Antirrhinum CENTRORADIALIS contributed to spring growth habit and environmental adaptation in cultivated barley. *Nat Genet* 44(12):1388–1392.

18 Turner A, Beales J, Faure S, Dunford RP, Laurie DA (2005) The pseudo-response regulator Ppd-H1 provides adaptation to photoperiod in barley. *Science* 310(5750):1031–1034.

19. Worland AJ, et al. (1998) The influence of photoperiod genes on the adaptability of European winter wheats. *Euphytica* 100(1-3):385–394.

20. Diaz A, Zikhali M, Turner AS, Isaac P, Laurie DA (2012) Copy number variation affecting the photoperiod-B1 and vernalization-A1 genes is associated with altered flowering time in wheat (Triticum aestivum). *PLoS One* 7(3). doi:10.1371/journal.pone.0033234.

21. Takenaka S, Kawahara T (2012) 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(5):999–1014.

22. Darapuneni MK, Morgan GD, Ibrahim AMH, Duncan RW (2014) Effect of vernalization and

photoperiod on flax flowering time. *Euphytica* 195(2):279–285.

23. Uysal, H., Kurt, O., Fu, Y.B., Diederichsen, A. & Kusters, P. (2012) Variation in phenotypic characters of pale flax (Linum bienne Mill.) from Turkey. *Genetic Resources and Crop Evolution* **59**, 19-30.

24. Uysal, H. *et al.* (2010) Genetic diversity of cultivated flax (*Linum usitatissimum* L.) and its wild progenitor pale flax (*Linum bienne* Mill.) as revealed by ISSR markers. *Genetic Resources and Crop Evolution* **57**, 1109-1119

25. Prusinkiewicz P, Erasmus Y, Lane B, Harder LD, Coen E (2007) Evolution and development of inflorescence architectures. *Science* 316(5830):1452–1456.

26. Karlgren A, et al. (2011) Evolution of the PEBP Gene Family in Plants: Functional Diversification in Seed Plant Evolution. *Plant Physiol* 156(4):1967–1977.

27. Igasaki T, Watanabe Y, Nishiguchi M, Kotoda N (2008) The FLOWERING LOCUS T/TERMINAL FLOWER 1 family in Lombardy poplar. *Plant Cell Physiol* 49(3):291–300.

28. Krieger U, Lippman ZB, Zamir D (2010) The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato. *Nat Genet* 42(5):459–463.

29. Lifschitz E, et al. (2006) The tomato FT ortholog triggers systemic signals that regulate growth and flowering and substitute for diverse environmental stimuli. *Proc Natl Acad Sci U S A* 103(16):6398–6403.

30. Shalit A, et al. (2009) The flowering hormone florigen functions as a general systemic regulator of growth and termination. *Proc Natl Acad Sci U S A* 106(20):8392–8397.

31. Baird NA, et al. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3(10). doi:10.1371/journal.pone.0003376.

32. Etter PD, Bassham S, Hohenlohe P lA, Johnson EA, Cresko WA (2011) SNP discovery and genotyping for evolutionary genetics using RAD sequencing. *Molecular Methods for*

Evolutionary Genetics, Methods in Molecular Biology., ed Rockman MV (Humana Press, Orgogozo, Virginie), pp 157–178.

33. Wang Z, et al. (2012) The genome of flax (Linum usitatissimum) assembled de novo from short shotgun sequence reads. *Plant J* 72(3):461–473.

34. Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19(9):1655–1664.

35. Fu YB (2011) Genetic evidence for early flax domestication with capsular dehiscence. *Genet Resour Crop Evol* 58(8):1119–1128.

36. Fu YB, Diederichsen A, Allaby RG (2012) Locus-specific view of flax domestication history. *Ecol Evol* 2(1):139–152.

37. Patterson, N. *et al.* (2012) Ancient admixture in human history. *Genetics* **192**, 1065-1093.

38. Rosenberg, N.A., Li, L.M., Ward, R., and Pritchard, J.K.. (2003) Informativeness of genetic markers for inference of ancestry. Am. J. Hum. Genet. *73*, 1402–1422.

39. Allaby, R.G., Stevens, C., Lucas, L., Maeda O., Fuller, D.Q. (2017) Geographic mosaics and changing rates of cereal domestication. *Phil. Trans. Roy. Soc. B in press.*

40. Shannon S, Meekswagner DR (1991) A mutation in the Arabidopsis TFL1 gene affects inflorescence mersitem development. *Plant Cell* 3(9):877–892.

41. Bradley D, et al. (1996) Control of inflorescence architecture in Antirrhinum. *Nature* 379(6568):791–797.

42. Bradley D, Ratcliffe O, Vincent C, Carpenter R, Coen E (1997) Inflorescence commitment and architecture in Arabidopsis. *Science* 275(5296):80–83.

43. Tian Z, et al. (2010) Artificial selection for determinate growth habit in soybean. *Proc Natl Acad Sci U S A* 107(19):8563–8568.

44. Yoo SJ, et al. (2010) BROTHER OF FT AND TFL1 (BFT) has TFL1-like activity and

functions redundantly with TFL1 in inflorescence meristem development in Arabidopsis. *Plant J* 63(2):241–253.

45. Mimida N, et al. (2001) Functional divergence of the TFL1-like gene family in Arabidopsis revealed by characterization of a novel homologue. *Genes Cells* 6(4):327–336.

46. Alvarez J, Guli CL, Yu X-H, Smyth DR (1992) terminal flower: a gene affecting inflorescence development in Arabidopsis thaliana. *Plant J* 2(1):103–116.

47. Diederichsen A, Ulrich A (2009) Variability in stem fibre content and its association with other characteristics in 1177 flax (Linum usitatissimum L.) genebank accessions. *Ind Crops Prod* 30(1):33–39.

48. McGarry RC, Ayre BG (2012) Manipulating plant architecture with members of the CETS gene family. *Plant Sci* 188:71–81.

49. Harlan, J.R. et al. (1973) Comparative evolution of cereals. *Evolution* **27,** 311–325.

50. Meyer, R.S., Purugganan, M.D. (2013) Evolution of crop species: genetics of domestication and diversification. *Nature Rev. Genet.* **14,** 840-852.

51. Spahillari, M., Hammer, K., Gladis, T., Diederichsen, A., (1999) Weeds as part of agrobiodiversity. Outlook Agr. 28, 227e232.

52. Senda, T., Hiraoka, Y., Tominaga, T., (2006) Inheritance of seed shattering in Lolium temulentum and L. persicum hybrids. Genet. Resour. Crop Evol. 53, 449e451.

53. Howard, T., Archer, J.E., Turley, R.M., (2011) Evolution, physiology and phytochemistry of the psychotoxic arable mimic weed darnel (Lolium temulentum L.). Prog. Bot. 72, 73e104.

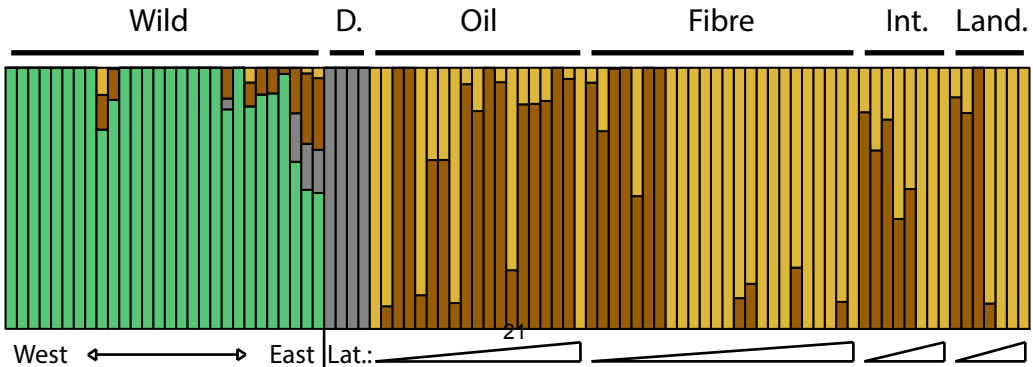
54. Milla, R., Osborne, C.P., Turcotte, M.M., Violle, C. (2015) Plant domestication through an ecological lens. *Trends Ecol. Evol.* **30,** 463-469.

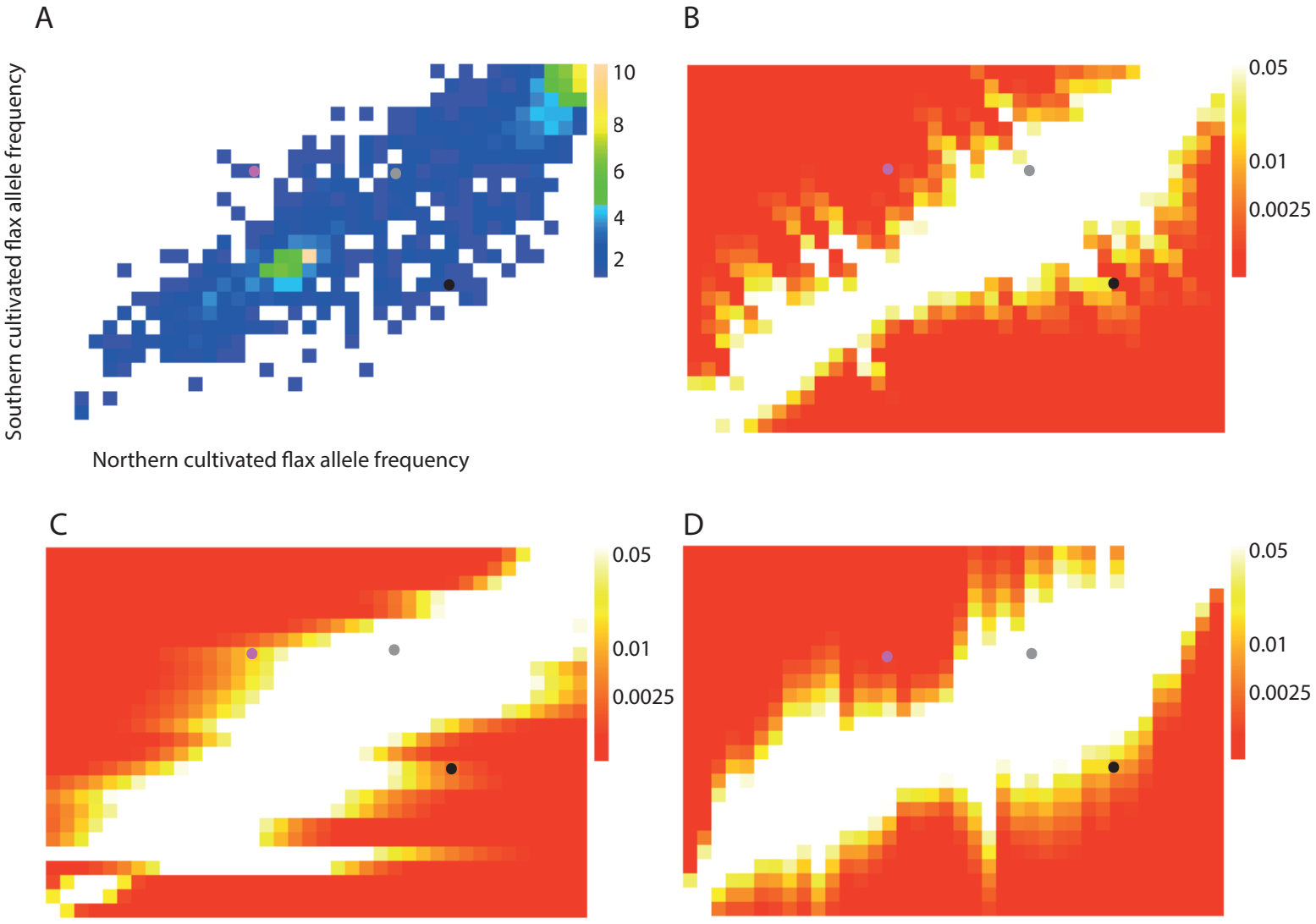
55. Milla, R. et al. (2013)Shifts in stomatal traits following the domestication of plant species. *J. Exp. Bot.* **64,** 3137–3146.

56. Coward F, Shennan S, Colledge S, Conolly J, Collard M (2008) The spread of Neolithic plant

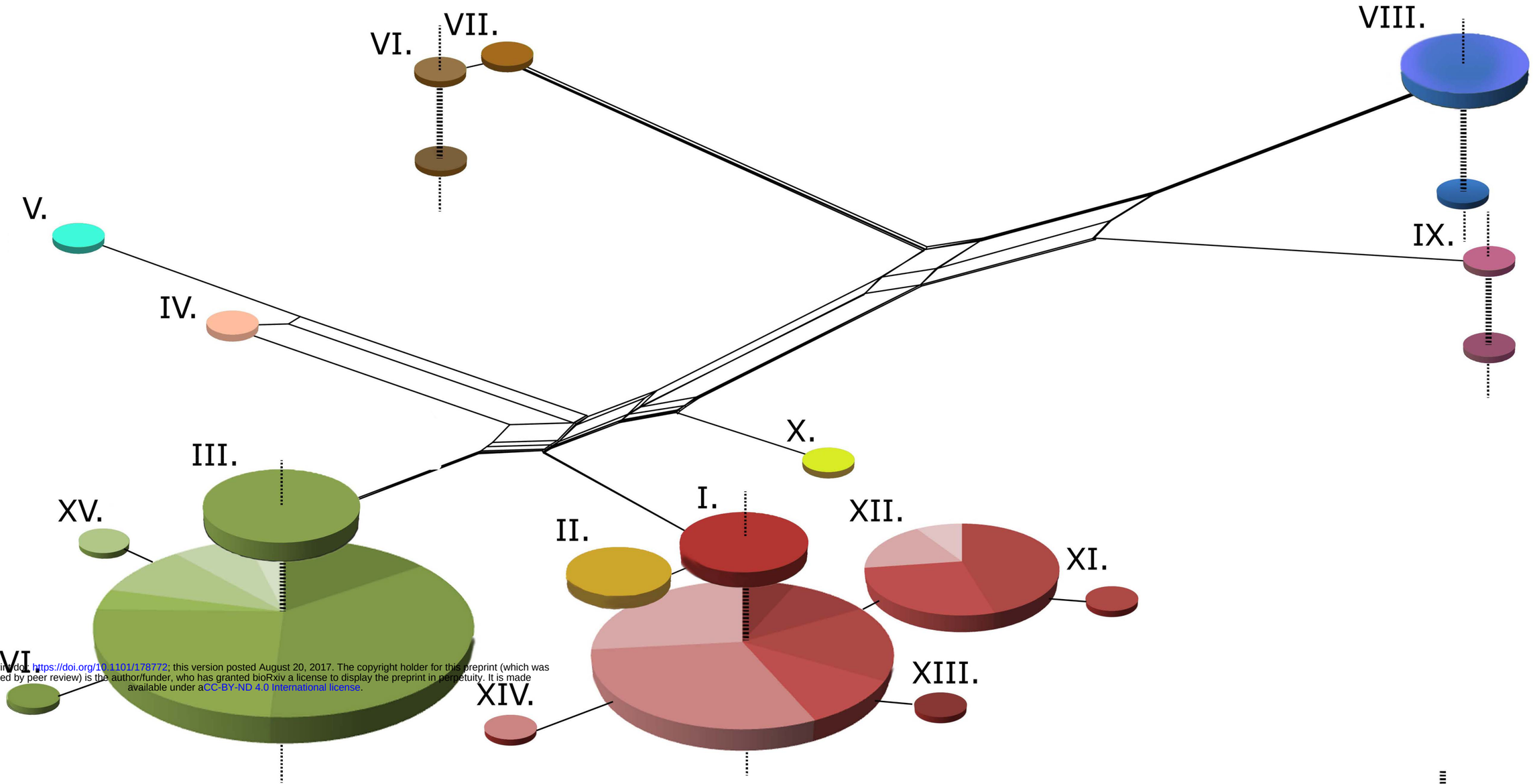
economies from the Near East to northwest Europe: a phylogenetic analysis. *J Archaeol Sci* 35(1):42–56.

57. Roffet-Salque M, Regert M, Evershed RP, Outram AK, Cramp LJE, Decavallas O, Dunne J, Gerbault P, Mileto S, Mirabaud S et al.. (2015) Widespread exploitation of the honeybee by early Neolithic farmers. *Nature* 527(7577): 226-230.

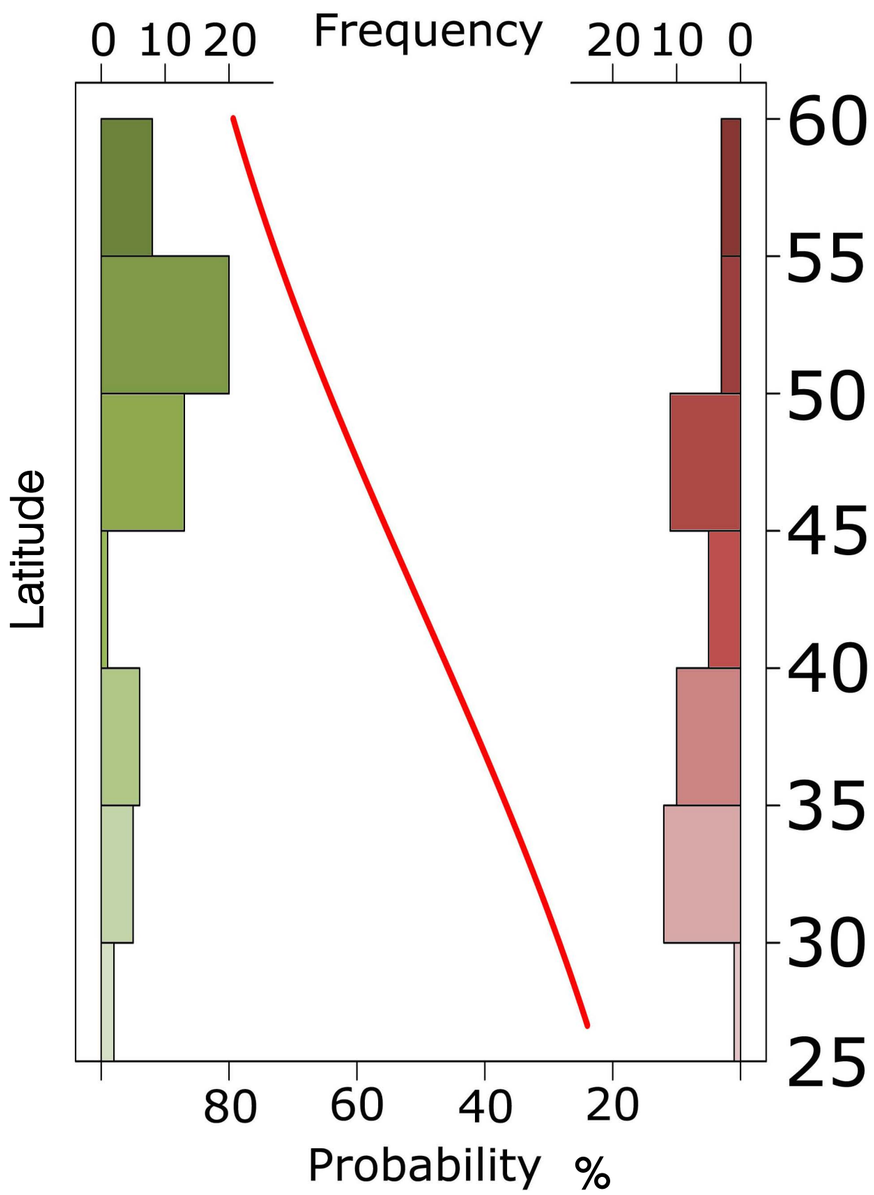

Wild
D.
Oil
Fibre
Int.
Land.
West
East
Lat.:
21



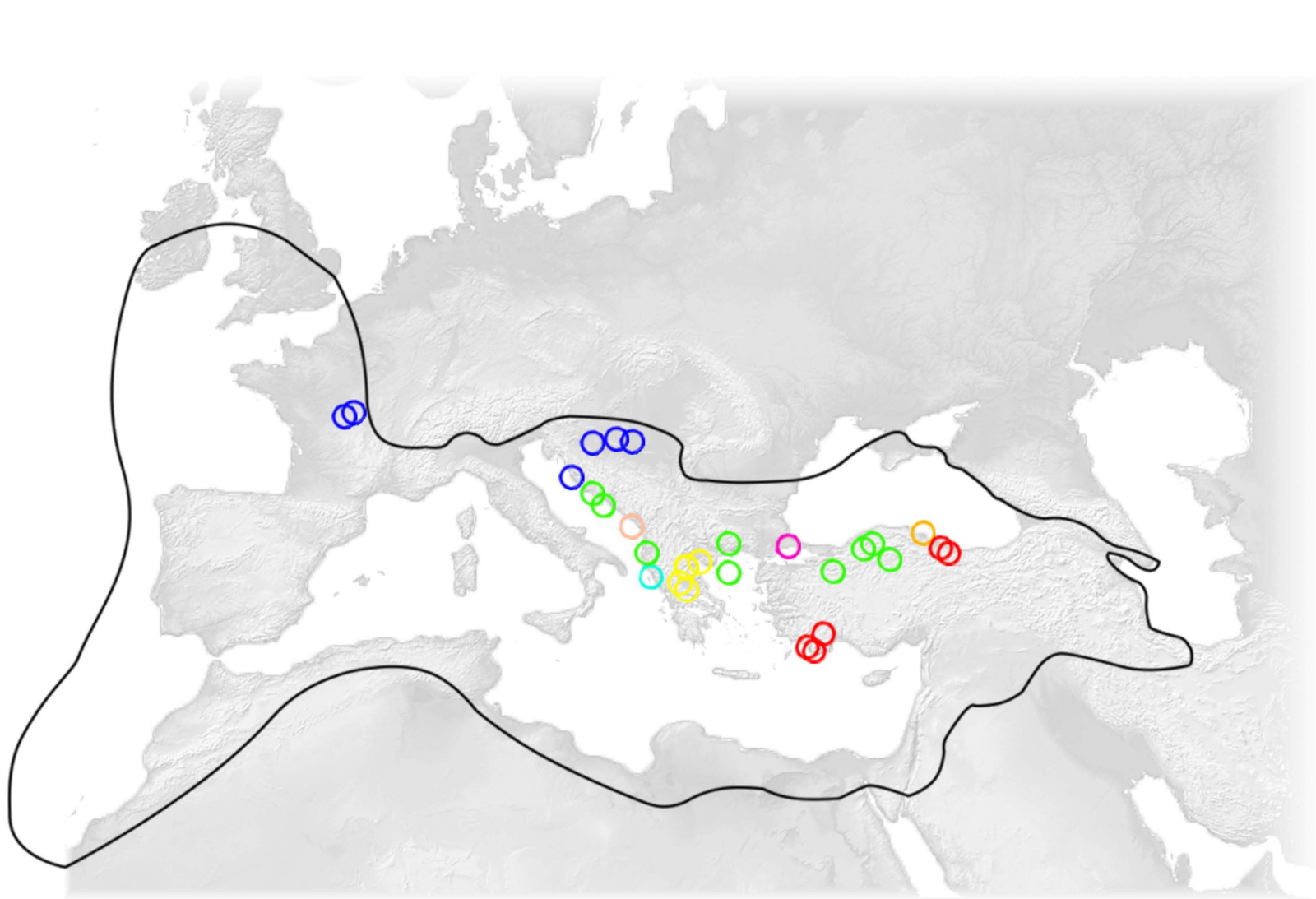
a



b



c




A
2x
B
1x
Ler-0
tfl1-2
35S::LuTFL1

Supplementary Information

Adaptation to European latitudes through assimilation of wild diversity at the *LuTFL1* locus altered architecture and promoted fiber production in flax.

Rafal M Gutaker[1,2], Maricris Zaidem[2], Yong-Bi Fu[3], Axel Diedrichsen[3], Oliver Smith[1,4], Roselyn Ware[1], Robin G Allaby[1*]

1. School of Life Sciences, University of Warwick, United Kingdom CV4 7AL
2. Max Planck Institute for Developmental Biology, Tübingen, Germany
3. Plant Gene Resources of Canada, Saskatoon Research and Development Centre, SK S7N 0X2, Canada
4. Centre for Geogenetics, Copenhagen University of Copenhagen, 1307 Copenhagen K, Denmark

Corresponding author r.g.allaby@warwick.ac.uk

1. Plant Materials

1.1 Pale Flax sample collection

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. Outlined in Table S1.

1.2. Cultivated sample collection

Seed material for 58 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. Outlined in Table S1.

1.3 Herbaria collections

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. Outlined in Table S1

2. RAD sequence analysis

2.1 Rationale of accessions used for RAD analysis

A subset of the flax accessions used in this study were used in the RADseq analysis (Table S1). In particular, for cultivated flaxes the sampling focused on historic samples collected by Vavilov during the early 20th century in order to establish phylogeographic patterns, which have not been disrupted by movements of modern crops subsequent to the Green Revolution.

2.2 RADseq library preparation

From the total collection, 90 plants were chosen for the RADseq experiment (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. Plant material was instantly frozen in liquid nitrogen and freeze-dried in -50°C for two days. Leaves were detached from stems and weighted. Between 10 and 15 mg of dried leaves was used for DNA extraction. Material was ground with 3 glass beads (3 mm diameter) in TissueLyser machine (Qiagen) for 1 min with 30 Hz shaking frequency. Powdered tissue was then used for DNA isolation with DNeasy® Plant Mini Kit (Qiagen) following the manufacturer's manual. In the final step, DNA solution was collected in 3 subsequent elution steps, each with 50 µl EB buffer. DNA concentration was measured with the Quant-it^{TM} dsDNA Broad-Range Assay kit (Invitrogen) in Qubit® fluorometer (Invitrogen).

For RADseq library preparation, 40 µl of genomic DNA at concentration of 25 ng/µl was prepared following previous methods[57]. DNA was digested with 2U of SbfI HF restriction enzyme (New England Biolabs) for one hour and then the reaction was heat inactivated. DNA fragments were ligated with 1 µl of custom-made, uniquely barcoded P1 adapters at 100 nM concentration for 90 min using T4 DNA ligase, rATP, NEB 2 buffer followed by heat inactivation. Samples 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 with following settings: 20% power, 20% pulses for 30 s, followed by 30 s break. Sheared DNA was purified using AMPure XP beads (Agencourt) in proportion 1 to 1 with buffer. Random DNA overhangs that resulted from sonication were blunted and phosphorylated using Quick Blunting Kit (NEB) followed by magnetic bead purification. Klenow Fragment 3'-5' exo (NEB) was used to create A overhangs and the library was once again bead-purified. Custom made P2 adapters were ligated to the A overhangs. Approximately 20 ng of library was amplified with Phusion HF Master Mix (NEB) in total of 14 cycles. The amplified library was visualized using HS DNA Analysis Kit in 2100 Bioanalyzer (Agilent Technologies). Approximately 5 ng of amplified library was sequenced. The first 50 samples were sequenced on HiSeq 2000 Illumina platform at Oxford Genomics Centre using the TruSeq reagents. Another 40 samples were submitted for sequencing on Genome Analyzer II Illumina platform in Genomic Centre at University of Warwick. During the second run the TruSeq reagents were also used.

2.3 *Basic bioinformatics processing of RAD tags*

Raw reads from Illumina sequence runs were assembled into FastQ format and assessed in FAST QC software v0.10.1 to check for standard quality measures: quality scores, overrepresented sequences, GC content and N content[58]. Reads were demultiplexed using fastq-multx script available in github (https://github.com/brwnj/fastq-multx) and mapped to reference genome[33] using the bwa mem (v0.7.10) algorithm[59] and sorted using samtools (v1.2)[60] Mapped reads with mapping quality (MQ) below 30 were removed from alignments. Variant calling was conducted following Genome Analyses ToolKit (GATK) best practices[60,61]. For each individual separately gatk v.3.7 was used to realign mapped reads around indels and to discover variants with HaplotypeCaller programme. Following to

that genotyping was done jointly for all samples using GenotypeGVCFs programme. Polymorphic sites were filtered with following specification: quality by depth (QD) should be larger than 10; genotype call should be present in at least 80% of individuals; site should be within 400 bp to SbfI restriction site in the reference genome.

For site frequency spectra we additionally generated SNPs from *de novo* assembled RADtags. Scripts from the STACKS pipeline v1.05 was employed to de- multiplex and discover RAD sequence markers[62]. Low quality sequences and those characterized by erroneous barcodes were discarded. Remaining sequences were sorted according to their barcodes into single-end (P1) and paired- end (P2) sequence files. The DENOVO_MAP.PL script of the STACKS package was used to call stacks of loci for all the multiplexed samples. The following settings were applied: 5 identical, raw reads required to create marker stack, 5 mismatches were allowed between alleles in single individual, 15 threads were executed and calling haplotypes from secondary reads was disabled. Marker stacks were then imported into a MYSQL database. For loci that were characterized by very high F_{ST} values between northern and southern cultivars, paired-end reads were collated using BOWTIE2 v2.1.0[63] and then transformed into indexed bam file with SAMTOOLS v0.1.19[64]. Loci that were present in all 90 samples and contained exactly 1 SNP were exported in STRUCTURE format.

2.4 Population structure analysis of RAD tags

Nucleotide diversity within flax subpopulations was measured by taking mean of diversity calculated for each RADtag with vcftools (v0.1.13)[65] Our samples were divided into the following subpopulations of flax: 1) all pale flax, 2) Turkish pale flax, 3) Greek pale flax, 4) Croatian pale flax, 5) all cultivated flax, 6) fibre types, 7) oil types, 8) intermediate types, 9) landraces, 10) dehiscent varieties, 11) northern cultivated flax and 12) southern cultivated flax, Table S1. Distances calculated from identity

by sequence between all individuals were measured using plink (v1.07)[66]. 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)[67]. We assumed existence of two to seven ancestral populations and chose the model, which was characterised with the lowest five-fold cross-validation error (Figure S2).

2.5 Northern and southern population differentiation analyses

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. The N40° latitude delimits Turkey and Greece from Central and Northern Balkans and is used here as approximate borderline between the distribution of domestication-associated *LuTFL1*.I allele and northern-associated *LuTFL1*.III allele in pale flax. We calculated F_{ST} value between northern and southern cultivated flax populations for each RADtag using F_{ST}() function in PEGAS[68] and build null distribution.

Allele frequency spectrums were generated from in house scripts from variant call format files generated from SAM tools in the case of RAD loci mapped to the flax genome (Figure S3), or STRUCTURE format files in the case of *de novo* assembly RAD tags (Figure 2). Residual frequency distributions were calculated by subtracting frequency values from corresponding cells either from the above or below the diagonal to give an assessment of the symmetry of the allele frequency spectrum as a means to detect potential biases in the data.

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 in either the rows (horizontal), columns (vertical) or perpendicular to the diagonal.

We utilized the ancestral information content concept[38] to employ a high variant test to explore possible gene movement between populations when there may be complex confounding movements masking a general signal that might be detected through f4 statistics. Our starting premise is that loci that have become highly differentiated between potential source populations through drift may be exploited as markers for that population to probe gene movements between specific populations in the face of multiple general movements. This concept has recently been used to investigate maize genome ancestry[69], and generally to identify ancestry informative markers[38]. The probability that a certain allele will be involved in a gene flow process is expected to be proportional to its frequency in the population, and the likelihood such alleles would result in a detectable impact in the recipient population would be expected to increase as the difference in frequencies (∂f) between the population increases. In a case in which two sets of two populations have differentiated from each other independently, we would expect highly differentiated loci in one population pair to be independent to the extent of differentiation of the corresponding loci in the other population, Figure S4. Using the terms of Figure S4, we assume the relationship of branch lengths to be as follows:

$$i + j \approx w + x$$

$$k + l < y + z$$

We describe the subset of differentiated loci as ∂f where the delta term refers to the minimum extent of difference between frequencies in one population pair, and f_i the differences across all loci. However, instance in which the following relationship is satisfied:

$$i + j < w + x$$

would suggest a non-independent process of differentiation between the sets of populations. By

selecting subsets of loci that are specifically high in one population (e.g. population C in figure S4), we can explore the independence of differentiation with specific respect to that donor population.

The expected relationship between our flax populations is that firstly wild southern and northern populations diverged, subsequently flax became domesticated in the south, and then spread to the north diverging from its respective southern population independently from the wild diversification. It is reasonable to expect that the southern cultivated population should resemble the southern wild population, but that should not matter for this analysis. However, we first examined the differentiation of all loci through Fst values between the respective northern and southern populations of wild and cultivated flax to check for broad correlation of frequencies that would compromise the underlying assumption of independent differentiation and possibly indicate a hitherto undetected signature of structure, Figure S5. The Fst values between the wild populations and the cultivated populations appear to be uncorrelated ($r = 0.098$), supporting the notion that differentiation has been largely independent in wild and cultivated flaxes. We then identified three subsets of loci in which the difference in frequency (∂f) between northern and southern wild populations was more than 0.5, 0.4 or 0.3, with the higher allele in the northern population. This process was repeated in the wild southern population. The average Fst value across those loci subsets was then calculated for the cultivated population, Table S3. To compare the ∂f subset mean Fst obtained in cultivated flax populations with the total null background of markers we sampled 10000 times random subsets of the same sample size from all Fst values between cultivated populations and calculated means producing null mean Fst distributions, Figure S6. In the case of $\partial f > 0.5$ and higher frequency in southern wild populations, we resampled 1000000 times because of the low number of loci recovered that discriminate southern populations to this level. We found that loci with alleles that are of discriminatingly high frequency in southern wild populations were agnostic in cultivated populations. However, in the case of northern wild populations

we find a strongly significant relationship between alleles that discriminate wild populations and those that discriminate cultivated populations, suggesting genetic contact between the northern cultivated and wild populations. We examined the ancestry information content of the ∂f loci subsets using the I_a ($I_{a\partial f}$) statistic[38] relative to the information content of all loci (I_{at}), Table S3. These results echoed our variant allele test, with increasing high values of ∂f being associated with an increased ancestral information content of the northern but decreasing in the southern wild population. This results are congruent with a process in which cultivated populations passed through the southern wild flax range in which higher frequency alleles were more more likely to become incorporated in both the resultant southern and northern cultivated flax populations and hence be poor ancestral discriminators. The high ancestral information content of the discriminatingly higher northern wild populations is congruent with some distinct ancestral contribution to northern cultivated flaxes. Together these data suggest a process of gene movement from the northern wild flaxes into northern cultivated populations.

3. PEBP Gene Survey

3.1 PEBP gene homolog identification

Based on multiple alignments that contain sequences from different eudicots, degenerate primers were designed in exonic regions to cover between 400 to 1800 bp sized target regions. Primers together with their respective sequences used for multiple alignments are listed in Table S5. All loci were amplified by the polymerase chain reaction (PCR) with Platinum® Taq Polymerase (Invitrogen). PCR was carried out in a TaKaRa Gradient Thermocycler with the following programme: initial denaturation in 94°C for 3 minutes, 32 cycles with sequence: 94°C for 45s, gradient temperature (55°C - 60°C) for 45s and 72°C for 75s, completed with the final elongation step in 72°C for 5 minutes. Samples of PCR product along with 100bp size ladder (Fermentas) were visualized under UV light

after electrophoresis in GelRedTM-stained, 1.5% agarose gel. PCR products were purified using QIAquick® PCR Purification Kit (Qiagen) followed by sequencing reaction with BigDye® Terminator. Sequencing fragments were then analyzed in ABi Prism® Genetic Analyzer (Applied Biosystems). The resultant sequences were then used to search for similar genes within NCBI database using BLASTN algorithm[70] in order to confirm that amplified products were homologous to the PEBP gene family.

The putative sequences of *TFL1* were subject to BLAST searches against the cultivated flax genome scaffolds v1.0 and CDs v1.0 databases[71] in order to identify full sequences of genes of interest and their flanking regions. The sequences found in this way were then subject to comparison of gene structure: sequences with the same number of exons (plus/minus one) as in the original *A. thaliana* gene were kept for further analysis. This allowed for filtering out pseudogenes and assembly artefacts. Open reading frames (ORFs) were estimated using FGENESG tool from Softberry® package[72].

A total of 8 putative paralogues of *TFL1* were identified. The apparent expansion of *TFL* homologues subsequent to the split of the *Populus* and *Linum* lineages of the Malpighiales was unexpected and could indicate an episode of adaptive evolution related to flowering time, Figure S9. We use the terminology *LuTFL* to specify *TFL* homologs from *L. usitatissimum.*

To identify the likely functions of the putative *LuTFL* homologs, the predicted amino acid sequences were compared sequences from *A. thaliana* and *P. nigra* (NM_101672, NM_125597, NM_101672, NM_118156, NM_120465, NM_128315, AB161110, AB181183, AB181185). All sequences were aligned in CLUSTALX v2.1 . The best model of protein evolution was selected using BIONJ algorithm[73] and Bayesian Inference Criterion[74]. Phylogenetic trees were estimated using Bayesian Inference in MRBAYES v3.2[75] with four chains in 1.000.000 generations using best-fitted model of evolution. The final tree with posterior probabilities for each clade was obtained by

summarizing sampled trees excluding 25% of initial results, Figure S9. None of the flax sequences showed similarity to ancestral *MFT* genes. Well-supported clades separated *FT*-related genes from *TFL* homologs. Only one of the putative homologs in flax – *LuTFL8* was assigned to *FT* clade, while remaining 7 grouped with *TFL*-like genes. Within those, both of the 3-exonic putative genes in flax (*LuTFL5* and *LuTFL6*) are clustered with *BFT*, while *LuTFL7* is most similar to *TFL1*. The remaining putative *TFL* homologs in flax (*LuTFL1*, *LuTFL2*, *LuTFL3* and *LuTFL4*) form a clade with *CEN* in *Arabidopsis* and *Populus*, but with a branching order conflicting with the species relationships so are considered here to be paraphyletic to CEN with high posterior probability. It is likely that this group of flax *TFL*s have functions similar to that of *CENTRODADIALIS* and/or *TFL1*.

The predicted amino acid sequences of flowering gene homologs in flax were compared to the models of active centres and binding pockets within respective genes in *A. thaliana* and *P. nigra*[76] to assess if crucial amino acids were conserved (Figure S10). We found that all the amino acids forming binding pocket and activity-specific residues in *A*. *thaliana* were conserved in flax, which further supports functional conservation between *TFL1* and *LuTFL* homologs.

Candidate genes were then tested to see if they showed evidence of differentiation between northern and southern flax populations. Specific primers were designed to cover all the exons and introns in the putative *TFL* genes. PCR and sequencing reactions were carried out with the same reagents and equipment as described above, using primers outlined in Table S4. Sequences were obtained where possible for 6 samples representing 4 distinct types of cultivated flax (oil, fibre, intermediate and indehiscent) and four pale flax populations – from northern Balkans and Turkey (M002, M005, M024, M065, W042, W046, W065, W069). We found that the sequences of amplicons from *LuTFL5* and *6* were indistinguishable. *LuTFL4* showed no evidence of diversity segregating north and south populations and *LuTFL7* showed no evidence of diversity in pale flaxes in our survey.

However, *LuTFL1*, *LuTFL2* and *LuTFL3* showed high levels of diversity, Table S5. Cultivated flaxes appear more like northern pale flax populations at *LuTFL2* and southern populations in *LuTFL3.* *LuTFL1* showed evidence of mixed ancestry in cultivated flax populations from both northern and southern pale flax populations. We therefore decided to focus on the two loci that showed some evidence of gene flow from northern populations of pale flax into cultivated populations, *LuTFL1* and *LuTFL2*. A new PCR system was designed for these loci based on the sequence information retrieved

3.2 Phylogeographic Analysis of LuTFL1 and LuTFL2

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 *L. decumbens* accessions, Table S1. The total DNA was isolated using modified DNEasy® Plant Mini Kit (Qiagen) protocol. 20 mg of seeds or in case of herbarium specimens – mixed plant tissue was ground and a mixture of 660 µl of 2% CTAB solution with 340 µl of MilliQ water was added followed by incubation in water bath at 65°C for 90 minutes. After centrifugation supernatant was transferred to 800 µl of chloroform:isoamyl alcohol (24:1) and mixed gently. This mix was centrifuged at high speed and supernatant was moved to tube with 2 volumes of AP3/E buffer followed by standard protocol. DNA concentration was measured with Quant-itTM dsDNA Broad-Range Assay kit (Invitrogen) in Qubit® fluorometer (Invitrogen).

A 1367 bp target region was amplified from *LuTFL1* spanning exons 1-4 using primers (5-TTACAACTCCACCAAGCAAGTC, and 5- TGTCTCGCGCTGAGCATT). In the case of *LuTFL2* a 383 bp target region was amplified from *LuTF2* spanning exons 1 and 2 using primers (5-TCGTCAGATCCTCTAGTGATA, 5- TGATTGATGTTATGTGTGTATGG). The conditions for PCR steps for *LuTFL1* were: 10 cycles with 94°C (45s), 59°C (45s) decreased by 0.1 with each step and 72°C (75s) followed by 25 cycles with 94°C (30s), 58°C (40s) and 72°C (75s). Similarly, cycling steps

for *LuTFL2* are: 10 cycles with 94°C (45s), 55°C (45s) decreasing by 0.1 with each step and 72°C (60s) followed by 25 cycles with 94°C (30s), 54°C (40s) and 72°C (60s).

Haplotype networks for *LuTFL1* and *LuTFL2* were made using uncorrected_P character transformation and Rooted Equal Angle splits transformation in *SplitsTree4*[77]. The size of the network nodes was made proportional to the number of samples that share a particular haplotype, Figure 3A, Figure S11A, Table S1 and S6. Almost half of pale and cultivated flax samples were characterised with precise geographic location, Table S1. For the remaining samples of cultivated flax only the country of origin was known. In these cases averaged latitudes (centroids) were established based on the locations provided by the Geographic Names Server (GNS), maintained by the National Geospatial-Intelligence Agency (https://www.cia.gov/library/publications/the-world-factbook/fields/2011.html). The relationship between latitude and *LuTFL1* and *LuTFL2* haplotypes separately was modelled using logistic regression in *R* programme and plotted using *popbio* package, Figure 3B and S11B. Increase of haplotype frequency per grade of latitude was calculated using logit transformation.

For the majority of wild flax samples (28/32) precise geographical data are available, these were used in testing spatial autocorrelation against *LuTFL1* and *LuTFL2* haplotype data, Figure 3C and S11C. A Classic Mantel's test of matrix congruence[78] was carried out together with custom-made permutation test. This test written in *R* by RG computes the sum of average pairwise distances within haplogroups (s-score), the test is available for download from (http://www2.warwick.ac.uk/fac/sci/lifesci/research/archaeobotany/downloads/flax). It compares the s-score for observed data to the s-scores for permuted data, in which haplotypes are randomly dispersed on spatial scale. When the s-score for observed data is lower than 5% of s-scores from permuted data then the null hypothesis of randomly distributed haplotypes is rejected. The test outputs are in Table S7.

3.3 Tests for selection

In order to test if *LuTFL1* and *LuTFL2* were inherited neutrally or were under selection, neutrality tests with variety of estimators were carried out, Table S9. Tajima's D[79], Fu and Li's D, Fu and Li's F[80] statistics were calculated in *INTRAPOP*[81] and R2 statistic[82] was calculated in *R* using the *PEGAS* package.

3.4 Linkage disequilibrium analysis

The extent of linkage disequilibrim was examined between the alleles of the *LuTFL1* and *LuTFL2* loci, Table S8. Expected frequencies of genotypes were calculated from the observed individual allele frequencies. The correlation coefficient of *D*, *r*, was calculated from observed genotyping data (Table S6) where:

$$r = \frac{[p1p2]\times[q1q2]-[p1q2]\times[q1p2]}{\sqrt{p1p2q1q2}} \qquad (1)$$

Where square brackets indicate the observed proportion of the sample of genotype combinations, *p1, p2, q1 and q2* 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*.

There is little overlap between genotypes found in the wild and in cultivated populations, with notably over-represented genotypes in both sets. *LuTFL1* III showed linkage with *LuTFL2 I* in pale flaxes and *LuTFL2* X in cultivated flaxes, but neither genotype occurs in both sets. In pale flaxes a strong association was also found between *LuTFL1* VIII and *LuTFL2* III. The positive correlations in pale flax can be explained by phylogeographic co-occurrence. Conversely, a number of genotypes did not occur in the pale flaxes, such as *LuTFL1* VIII and *LuTFL2* I, most likely because these occur in different regions. The high frequency of genotype of *LuTFL2* X in cultivated flaxes despite its relative low frequency and restricted distribution in the pale flaxes is suggestive of a founder effect in

cultivated flaxes. The high degree of association of *LuTFL2* X with *LuTFL1* III is suggestive that the founder effect could have occurred as a result of the selection of the *LuTFL1* III allele, with the *LuTFL2* X allele being carried by a hitchhiking effect. If this is the case, then it is likely that the selection for *LuTFL* III was in a geographically restricted region in the north of Turkey close to the Black Sea coast, which would have been when the spreading cultivated flaxes first came in to contact with the allele.

3.5 Structure of the LuTFL1 locus

The analysis of multiple alignment of *LuTFL1* revealed that the length of the target region of the locus used for phylogeographic analyses varies from 1284 to 1306 bp (Figure S14A). The full gene (as measured from start to stop codon) is 1498 bp long and consists of four exons with total length of 552 bp. Exons 2 and 3 are of exact same length as in *TFL1* in *Arabidopsis thaliana*, 62 and 41 bp respectively, while exon 1 is 15 bp longer (225 bp) and exon 4 - 3 bp shorter (221 bp). In 5' UTR of *LuTFL1*, we identified motifs characteristic to promoter region, including TATA-box and TCA triplet (Figure S14B). Further to that, we discovered a transcription factor binding site 20 bp upstream from start codon through GOMO[83] search in MEME suite[84]. This binding site carries mutation in *LuTFL1* I haplotype.

3.6 Selection coefficient model

Each latitude in Europe is associated with a different time of arrival of agriculture based on archaeological data. The program SELECTION_TIME.pl[39] estimates selection coefficients from dated frequencies (available for download from http://www2.warwick.ac.uk/fac/sci/lifesci/research/archaeobotany/downloads/flax). The model takes as

input a series of dated frequencies, mating strategy of the organism, whether the data is phenotype or allele frequency and whether the trait under selection is dominant or recessive. The estimation of allele frequency from phenotype function was not required in this study because the input data was allele frequency. To calculate the inbreeding coefficient associated with the mating strategy the program begins with Hardy Weinberg frequencies based on 50% frequency of alleles in a two-allele system, and then iterates through generations making genotypes based on the mating strategy using equations *2*, *3* and *4*.

$$P' = m(P + 0.25Q) + (1 - m)(P^2 + PQ + 0.25Q^2) \quad (2)$$

$$Q' = m(0.5Q) + (1 - m)(2PR + 0.5Q^2 + PQ + QR) \quad (3)$$

$$R' = m(R + 0.25Q) + (1 - m)(R^2 + QR + 0.25Q^2) \quad (4)$$

Where P, Q and R denote the proportions of the dominant homozygote, heterozygote and recessive homozygote individuals in a population respectively, m denotes the mating strategy defined as the probability of a selfing event, and P', Q' and R' denote the proportions of P, Q and R in the next generation. This iteration causes genotype frequencies to equilibrate to their expected frequencies under the mating strategy m. Each generation the inbreeding coefficient F is calculated as determined by the under-representaton of heterozygotes and over-representation of homozygotes in the case of inbreeders following equation *5*.

$$F = \frac{R - q^2}{pq} \quad (5)$$

Where p and q are the proportions of the dominant and recessive alleles respectively. The iterations continue until consecutive values for F differ by less than 1×10^{-10}, and the frequency equilibrium of genotypes has been reached.

The program then takes a pair of input observed dated allele frequencies as start and stop frequencies. The inbreeding coefficient F is used to determine the initial genotype frequencies using

the equations *6, 7* and *8*.

$$P = p^2 + Fpq \quad (6)$$

$$Q = 2pq(1 - F) \quad (7)$$

$$R = q^2 + Fpq \quad (8)$$

Beginning with the initial (lower frequency), a number of generations equal to the date difference are iterated through using equations *9, 10* and *11*.

$$P' = (1 - k)m(P + 0.25Q) + (1 - k)(1 - m)(P^2 + PQ + 0.25Q^2) \quad (9)$$

$$Q' = (1 - k)m(0.5Q) + (1 - k)(1 - m)(2PR + 0.5Q^2 + PQ + QR) \quad (10)$$

$$R' = m(R + 0.25Q) + (1 - m)(R^2 + QR + 0.25Q^2) \quad (11)$$

Where *k* is the selection differential taking values between -0.000001 and -0.999999, and selection is of the dominant allele *p* in this case to reflect the existence of the transcription-factor binding site in the 5' upstream region of the *LuTFL1* III allele under selection. The selection coefficient *s* is then described by equation *12*.

$$s = 1 - \frac{1}{(1-k)} \quad (12)$$

The variable k_i is designated as either k_u or k_o depending on whether it's implementation in equations *9* and *10* lead to a final gene frequency of *p* (p_f) that is under or over the real observed final gene frequency (p_{fo}). The initial values of k_u and k_o are -0.000001 and -0.99999 respectively. Iterative values of *k* are calculated using equation *13*.

$$k_i = k_u + \frac{k_o - k_u}{2} \quad (13)$$

$$k_{u+1} = k_i \ for\ p_{fi} < p_{fo} \quad (14)$$

$$k_{o+1} = k_i \ \ for\ p_{fi} > \ p_{fo} \qquad (15)$$

The iterations continue until consecutive estimates of k_u and k_o differ by less than 1x 10^{-6}. Estimates of *k* were made for all pairs of dated frequencies. The average value of *s* found for the *LuTFL1* III haplotype is given in Figure S12A.

To account for sample error, we generated upper and lower bounds of *s* for dated frequencies. To do this we used a Beta distribution to estimate the 95% range of possible underlying frequencies that could have given rise to the observed number of genotypes in our samples assuming a binomial process of *x* observations in *n* samples, where x equates to the number of samples carrying the *LuTFL1* III allele, and *n* equates to the total number of samples at a specific latitude. We then calculated all possible values of *s* using the upper and lower sample number bounds respectively, Figure S12A.

4. Phenotypic analyses

4.1 RNA expression analysis

Plants of six populations (W042, W043, W067, W069, W077 and W094), representative of three geographical regions (Turkey, Greece and Croatia) were chosen to form the core experimental plants and sown in 100 replicates each. Seeds were stratified for 3 days in the dark at 4°C. For the subsequent ten days, they were grown at room temperature with a 16 hour photoperiod. These growth cabinets were moved to a cold room at 4°C a vernalization period of 40 days. Plants were moved after 40 days to MLR-351 Versatile plant growth (SANYO Electric Co., Ltd.) cabinets set up to emit light for 16h in temperature of 24°C during a day and 16°C during a night. Plants were grown in these conditions for another 40 days and then moved to the larger pots and into the glasshouse (daylength between 16 h 42 maximum and 15 h 55 minimum).

During the time course experiment, at time intervals of 0, 15, 17, 19 and 21 days samples of three

plants were taken from each of the core pale flax accessions. Sampled material included the main meristem, 1 cm of stem from the meristem and adjacent leaves. Tissue was rapidly frozen in liquid nitrogen and stored at -80°C. Frozen tissue was then ground in a pre-chilled mortar on dry ice. The total RNA extraction was carried out from 20 mg of ground tissue using mirVana™ miRNA isolation kit (Invitrogen) following the standard protocol. RNA intactness was assessed using electrophoresis. RNA concentration was measured in a NanoDrop spectrophotometer. DNA contaminants were digested in reaction with 2 units of DNase I (Invitrogen) and 2 µg of total RNA in DEPC-treated water. Further to that, total cDNA was synthesized from 1µg of DNase-treated RNA with use of SuperScript® II reverse transcriptase and unspecific Oligo dT primer following the manufacturer instructions. The concentration of each cDNA sample was adjusted to 100 ng/µl. Specific primers were designed in exonic regions of *LuTFL1*, *LuTFL2* and *LuFT* to cover approximately 150 bp long fragment (Table S11). The *GAPDH* housekeeping gene was chosen as an expression control (Huis et al. 2010). Semi-quantitative PCR was carried out for each sample with 30 cycles. Each cycle consisted of the following steps: 30 s in 92°C for denaturation, 30 s in 56-59.5°C for annealing and 30 s in 72°C for elongation stage. PCR products were visualized in UV on 2% agarose gel after electrophoresis with size marker, Figure S15.

4.2 Heterologous expression analysis of LuTFL1

We have used Landsberg erecta (Ler-0) ecotype with functional *TFL1* gene as a reference and its tfl mutant, *tfl1-2*[85] 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 by PCR with Phusion® High-Fidelity

DNA Polymerase from cDNA samples of pale flax accession W067 (Table S1) using the primers (5' ATGCAGGAGAAATCGATGGGGAAAGTG 3' and 5' CTAACGCCTCCTGGCTGCTGTCTC 3'). The PCR product was then introduced into a Sma1-digested Gateway entry vector, pJLBlue[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-IIS[86]. This recombination reaction effectively placed the *LuTFL1*.III allele behind the constitutive CaMV 35S 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[87]. 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 35S::*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, Figure S16. The difference in mean phenotypes for 35S::*LuTFL1*.III, Ler-0 and tfl1-2 was tested using t-test in R. Plots were generated using pirate plot package (v2.0).

4.3 Flowering time and stem height analysis

Thirty two populations of pale flax were sown in five replicates for phenotyping experiment in the same experiment as described in section 4.1. Flowering time was recorded when floral primordia appeared for the first time. After maturation, plant height was measured along with weight of 1000

seeds for each pale flax population. The data for Turkish populations were obtained from authors of phenotypic study in Turkish pale flax[23]. The data for cultivated flax was already published in an earlier article[36]. Flowering time for cultivated and pale flax, stem height and branching ,its associations with latitudes and *LuTFL1* alleles were summarised in Figure S18 and Table S12. A similar plot for *LuTFL2* alleles is presented in Figures S19.

4.4 Seed size, flowering time and fibre quality correlation analysis

Correlation between mass of 1000 seeds, height, fibre content, branching pattern and flowering time in cultivated flax was measured in R using Pearson's and Kendall's ranks. Flowering time was significantly, positively correlated with height and branching pattern and less significantly, negatively correlated with seed size (Figure S20).

5. Developmental growth model

The *PGROWTH* model simulates flax growth and development (http://www2.warwick.ac.uk/fac/sci/lifesci/research/archaeobotany/downloads/flax). The key event in the model determining plant architecture is flowering time. A change in the flowering time impacts on plant stature and branching pattern. A script in the *R* programming language called *PGROWTH* has been written by RG to simulate the impact of the photoperiod-dependent pathway on flax architecture during its development. It is an implementation of observations made by McGarry and Ayres[48] on the impact of *FT*/*TFL* expression on tomato plant architecture, and represents a simpler system than that of Prusinkiewicz[32.] In order to simulate flowering on second order branches we used a three-threshold approach, the first to initiate branching, the second to initiate flowering in lateral branches and the third

to terminate flowering. This contrasts with Prusinkiewicz[25] who used a system of reversions between meristem states to achieve a wider range of architectures than occurs in our model, or with flax. The thresholds we used were trained to recapitulate type of growth and inflorescence architecture characteristic for flax. The molecular basis of the model assumes that *TFL* and *FT* compete for the same binding site, and that when *FT* exceeds *TFL* expression floral initiation will occur. The model takes as input variables, the germination date, the season length, the minimum daylength required to express *FT,* the level of TFL expressiom, and the geographical latitude of the plant simulated. The model components were as follows:

- *Space*: Latitude is an input variable to the model in order for daylength to be calculated. Latitude is input as the variable α in radians as described in equation (*16*)

$$\alpha = \frac{lat}{57.296^{\circ}} \qquad (16)$$

- *Time*: The model increments on a daily basis, with the first day equating to the 1st March in this study. The 'date' is converted to the variable β in radians using equation *17*.

$$\beta = \frac{2\pi(x+59)}{356} \qquad (17)$$

- *Daylength*: This is calculated using an empirical formula[88] based on the date in radians (β), and input latitude in radians (α):

$$daylength = tan(\alpha) * tan\left(-\frac{\pi\left(22.9133 * cos(\beta) + 4.02543 * sin(\beta) - 0.3872 * cos(2 * \beta) + 0.052 * sin(2 * \beta)\right)}{180}\right) * \frac{180}{\pi} \qquad (18)$$

- *Vegetative growth*: During this phase of simulated plant growth stems and adjacent leaves grow indeterminately and uniformly. In this study scenario, plants grew 1 cm per day and leaves were produced every two days.

- *Transition*: The vegetative phase ends when the expression of *FT* is greater than the expression of *TFL* triggering the transition from vegetative to generative growth.
- *TFL and FT expression*: *TFL* is assumed to be constantly expressed in the model at a set number of units per day, while *FT* is expressed only when the user-defined daylength threshold is met. *TFL* expression levels take a value between 0-100 units in this model. *FT* expression is a linear function of the number of leaves below each meristem, each leaf contributing a single unit, therefore increasing expression levels as vegetative growth progresses. Note that it is a simplification in this model that *FT* producing leaves are insensitive to increasing daylength in regard to the quantity of *FT* produced. Given the close recapitulation of real flax architectures produced by the model, this appears to be a reasonable simplification.
- *Generative growth*: During the generative phase the rates of stem growth remains the same as with the vegetative phase, but instead of producing leaves every second day the main stem produces leafy-bracts and axillary stems (floral branches) alternately. Axillary branches grow in a vegetative manner until a second threshold is reached, which was set to when *FT* expression is 1.1 times higher than the *TFL*. Once this second threshold is reached on the axillary stems leafy-bracts and pedicels are generated alternately every 2 days. The generative growth phase continues producing axillary branches from the main stem and pedicels on the axillary stems until the third termination threshold has been reached, which is when *FT* is 1.4 times *TFL*. Alternatively, all meristems are terminated with an inflorescence after a defined period of time, in this analysis the season length was set to 150 days. The combination of thresholds used in this model were selected to produce a racemose corymb that closely reflects flax inflorescence.

The model executes beginning at day 1, which corresponds to 1st March. The day length of each day is

calculated using equation (18) based on latitude and date as defined by the day number. At a user-defined date a simulated germination occurs which is defined as the initiation of vegetative growth in this model. During this phase each day the plant accrues 1cm length along a single un-branched main stem axis, and every other day a leaf is produced. In this model *TFL* is user-defined to a value between 0-100 units and produced from the first day of vegetative growth at a constant level. The expression level of *TFL* sets the threshold that *FT* expression must reach to trigger generative stage. When a threshold daylength is reached, defined by 14 hours in this study, the vegetative plant begins to express *FT*. The quantity of *FT* is dependent on the number of leaves in the plant, with each leaf producing one unit of *FT* daily. The vegetative phase of growth continues until the leaf production of *FT* is greater than *TFL* expression, upon which point the generative mode of growth is assumed. In the generative phase the simulated plants generate axillary branches on the main stem. Axillary stems after reaching 1.1 *FT*:*TFL* ratio switch to generative phase in which flowers are produced. In the generative phase the simulated plant on the main stem produces leafy-bracts and axillary stems (floral branches) alternately every fifth day. On the axillary stems leafy-bracts and pedicels are generated alternately every 5 days. In this simulation growth was terminated with a terminal inflorescence either by end of season defined as 150 days after the germination date (user-defined date) or by the ratio of *FT*:*TFL* reaching 1.4.

The script includes graphical functions to visualize the output architecture. In this study the following parameters were explored: latitudes N35° and N60°, *FT* expression threshold set to 14 hours and *TFL* expression was set from 0 to 100 which should cover all possible architectures achievable under the combination of thresholds used in this study.

Different input settings yielded different inflorescence architecture, Figure S21. These architectures

reflected closely sketches of different flax convarieties made by Kulpa and Danert[89]. Generally, the effect of increased latitude was to produce shorter bushier plants for a given *FT/TFL* floral initiation threshold, while an increase in the threshold value gave rise to taller, less branched plants at any one latitude, Figure S21. This finding is consistent with the notion that an increased expression of *TFL* or reduced expression of *FT* would likely give rise to tall phenotypes more suitable for fibre use. However, the model also demonstrated the expectation that a larger range of architectures would be expected to be viable at higher latitudes because at low floral initiation ratios short plants would flower earlier in the year, and at high initiation ratios tall phenotypes would be able to flower later in the year. Therefore given a selection pressure to adapt to latitude by strengthening the signal to delay flowering in order to achieve a similar date of flowering, we would expect taller phenotypes to prevail, Figure S22.

References

57. Etter, P.D., Bassham, S., Hohenlohe, P.l., Johnson, E.A., Cresko, W.A. SNP discovery and genotyping for evolutionary genetics using RAD sequencing. In: *Molecular Methods for Evolutionary Genetics* (ed. Rockman MV), pp. 157-178. Humana Press, Orgogozo, Virginie (2011)

58. Andrews, S. (2012) *FastQC A Quality Control tool for High Throughput Sequence Data.* http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

59 Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM arXiv 1303.3997.

60. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G et al. (2013) From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline *Curr. Protoc. Bioinformatics*, 43:11.10.1-33.

61. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data *Nat. Genet.* 43(5):491-8.

62. 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).

63. Langmead, B., Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357-359 (2012).

64. Li, H., Handsaker, B., Wysoker, A.*, et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).

65. Petr Danecek, Adam Auton, Goncalo Abecasis, Cornelis A. Albers, Eric Banks, Mark A.

DePristo, Robert E. Handsaker, Gerton Lunter, Gabor T. Marth, Stephen T. Sherry, Gilean McVean, Richard Durbin (2011) 1000 Genomes Project Analysis Group; The variant call format and VCFtools. *Bioinformatics* 27 (15): 2156-2158

66. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M, Bender D, Maller J, Sklar P, de Bakker P, Daly MJ, Sham PC (2007) PLINK: A Tool Set for Whole-Genome and Population-Based Linkage Analyses. *American Journal of Human Genetics*, 81. : 559–575.

67. Alexander DH, Novembre J, Lange K. (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19(9):1655-1664.

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

69. Ramos-Madrigal, J., Smith, B.D., Moreno-Mayar, J.V., Gopalakrishnan, S., Ross-Ibarra, J., Gilbert, M.T., Wales, N. (2016) Genome Sequence of a 5,310-year-old maize cob provides insights into the early stages of maize domestication. *Current Biology* 26(23):3195-3201.

70. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403-410.

71. Rowland, G., Cloutier, S. Linum.ca database, http://anther.ccis.ualberta.ca/"I" a (2012).

72. Solovyev, V., Kosarev, P., Seledsov, I., Vorobyev, D. (2006) Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biology* **7**, 1-12.

73. Gascuel, O. (1997) BIONJ: An improved version of the NJ algorithm based on a simple model of sequence data. *Molecular Biology and Evolution* **14**, 685-695.

74. Schwarz, G. (1978) Estimating the dimension of a model. *Annals of Statistics* **6**, 461- 464.

75. Ronquist, F., Teslenko, M., van der Mark, P.*, et al.* (2012) MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61**,

539-542.

76. Igasaki, T., Watanabe, Y., Nishiguchi, M., Kotoda, N. (2008) The *FLOWERING LOCUS T/TERMINAL FLOWER 1* family in Lombardy poplar. *Plant and Cell Physiology* **49**, 291-300.

77. Huson, D.H., Bryant, D. (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**, 254-267.

78. Mantel, N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research* **27**, 209–220.

79. Tajima, F. (1989) Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585-595.

80. Fu, Y.X., Li ,W.H. (1993) Statistical test of neutrality of mutations. *Genetics* **133**, 693- 709.

81. Achaz, G. (2013) http://wwwabi.snv.jussieu.fr/achaz/neutralitytest.html.

82. Ramos-Onsins, S.E., Rozas, J. (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution* **19**, 2092-2100.

83. Buske, F.A., Bodén, M., Bauer, D.C., Bailey, T.L. (2010) Assigning roles to DNA regulatory motifs using comparative genomics. *Bioinformatics* **26(7)**, 860-866.

84. Bailey, T. L., Bodén, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., Noble, W.S. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research* **37**: 202-208.

85. Alvarez J, Guli CL, Yu X-H, Smyth DR. 1992. terminal flower: a gene affecting inflorescence development in Arabidopsis thaliana . *The Plant Journal* 2: 103–116.

86. Hellens RP, Edwards EA, Leyland NR, Bean S, Mullineaux PM (2000) pGreen: a versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation. *Plant Mol Biol* 42: 819–832.

87. Weigel D, Glazebrook J (2002) Arabidopsis: A Laboratory Manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.. 354 p.D. WeigelJ. Glazebrook2002Arabidopsis: A Laboratory Manual.Cold Spring Harbor, NYCold Spring Harbor Laboratory Press.

88. Kirk, J.T.O. (1994) *Light and photosynthesis in aquatic ecosystems*, 2 edn. Cambridge University Press, Cambridge.

89. Kulpa, W., Danert, S. (1962) Zur systematik von *Linum usitatissimum* L. *Kulturpflanzen* **3**, 341-388.

Supplementary Figure legends

Figure S1 Multidimensional Scaling Plot (MDS) of pairwise nucleotide distances between flax accessions at RAD loci.

Figure S2 Graphs showing delta K for different K number of populations in ADMIXTURE analysis.

Figure S3 Allele frequency and residual spectra analyses of RAD loci. A. SNPs called against RAD loci mapped to the flax genome including the reference state (ie not variants), 18,314 SNPs. B Variant SNPs called against RAD loci mapped to the flax genome, 18,314 SNPs. C. SNPs called by denovo assembly of RAD loci with no reference genome, 1684 SNPs. Top row: allele frequency spectra. Middle row: residual spectra showing bottom right cells subtracted from top left cells (A, 9989 SNPs; B, 418 SNPs; C, 358 SNPs). Bottom row: residual spectra showing top left cells subtracted from bottom right cells (A, 43 SNPs; B, 6552 SNPs; C, 280 SNPs).

Figure S4 Variant allele test model. AB and CD represent two independently differentiating pairs of populations. Considering all loci allele frequencies (f_l) populations AB and CD are separated by branch lengths ($i + j$) and ($k + l$) respectively. Considering only loci at which populations C and D have differing frequencies by a minimum magnitude, with C being the population of the higher frequency (∂f[C-D]), populations AB and CD are separated by branch lengths ($w + x$) and ($y + z$) respectively.

Figure S5 Fst values for 1684 SNPs between the northern and southern wild and cultivated flax populations respectively. Product moment correlation coefficient value (r), 0.098.

Figure S6 Variant allele test. Average Fst values in cultivated flax of ∂f loci subsets defined in wild populations shown by red bars. Null distributions shown in grey bars generated by random sampling 10^4 or 10^6 times equal subset sizes from all cultivated flax values. Significance values shown in Table S3.

Figure S7 Exon structure of the PEBP family in *Arabidopsis thaliana* and *TFL* homologues in *Linum usitatissimum.*

Figure S8 PCR amplicons of putative *LuTFL* targets using systems outlined in Table S4.

Figure S9 Phylogenetic tree obtained through Bayesian inference represents relationship of PEBP family genes in *Linum usitatissimum*, *Arabidopsis thaliana* and *Populus nigra*. Values by the internal nodes represent clade posterior probability. Gene prefixes indicate the organism in which locus was identified (At – *A. thaliana*, Lu – *L. usitatissimum*, Pn – *P. nigra*).

Figure S10 Conservation of amino acids in active sites of *FT*, *TFL1* and *ATC* genes and their putative homologs in flax. Asterisks indicate the potential binding pocket amino acids, in red the residues that are typical for proteins with *TFL1* activity, in green – with *FT* activity.

Figure S11 Relationship and geography of cultivated and pale flax according to *LuTFL2* data: 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, vertical dotted lines link different species within same haplotype. B. Histogram showing latitudinal gradient of

LuTFL2 alleles in cultivated flax, to the left frequency of haplotype XI (green), to the right frequency of all other haplotypes (grey) with fitted logistic regression curve (p-value of 0.000291) 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).

Figure S12 Strength of selection of *LuTFL1* III in modern and historic flaxes. A. Estimates of selection coefficient over latitude and time, where latitude correlates to time of arrival of agriculture, using the selection_time.pl algorithm (48), see methods. B. Histogram showing latitudinal gradient of *LuTFL1* alleles in historic and modern 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.

Figure S13. Comparison of Fst based on *LuTFL1* I and III differences between cultivated flax populations to the null distribution of Fst values generated from RAD loci.

Figure S14 Alignment of *LuTFL1* sequences used in this study. A. Alignment representing 15 different haplotypes of *LuTFL1* gene. Putative exons were marked with yellow highlights. B. Alignment of 5'UTR region of *LuTFL1* gene as extension of four important haplotypes (I, II, III and VIII). Promoter motifs are marked with green, cyan and purple respectively. Transcription factor binding site with putative role in flower development is marked in red.

Figure S15 Semi-quantitative RT-PCR expression analysis in leaf and meristem tissue in individuals from six different populations of pale flax. Accession ID and associated *LuTFL1* haplotypes are indicated in the top. Analysis was carried out for *GAPDH* (housekeeping gene as a positive control for constant expression), *LuTFL1*, *LuTFL2* and *LuFT* genes.

Figure S16 Amplification of *35S::LuTFL1*.III construct in *A. thaliana* plants. Presence/absence polymorphism in: A. segregating T2 population, asterisks mark lines that were characterized with late flowering. B – *tfl1-2* mutant plants and C. in positive control (vector backbone with insert in *E. coli*) and negative control (vector backbone without insert in *E. coli*).

Figure S17 Phenotypic traits of *Arabidopsis thaliana* genotypes *TFL* (Ler-0), no *TFL* (*tfl1-2*) and *LuTFL1 III* (T2). A. Days to bolting. B. Days to flower opening. C. Days to end of flowering. D Height at end of flowerinf.

Figure S18 Phenotypic traits of *Linum usitatissimum* genotypes for the *LuTFL1* locus. A-C cultivated flax, D pale flax. A. Main stem height. B. Branch number. C and D Days to flowering.

Figure S19 Phenotypic traits of *Linum usitatissimum* genotypes for the *LuTFL2* locus. A-C cultivated flax, D pale flax. A. Main stem height. B. Branch number. C and D Days to flowering.

Figure S20 Scatter plots showing pairwise association between flowering time and morphological traits in 1146 accessions of cultivated flax. Morphological traits consist of mass of 1000 seeds, fibre content, overall plant height and branching pattern displayed as a proportion of branched stem length to not branched.

Figure S21 Plots summarizing parameters of flax inflorescence simulated with *PGROWTH* model over

different base levels of *TFL* expression. Shown below charts are sampled flax architectures associated with base level of *TFL* expression equal to 10, 30, 50 and 70 from left to right. A. Chart and architecture diagrams for plants grown at N 60° latitude with assumed daylength threshold set to 14 hours. B. Chart and architecture digarams for plants grown at N 35° latitude with assumed daylength threshold set to 14 hours.

Figure S22 Architecture diagrams of flax plants simulated with pgrowth over the course of 15 days in different latitudes. Base *TFL* expression was set to 20, daylength threshold for *FT* expression was set to 14 hours.

Table S1 Pale and Cultivated flaxes used in this study.

											Genotypes	
	Accession name	Donor	Donor accession	Country of origin	Collection date	Latitude	Longitiude	Improvement status	Cultivation purpose	Latitude data	*LuTFL1*	*LuTFL2*
Pale flaxes												
¶	W042	PGRC	CN113618	Turkey	2007	37.567	27.467	wild	NA	GPS	I	I
¶	W043	PGRC	CN113630	Turkey	2007	41.183	33.733	wild	NA	GPS	III	I
	W044	PGRC	CN107295	Greece	2000	39	22	wild	NA	Centroid	II	I
¶	W045	PGRC	CN107296	Greece	2000	39	22	wild	NA	Centroid	II	I
	W046	PGRC	CN19021	France	1998	46	2	wild	NA	Centroid	VIII	III
¶	W047	PGRC	CN107257	France	1998	46	2	wild	NA	Centroid	VIII	III
¶	W048	PGRC	CN113603	Turkey	2007	41.35	36.167	wild	NA	GPS	I	X
¶	W049	PGRC	CN113606	Turkey	2007	41.383	36.183	wild	NA	GPS	I	I
¶	W050	PGRC	CN113610	Turkey	2007	37.817	29.65	wild	NA	GPS	I	III
¶	W051	PGRC	CN113621	Turkey	2007	37.033	27.367	wild	NA	GPS	I	I
¶	W052	PGRC	CN113626	Turkey	2007	41.583	35.333	wild	NA	GPS	X	III
¶	W053	PGRC	CN113629	Turkey	2007	41.217	33.233	wild	NA	GPS	III	I
¶	W054	PGRC	CN113633	Turkey	2007	41.833	31.817	wild	NA	GPS	III	nd
¶	W055	PGRC	CN113636	Turkey	2007	40.433	29.933	wild	NA	GPS	III	I
¶	W056	PGRC	CN113639	Turkey	2007	40.817	26.633	wild	NA	GPS	IX	I
¶	W057	PGRC	CN113642	Turkey	2007	40.75	39.55	wild	NA	GPS	I	XI
¶	W065	Collected	W065	Croatia	2011	43.9	16.45	wild	NA	GPS	VIII	III
¶	W066	Collected	W066	Croatia	2011	43.467	16.833	wild	NA	GPS	III	III
¶	W067	Collected	W067	Croatia	2011	43.017	17.45	wild	NA	GPS	III	I
¶	W068	Collected	W068	Montenegro	2011	42.233	18.883	wild	NA	GPS	IV	I
¶	W069	Collected	W069	Albania	2011	41.167	19.467	wild	NA	GPS	III	I
¶	W070	Collected	W070	Greece	2011	39.9	20.367	wild	NA	GPS	VI	II
¶	W072	Collected	W072	Greece	2011	38.983	21.15	wild	NA	GPS	V	II
	W074	Collected	W074	Greece	2011	39.05	21.867	wild	NA	GPS	nd	I
¶	W076	Collected	W076	Greece	2011	39.783	21.633	wild	NA	GPS	II	I
¶	W077	Collected	W077	Greece	2011	39.967	21.5	wild	NA	GPS	II	I
¶	W081	Collected	W081	Greece	2011	40.35	23.933	wild	NA	GPS	III	I
¶	W082	Collected	W082	Greece	2011	40.583	23.783	wild	NA	GPS	III	I
¶	W085	Collected	W085	Bulgaria	2011	41.583	23.717	wild	NA	GPS	VII	III
	W086	Collected	W086	Bulgaria	2011					GPS	nd	nd
¶	W094	Collected	W094	Croatia	2011	45.13	18.233	wild	NA	GPS	VIII	III
¶	W095	Collected	W095	Croatia	2011	45.367	16.267	wild	NA	GPS	VIII	III
	W096	Collected	W096	Croatia	2011	45.417	15.367	wild	NA	GPS	VIII	III
Cultivated flaxes												
¶	M056	PGRC	CN97321	Romania	1977	46§	25§	Cultivar	Oil	Centroid	nd	VI
	M057	PGRC	CN97322	Romania	1977	46	25	Cultivar	Oil	Centroid	nd	VI
	M068	PGRC	CN97044	Spain	1977	40	4	Cultivar	Oil	Centroid	nd	VI
	M021	PGRC	CN18990	Czech Republic	1999	49.5	15.5	Cultivar	Fiber	Centroid	nd	X
	M036	PGRC	CN19011	Tunisia	1999	34	9	Cultivar	Fiber	Centroid	nd	X
¶	M010	PGRC	CN98561	Turkey	1977	39	35	Cultivar	Fiber	Centroid	I	nd
¶	M043	PGRC	CN113611	Turkey	2007	39	35	Landrace	Unknown	Centroid	I	nd
¶	M051	PGRC	CN98164	Iran	1977	32	53	Cultivar	Oil	Centroid	I	nd
	M042	PGRC	CN100893	Iran	1999	32	53	Landrace	Unknown	Centroid	I	III
	M053	PGRC	CN97017	Turkey	1977	39	35	Cultivar	Oil	Centroid	I	III
	M059	PGRC	CN98278	Hungary	1977	47	20	Cultivar	Oil	Centroid	I	IV
¶	M011	PGRC	CN98562	Turkey	1977	39	35	Cultivar	Fiber	Centroid	I	V
¶	M001	PGRC	CN98833	Russia	1977	60	100	Dehiscent landrace	Unknown	Centroid	I	VI
¶	M003	PGRC	CN100852	Portugal	2002	39.5	8	Dehiscent landrace	Unknown	Centroid	I	VI
¶	M004	PGRC	CN97606	Spain	1977	40	4	Dehiscent landrace	Unknown	Centroid	I	VI
¶	M005	PGRC	CN98507	Israel	1977	31.5	34.75	Cultivar	Fiber	Centroid	I	VI
	M007	PGRC	CN97048	Iran	1977	32	53	Cultivar	Fiber	Centroid	I	VI
	M012	PGRC	CN96889	Turkey	1977	39	35	Cultivar	Fiber	Centroid	I	VI
	M013	PGRC	CN98299	Hungary	1977	47	20	Cultivar	Fiber	Centroid	I	VI
	M015	PGRC	CN98303	Hungary	1977	47	20	Cultivar	Fiber	Centroid	I	VI
	M020	PGRC	CN97326	Poland	1977	52	20	Cultivar	Fiber	Centroid	I	VI
¶	M044	PGRC	CN113613	Turkey	2007	39	35	Landrace	Unknown	Centroid	I	VI
¶	M063	PGRC	CN97176	Czech Repbulic	1977	49.5	15.5	Cultivar	Oil	Centroid	I	VI
	M067	PGRC	CN98552	UK	1977	54	2	Cultivar	Oil	Centroid	I	VI
	M037	PGRC	CN100960	Tunisia	1999	34	9	Cultivar	Fiber	Centroid	I	IX
¶	M050	PGRC	CN98163	Iran	1977	32	53	Cultivar	Oil	Centroid	I	IX
¶	M025	PGRC	CN32544	Russia	1978	60	100	Cultivar	Fiber	Centroid	I	X
	M049	PGRC	CN101028	Tunisia	1999	34	9	Landrace	Unknown	Centroid	I	X
	M029	PGRC	CN101020	Kazakhstan	1999	48	68	Cultivar	Fiber	Centroid	III	nd
	M071	PGRC	CN98880	Morocco	1977	32	5	Cultivar	Oil	Centroid	III	nd
	M017	PGRC	CN18978	Germany	1990	51	9	Cultivar	Fiber	Centroid	III	III
¶	M066	PGRC	CN98752	France	1977	46	2	Cultivar	Oil	Centroid	III	III
	M016	PGRC	CN98475	Germany	1977	51	9	Cultivar	Fiber	Centroid	III	VI
	M019	PGRC	CN97325	Poland	1977	52	20	Cultivar	Fiber	Centroid	III	VI
	M023	PGRC	CN98479	Czech Republic	1977	49.5	15.5	Cultivar	Fiber	Centroid	III	VI
	M035	PGRC	CN101025	Morocco	1999	32	5	Cultivar	Fiber	Centroid	III	VII
¶	M052	PGRC	CN113643	Turkey	2009	39	35	Cultivar	Oil	Centroid	III	IX
	M014	PGRC	CN98300	Hungary	1977	47	20	Cultivar	Fiber	Centroid	III	X
	M018	PGRC	CN18991	Poland	1999	52	20	Cultivar	Fiber	Centroid	III	X
	M024	PGRC	CN18995	Russia	1999	60	100	Cultivar	Fiber	Centroid	III	X
¶	M026	PGRC	CN101099	Russia	1999	60	100	Cultivar	Fiber	Centroid	III	X
	M027	PGRC	CN101056	Ukraine	1999	49	32	Cultivar	Fiber	Centroid	III	X
¶	M030	PGRC	CN18988	France	1999	46	2	Cultivar	Fiber	Centroid	III	X
¶	M031	PGRC	CN18982	France	1999	46	2	Cultivar	Fiber	Centroid	III	X
	M038	PGRC	CN18999	The Netherlands	1999	52.5	5.75	Cultivar	Fiber	Centroid	III	X
¶	M039	PGRC	CN100930	The Netherlands	1999	52.5	5.75	Cultivar	Fiber	Centroid	III	X
¶	M040	PGRC	CN100929	The Netherlands	1999	52.5	5.75	Cultivar	Fiber	Centroid	III	X

	M041	PGRC	CN100909	Palestine	1999	31.5	34.75	Landrace	Unknown	Centroid	III	X
¶	M045	PGRC	CN101035	Russia	1999	60	100	Landrace	Unknown	Centroid	III	X
¶	M046	PGRC	CN101046	Russia	1999	60	100	Landrace	Unknown	Centroid	III	X
	M047	PGRC	CN100922	Malta	1999	35.5	14.583	Landrace	Unknown	Centroid	III	X
	M060	PGRC	CN97758	Germany	1977	51	9	Cultivar	Oil	Centroid	III	X
¶	M061	PGRC	CN97760	Germany	1977	51	9	Cultivar	Oil	Centroid	III	X
¶	M062	PGRC	CN97332	Poland	1977	52	20	Cultivar	Oil	Centroid	III	X
	M064	PGRC	CN97344	Czech Republic	1977	49.5	15.5	Cultivar	Oil	Centroid	III	X
¶	M069	PGRC	CN98826	Egypt	1977	27	30	Cultivar	Oil	Centroid	III	X
¶	M070	PGRC	CN97439	Egypt	1977	27	30	Cultivar	Oil	Centroid	III	X
	M073	PGRC	CN30841	Kazakhstan	1976	48	68	Wild	NA	Centroid	VI	X
¶	M065	PGRC	CN18996	France	1999	46	2	Cultivar	Oil	Centroid	VIII	X
¶	M076	PGRC	CN30842	Kazakhstan	1976	48	68	Wild	NA	Centroid	XI	VI
	M022	PGRC	CN98700	Czech Republic	1977	49.5	15.5	Cultivar	Fiber	Centroid	XII	?
	M033	PGRC	CN97043	Spain	1977	40	4	Cultivar	Fiber	Centroid	XII	VIII
¶	M006	PGRC	CN98511	Israel	1977	31.5	34.75	Cultivar	Fiber	Centroid	XII	VI
	M034	PGRC	CN101031	Morocco	1999	32	5	Cultivar	Fiber	Centroid	XII	VI
	M048	PGRC	CN100896	Egypt	1999	27	30	Landrace	Unknown	Centroid	XII	VI
	M054	PGRC	CN98987	Italy	1977	43	12.833	Cultivar	Oil	Centroid	XII	VI
¶	M055	PGRC	CN98988	Italy	1977	43	12.833	Cultivar	Oil	Centroid	XII	VI
	M058	PGRC	CN18992	Hungary	1999	47	20	Cultivar	Oil	Centroid	XII	VI
	M028	PGRC	CN30862	Ukraine	1976	49	32	Cultivar	Fiber	Centroid	XII	X
	M032	PGRC	CN18989	France	1999	46	2	Cultivar	Fiber	Centroid	XII	X
¶	M074	PGRC	CN30844	Ukraine	1976	49	32	Wild	NA	Centroid	XII	X
¶	M075	PGRC	CN30846	Ukraine	1976	49	32	Wild	NA	Centroid	XII	X
	M072	PGRC	CN30852	Russia	1976	60	100	Wild	NA	Centroid	XIII	V
¶	M009	PGRC	CN98162	Iran	1977	32	53	Cultivar	Fiber	Centroid	XIV	?
¶	M008	PGRC	CN97180	Iran	1977	32	53	Cultivar	Fiber	Centroid	XV	VI
Historic flaxes												
¶	H095	VIR	VIR1727	Turkey	1926	38.833	35.633	landrace	Oil	GPS	?	?
¶	H100	VIR	VIR2458	Turkey	1927	39.9	41.283	landrace	Oil	GPS	?	?
¶	H098	VIR	VIR1931	Tunesia	1927	34	9	landrace	Intermediate	Centroid	?	V
¶	H102	VIR	VIR3688	Latvia	1930	57.15	22.517	landrace	Fibre	GPS	?	X
¶	H101	VIR	VIR2526	Italy	1927	39.217	9.1	landrace	Intermediate	GPS	I	?
¶	H107	VIR	VIR1114	Israel	1923	32.05	34.75	cultivar	Intermediate	GPS	I	III
¶	H086	VIR	VIR900	Italy	1923	43	12.833	landrace	Intermediate	GPS	I	X
¶	H083	VIR	VIR776	Belarus	1923	55.167	30	landrace	Fibre	Centroid	III	?
¶	H089	VIR	VIR929	Kazakhstan	1923	40.8	68.35	landrace	Oil	GPS	III	?
¶	H106	VIR	VIR462	Turkey	1922	39	35	cultivar	Oil	Centroid	III	?
¶	H088	VIR	VIR925	Armenia	1923	40.75	44.867	landrace	Oil	GPS	III	VI
¶	H093	VIR	VIR1693	Turkey	1926	38.5	27.5	landrace	Oil	GPS	III	VI
¶	H108	VIR	VIR3683	Latvia	1930	57.15	22.517	cultivar	Fibre	GPS	III	VI
¶	H112	VIR	VIR5546	Yugoslavia	1948	45.333	17.683	cultivar	Intermediate	GPS	III	VI
¶	H084	VIR	VIR791	Belarus	1923	53.017	30.65	landrace	Fibre	GPS	III	X
¶	H085	VIR	VIR889	Ukraine	1923	49.717	28.85	landrace	Intermediate	GPS	III	X
¶	H087	VIR	VIR901	Ukraine	1923	48.667	38.833	landrace	Intermediate	GPS	III	X
¶	H090	VIR	VIR1028	Russia	1924	57.35	28.317	landrace	Fibre	GPS	III	X
¶	H091	VIR	VIR1032	Russia	1924	57.017	29.333	landrace	Fibre	GPS	III	X
¶	H092	VIR	VIR1042	Belarus	1924	55.183	30.183	landrace	Fibre	GPS	III	X
¶	H094	VIR	VIR1723	Turkey	1926	38.833	35.633	landrace	Oil	GPS	III	X
¶	H096	VIR	VIR1859	Iran	1926	35.25	60.617	landrace	Oil	GPS	III	X
¶	H103	VIR	VIR17	Russia	1922	60	100	cultivar	Fibre	Centroid	III	X
¶	H104	VIR	VIR193	The Netherlands	1922	52.5	5.75	cultivar	Fibre	Centroid	III	X
¶	H105	VIR	VIR253	The Netherlands	1922	52.5	5.75	cultivar	Fibre	Centroid	III	X
¶	H109	VIR	VIR3933	Germany	1930	51	9	cultivar	Fibre	Centroid	III	X
¶	H110	VIR	VIR4165	France	1932	46	2	cultivar	Intermediate	Centroid	III	X
¶	H111	VIR	VIR5339	Ukraine	1938	51.683	33.683	cultivar	Intermediate	GPS	III	X
Herbaria flaxes												
	H031	HUO	HUO07	Uk	1875	54	2	Unknown	Unknown	Centroid	nd	nd
	H050	HUNG	HUNG33	Turkey	1957	37	35.333	Unknown	Unknown	approximate GPS	I	I
	H042	HUNG	HUNG07	Macedonia	1918	42	21.75	Unknown	Unknown	approximate GPS	I	V
	H051	HUNG	HUNG34	Syria	1945	35	38.75	Unknown	Unknown	approximate GPS	I	V
	H047	HUNG	HUNG12	Germany	1950	49.267	10.49	Unknown	Unknown	approximate GPS	I	X
	H002	HPW	HPW03	Poland	1968	52.167	20.8	Unknown	Unknown	approximate GPS	III	X
	H009	HPW	HPW13	Poland	1821	52	20	Unknown	Unknown	approximate GPS	III	X
	H013	HPW	HPW17	Poland	1961	52	21	Landrace	Unknown	approximate GPS	III	X
	H048	HUNG	HUNG13	France	1872	47.133	3.9	Unknown	Unknown	approximate GPS	IX	I
	H012	HPW	HPW16	Poland	1963	52.2	20.717	Landrace	Unknown	approximate GPS	XVI	X

¶ samples used in RAD analysis
§ location estimated from centroid of country of origin

Table S2. Nucleotide diversity of all flax subpopulations and varieties based on RADseq data.

Flax population	No. samples	Nucleotide diversity
Wild:	**28**	**1.89e-05**
Turkish	10	2.09e-05
Greek	11	1.68e-05
Croatian	7	1.71e-05
Cultivated:	**62**	**1.14e-05**
Fibre:	23	1.05e-05
Modern	12	1.31e-05
Historic	11	1.03e-05
Oil:	20	1.21e-05
Modern	12	1.38e-05
Historic	8	1.24e-05
Intermediate	8	1.36e-05
Landraces	7	1.49e-05
Dehiscent	4	1.50e-05

	ft	∂f(Nw-Sw) > 0.3	∂f(Sw-Nw) > 0.3	∂f(Nw-Sw) > 0.4	∂f(Sw-Nw) > 0.4	∂f(Nw-Sw) > 0.5	∂f(Sw-Nw) > 0.5
mean Fst	0.021229179	0.026124288	0.02069484	0.033327468	0.021649736	0.038403004	0.007924316
p value		0.133313331	0.576515303	0.04110411	0.124812481	0.00230023	0.237801693
Ia∂f/Iat		1.208788859	0.987098562	1.18274407	0.384909389	1.300927191	0.308692213

Table S3 Fst values between northern and southern cultivated flax populations across all and ∂f subsets of RAD loci and enrichment of and

Table S4 list of putative flowering time gene homologs identified in the flax genome.

Locus	GBrowse CDS reference	E-value when querried with Atha CDS	Specific primer sequence 5' to 3'	Genomic position
LuTFL1	Lus10004886	9e-74	GGAGAAATCGATGGGGAAAG TGGGTCGGTACTAACGCCT	scaffold1821: 167638..169735
LuTFL2	Lus10043385	4e-65	AATCGTAAAATTAACGGCGTC AACCTCGACTATGTCGTCAGA	scaffold25: 2627196..2628849
LuTFL3	Lus10020600	2e-76	TGGGGCAAAATGGGGAAA ATCGATCAATCTAGCGCCG	scaffold77: 545496..548022
LuTFL4	Lus10004884	1e-34	ATTAGGCAAAATGGGGAAAG ATCGATCAGTTTAGCGCCG	scaffold1821: 157633..160714
LuTFL5	Lus10027442	9e-30	AGTTACAACAATGGCAAGAGG TATTATTAGCGTTTTCTGGCG	scaffold96: 366083..367848
LuTFL6	Lus10005753	2e-26	AGTTACAAAAATGGCAAGAGGA TGGCGGCATATTAGCGTT	scaffold1036: 75860..77589
LuTFL7	Lus10021372	2e-64	AAAAGATGGCAGCAGGTGC TAGATCGATGTCATCGACGC	scaffold1123: 745123..747438
LuTFL8 (*LuFT*)	Lus10004452	2e-50	GTTATCGAAAATGCCAAGGG GAGGATATCATCATCGCCGT	scaffold845: 154960..158446

	Polymorphic sites		Ancestry¶
	Stotal	*Swild*	
LuTFL1	22	19	mixed
LuTFL2	21	1	Northern
LuTFL3	21	3	Southern

Table S5 Polymorphism within *LuTFL* loci[62]

¶ Ancestry of cultivated SNPs estimated from partitioning in wild diversity

Table S6 Haplotype frequencies in wild and cultivated flax

Haplotype	wild	cultivated
LuTFL1		
I	0.1935	0.2913
II	0.129	0
III	0.2903	0.5146
IV	0.0323	0
V	0.0323	0
VI	0.0323	0.0097
VII	0.0323	0
VIII	0.1935	0.0097
IX	0.0323	0.0097
X	0.0323	0
XI	0	0.0097
XII	0	0.1165
XIII	0	0.0097
XIV	0	0.0097
XV	0	0.0097
XVI	0	0.0097
LuTFL2		
I	0.5484	0.0202
II	0.0645	0
III	0.3226	0.0505
IV	0	0.0101
V	0	0.0505
VI	0	0.3030
VII	0	0.0101
VIII	0	0.0101
IX	0	0.0303
X	0.0323	0.5151
XI	0.0323	0

Table S7 Output of spatial autocorrelation tests for pale flax samples.

Locus	Mantel's z-statistic	Permutational s-score
LuTFL1	43.01*	24.27***
LuTFL2	15.35	20.60

Table S8 Linkage Disequilibrium between the *LuTFL1* and *LuTFL2* loci

	wild				cultivated			
LuTFL1/LuTFL2 Genotypes	expected	observed	D	r	expected	observed	D	r
I/I	0.1061	0.0968	-0.0067	-0.0339	0.0059	0.0109	0.0046	0.0721
I/III	0.0624	0.0333	-0.0322	-0.1788	0.0147	0.0326	0.0172	0.1734
I/IV	-	-	-	-	0.0029	0.0108	0.0078	0.1716
I/VI	-	-	-	-	0.0883	0.1304	0.0475	0.2275
I/IX	-	-	-	-	0.0088	0.0217	0.0125	0.1608
I/X	0.0062	0.0333	0.0267	0.3820	0.1500	0.0435	-0.1040	**-0.4579**
I/XI	0.0062	0.0333	0.0013	0.0184	-	-	-	-
II/I	0.0708	0.1333	0.0622	0.3730	-	-	-	-
III/I	0.1592	0.2333	0.0989	**0.4378**	-	-	-	-
III/III	0.0937	0.0333	-0.0556	-0.2618	0.0260	0.0217	-0.0066	-0.0605
III/VI	-	-	-	-	0.1559	0.0761	-0.0770	-0.3354
III/VII	-	-	-	-	0.0052	0.0109	0.0052	0.1040
III/IX	-	-	-	-	0.0156	0.0109	-0.0061	-0.0789
III/X	0.0094	0.0000	-0.0089	-0.1108	0.2651	0.4022	0.1300	**0.5203**
IV/I	0.0177	0.0333	-0.0011	-0.0126	-	-	-	-
V/II	0.0021	0.0333	0.0311	0.7167	-	-	-	-
VI/II	0.0021	0.0333	0.0322	0.7423	-	-	-	-
VI/X	0.0010	0.0000	-0.0011	-0.0356	0.0050	0.0108	0.0052	0.1061
VII/III	0.0104	0.0333	0.0222	0.2691	-	-	-	-
VIII/I	0.1061	0.0000	-0.1000	**-0.5086**	-	-	-	-
VIII/III	0.0624	0.2000	0.1333	**0.7219**	0.0005	0.0000	-0.0007	-0.0323
VIII/X	0.0624	0.0000	-0.0067	-0.0955	0.0050	0.0108	0.0051	0.1061
IX/I	0.0177	0.0333	0.0156	0.1769	0.0002	0.0109	0.0106	0.7708
X/III	0.0104	0.0333	0.0222	0.2691	-	-	-	-
XI/VII	-	-	-	-	0.0029	0.0108	0.0077	0.1704
XII/VII	-	-	-	-	0.0353	0.0652	0.0301	0.2043
XII/VIII	-	-	-	-	0.0012	0.0109	0.0096	0.2983
XII/X	-	-	-	-	0.0600	0.0430	-0.0189	-0.1179
XIII/V	-	-	-	-	0.0005	0.0108	0.0104	0.4842
XV/VI	-	-	-	-	0.0029	0.0108	0.0077	0.1704
XVI/X	-	-	-	-	0.0050	0.0108	0.0052	0.1061

grey - single accession samples give rise to variant frequencies
blue - absent genotypes
bold - notable correlation

Table S9. Tests of neutrality in cultivated and wild flax.

Locus	Sample	Tajima's D test	Fu and Li's D_2 test	Fu and Li's F test	R2 test
LuTFL1	All	-0.7254	0.8105	0.2300	0.0375*
	Cultivated	-1.4701*	-4.6342**	-3.8800**	0.0401*
	WIld	-0.1403	0.0199	-0.0361	0.1099
LuTFL2	All	2.1806	0.1114	1.0638	0.1319
	Cultivated	1.5637	-0.0399	0.6442	0.1256
	WIld	-1.4443	1.2241	0.4099	0.0708

* rejection of null hypothesis with confidence of 0.05
** rejection of null hypothesis with confidence of 0.01

Table S10 *p*¶ values of *LuTFL1* I and III positions in the RAD allele frequency spectrum

allele	p_d	p_h	p_v
I	0.00021091	0.00326645	1.53E-06
III	0.02411791	0.00461858	0.04410392

¶ p_d *p* values relative to the diagonal distribution
p_h *p* values relative to the row
p_v *p* values relative to the column

Table S11 Primers used in specific amplification of cDNA for semi-qPCR analysis and their melting temperatures.

Gene	Primers	Melting temperature
LuTFL1	AGTGATTGGGAGAGTAATAGGAG TTAGGCATGTGGGAAACAAGA	56°C
LuTFL2	AGTGATAGGGAGAGTGATTGGAG TTAGGCTTCTGAGTAACAGCAGA	59.5°C
LuFT	TCAGTTGGGATACACCGTTTC AGACGGAAGCAACAGGCG	56°C
GAPDH	AGGTTCTTCCCGCTCTCAAT CCTCCTTGATAGCAGCCTTG	58°C

Table S12: The list of seeds of pale flax accessions

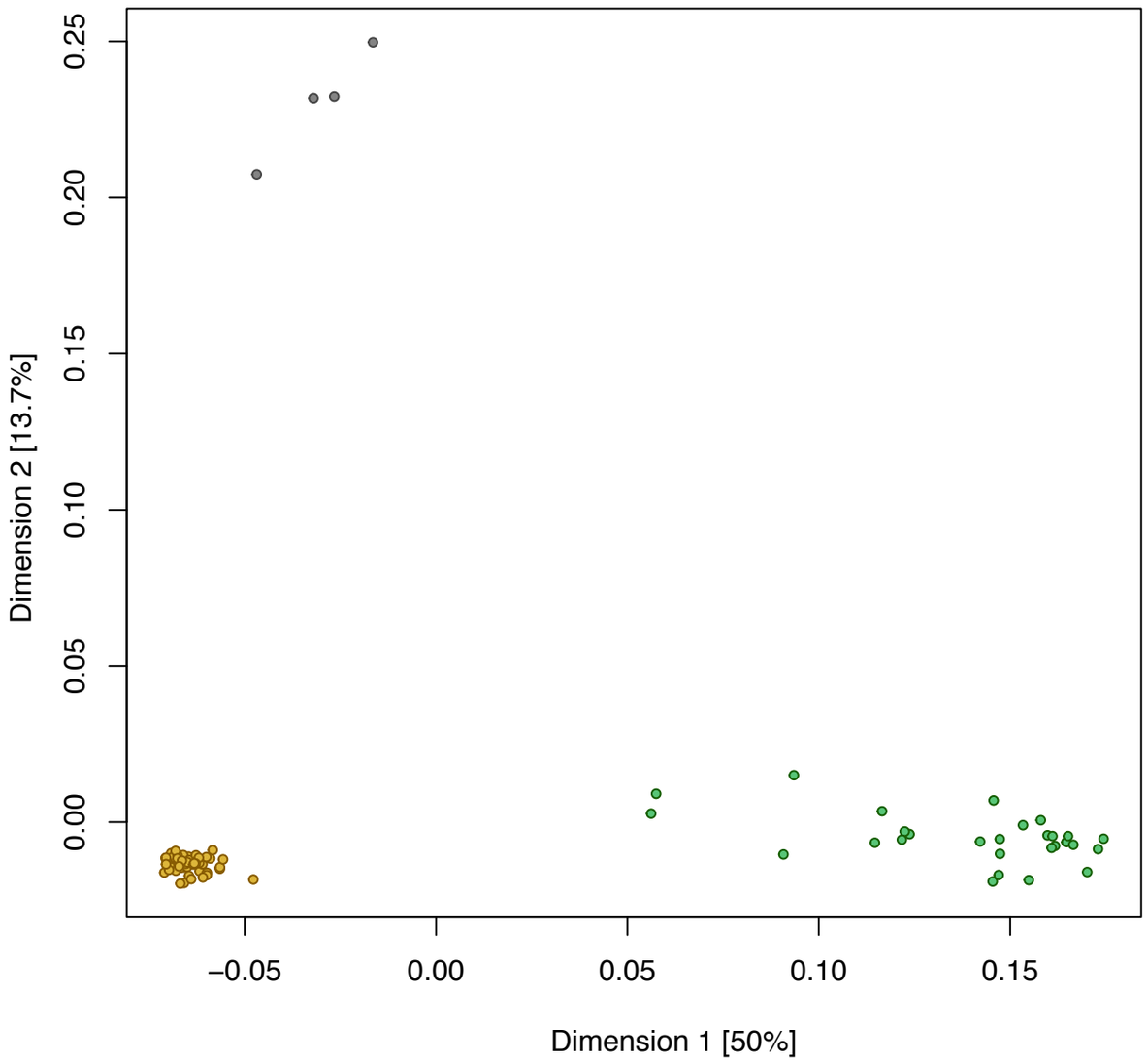
Name	Mean days to bolting	Mean days to open flowers	Mean days to end of flowering	Mean height at end of flowering
Ler-0	20.47	25.35	36.23	133
tfl1-2	14.51	20.45	23.32	69
35S::*LuTFL1*	21.33	28.02	40.33	220

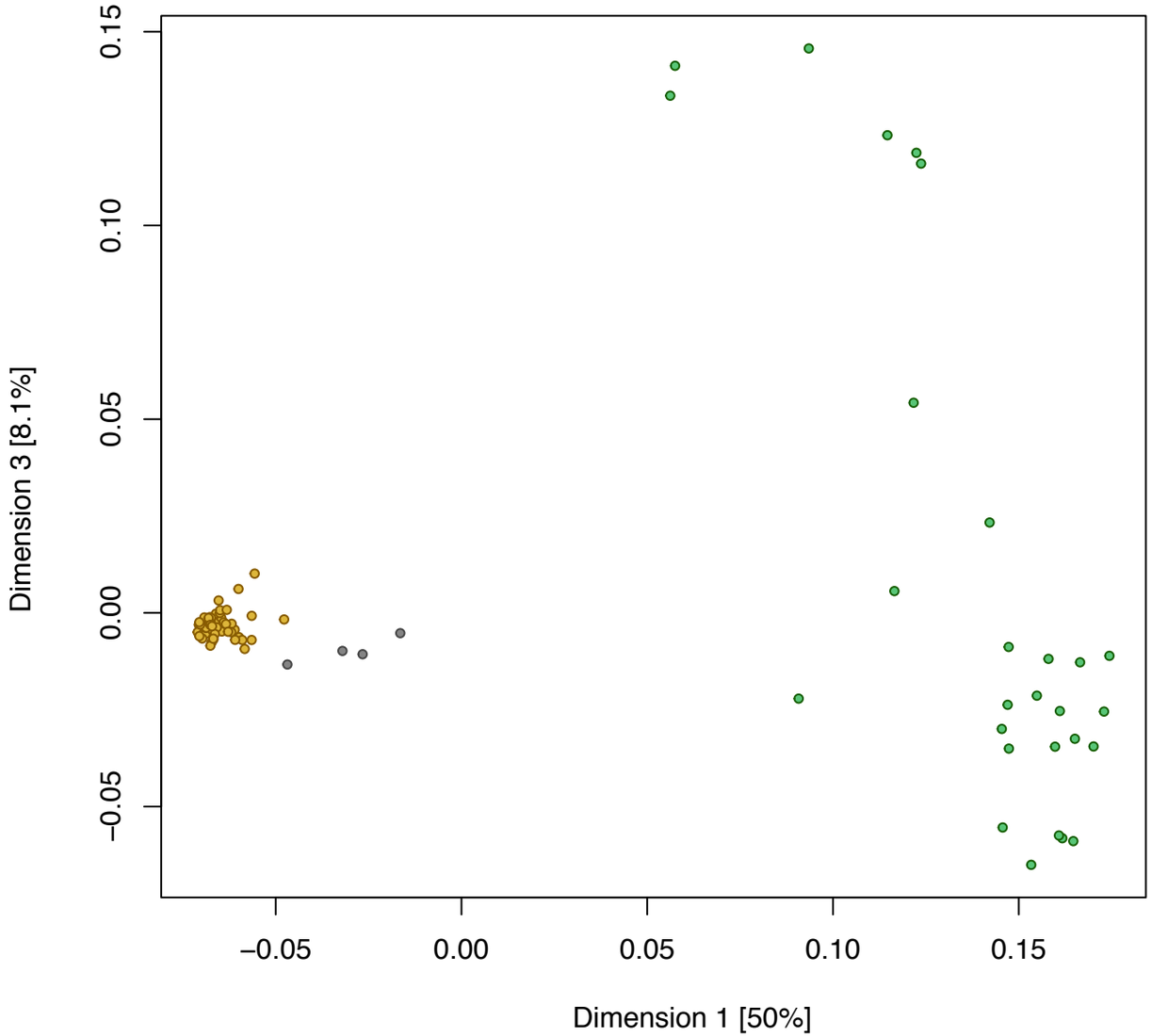
Mean for each phenotype was significantly different (with alpha level < 0.005) in all three pairwise comparisons between genotypes.

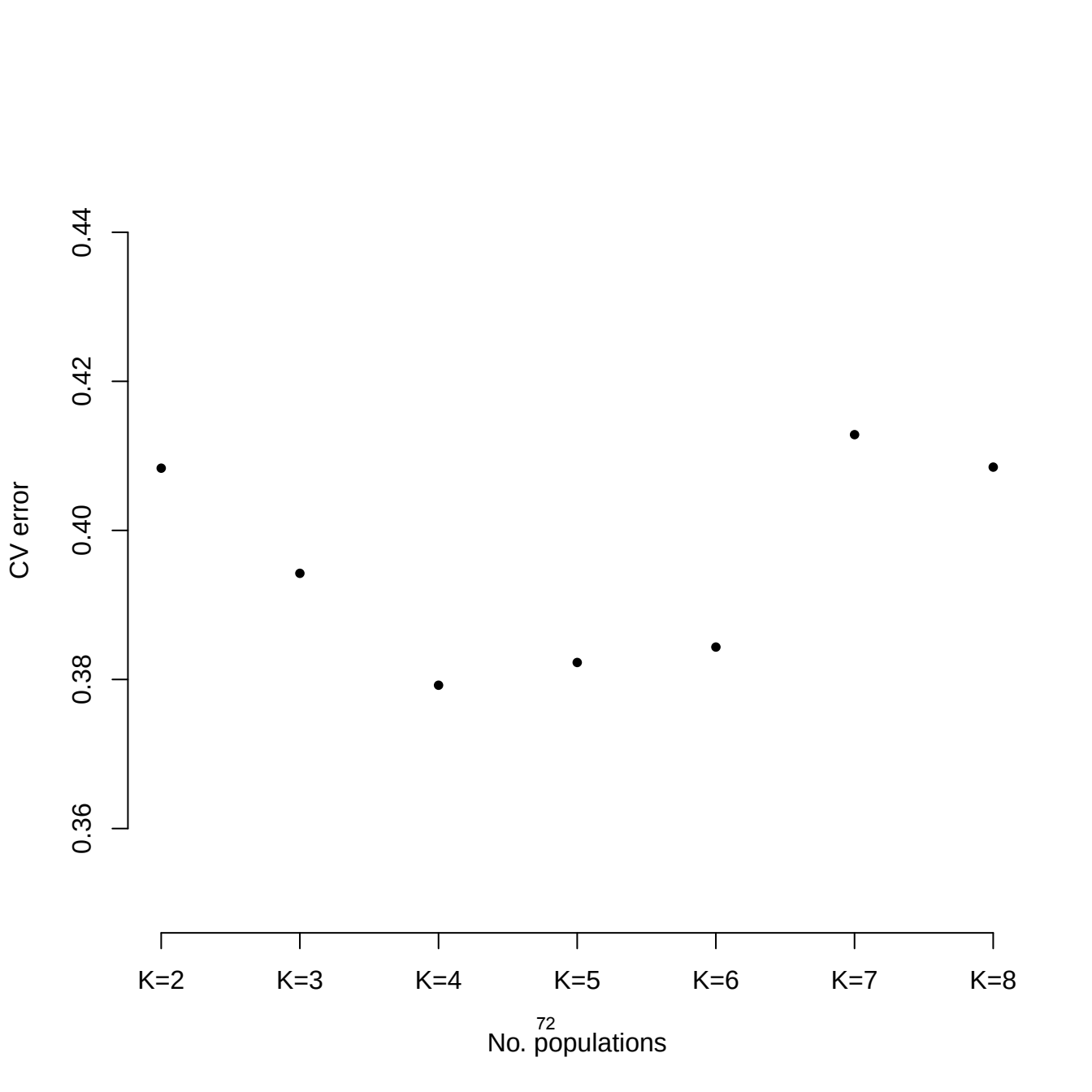
Table S13 Correlation coefficients between the start of flowering time and morphological traits in cultivated flax.

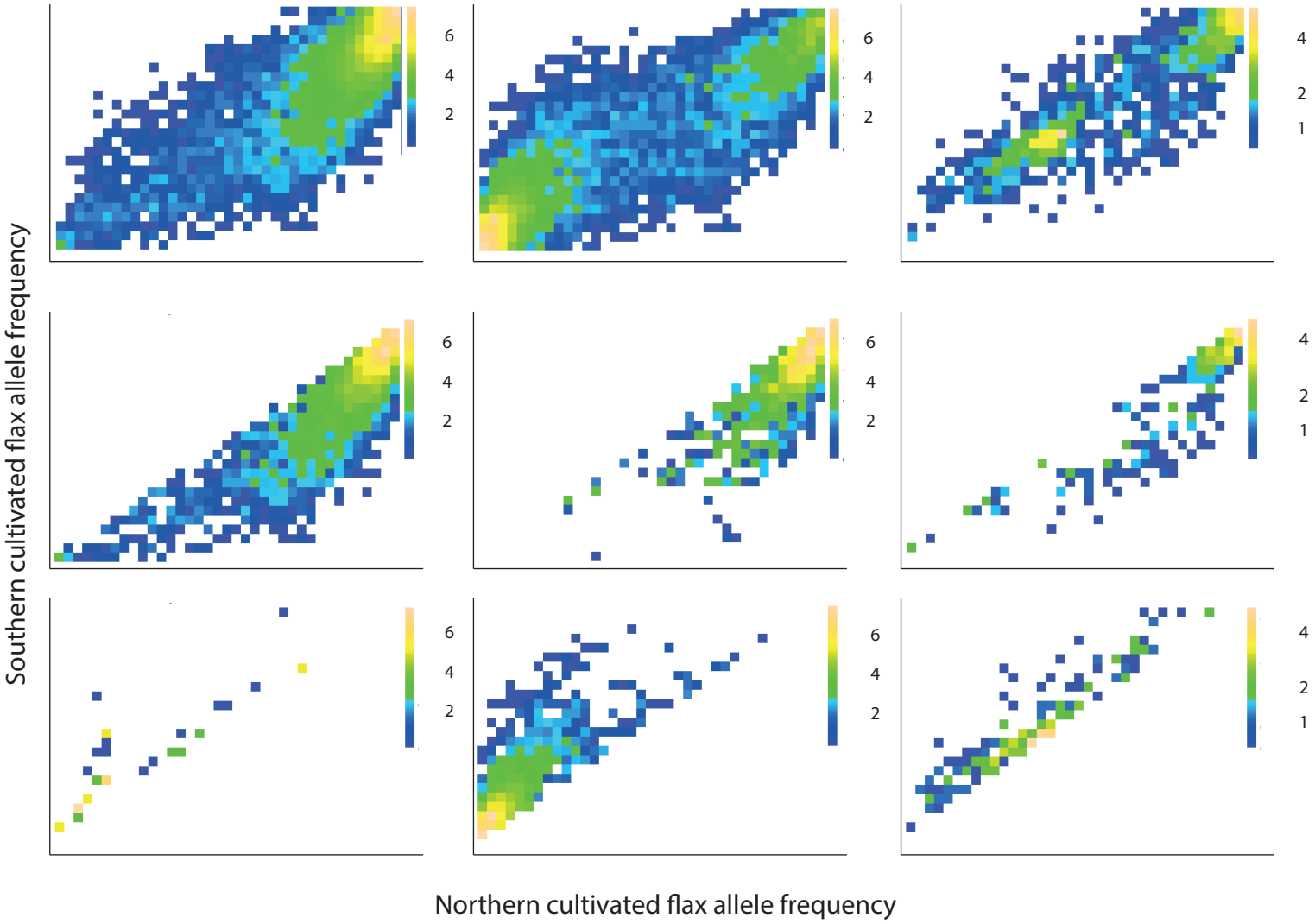
Rank correlation	Mass of 1000 seeds	Fibre content	Plant height	Branching pattern
Pearson's	-0.055	0.116***	0.166***	0.107***
Kendall's	-0.054*	0.075***	0.143***	0.161***

* result with significance level <0.05
** result with significance level <0.01
*** result with significance level <0.001


Dimension 2 [13.7%]
0.25
0.20
0.15
0.10
0.05
0.00
−0.05
0.00
0.05
0.10
0.15
Dimension 1 [50%]


Dimension 3 [8.1%]
0.15
0.10
0.05
0.00
−0.05
−0.05
0.00
0.05
0.10
0.15
Dimension 1 [50%]


CV error
0.36
0.38
0.40
0.42
0.44
K=2
K=3
K=4
K=5
K=6
K=7
K=8
No. populations



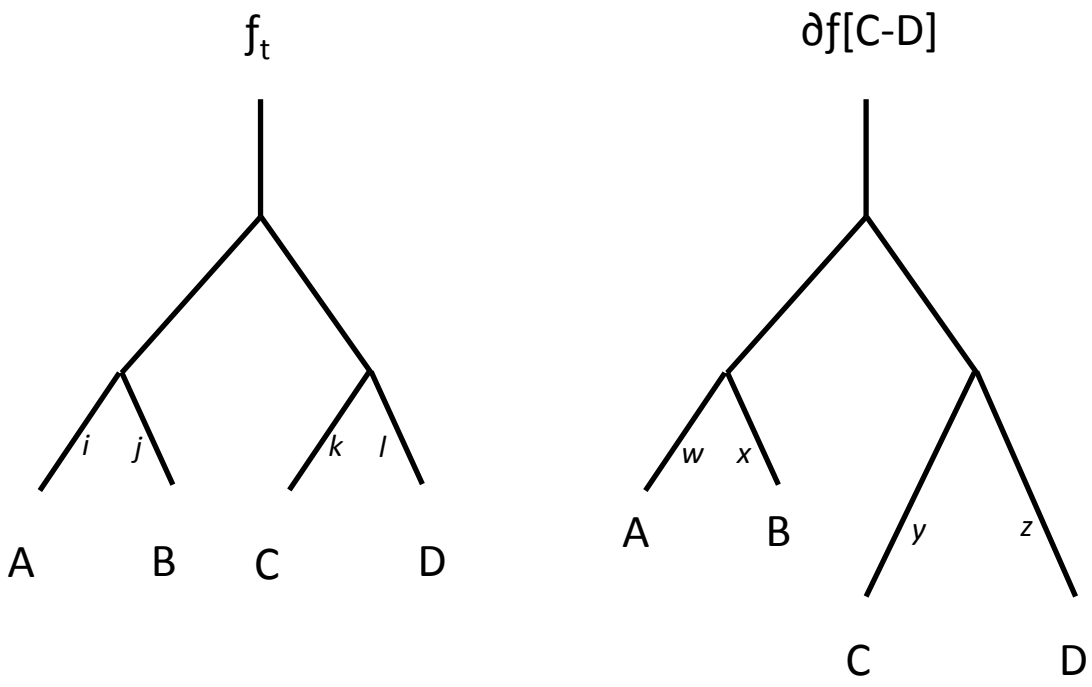
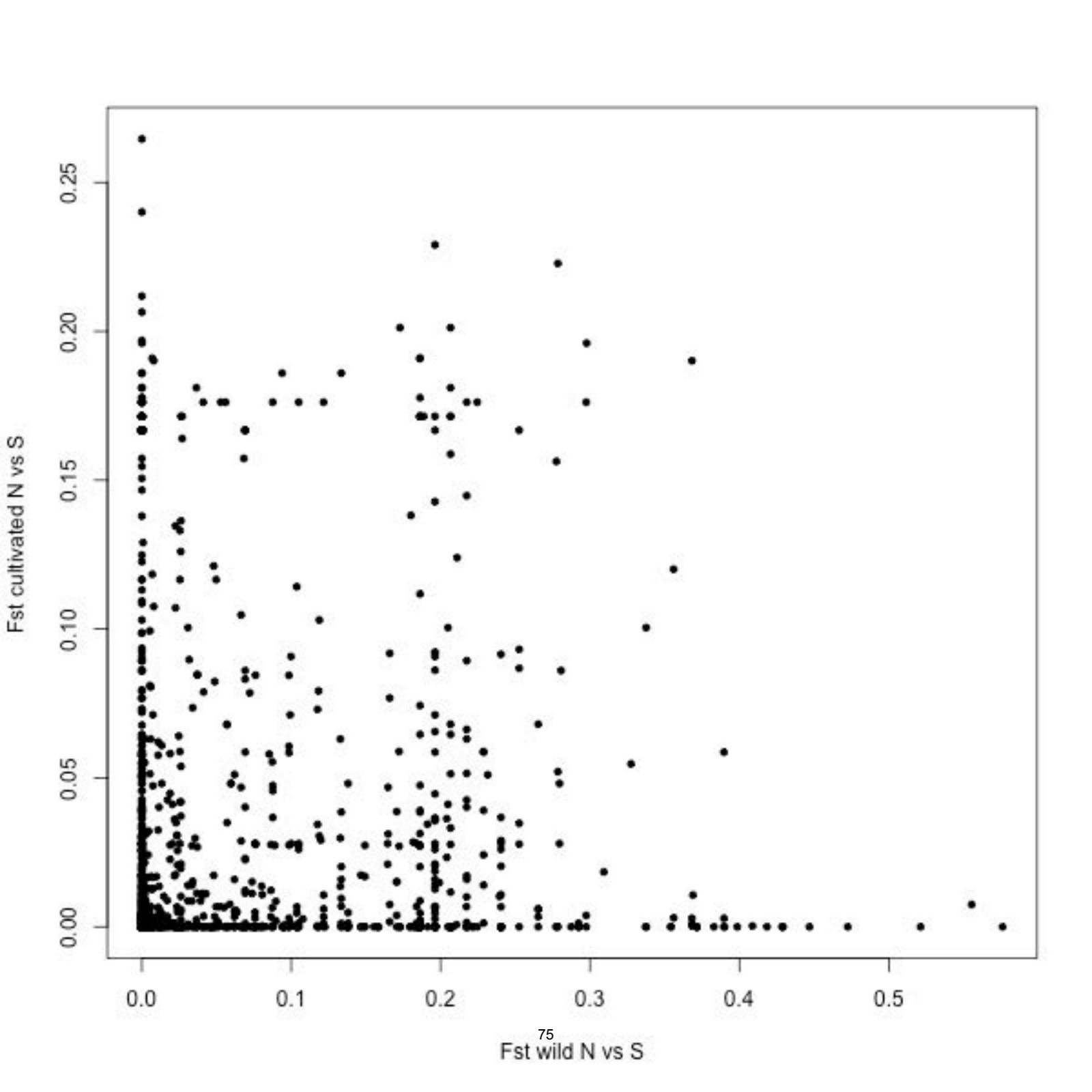
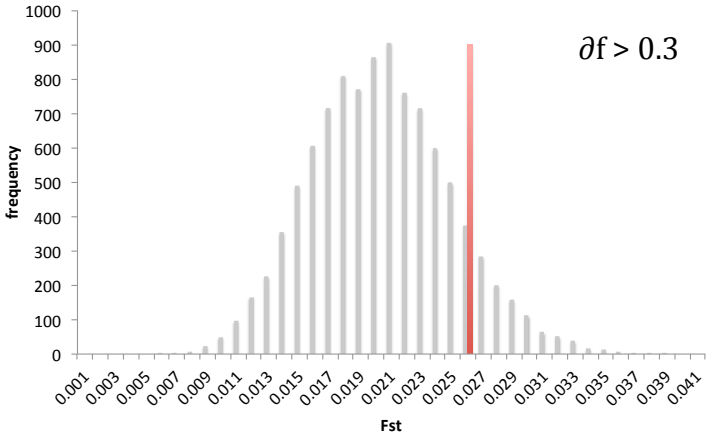
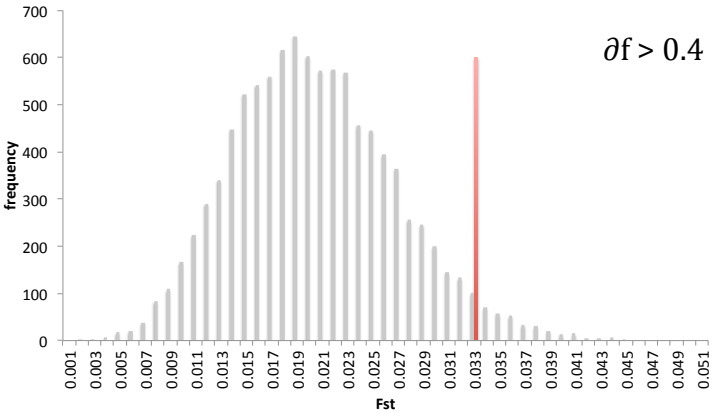


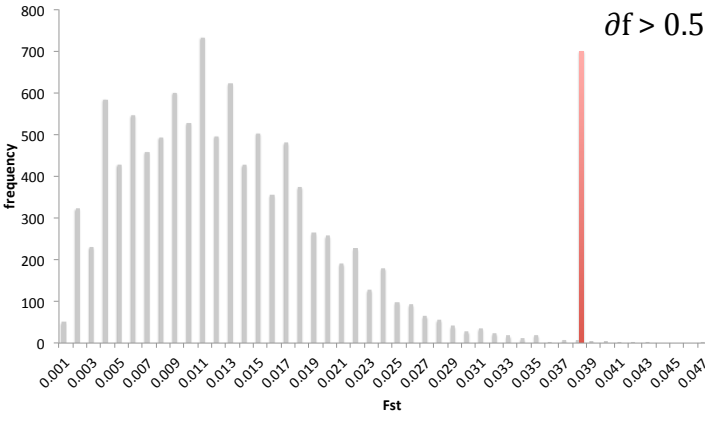
Figure S4


Fst cultivated N vs S
0.00
0.05
0.10
0.15
0.20
0.25
0.0
0.1
0.2
0.3
0.4
0.5
Fst wild N vs S

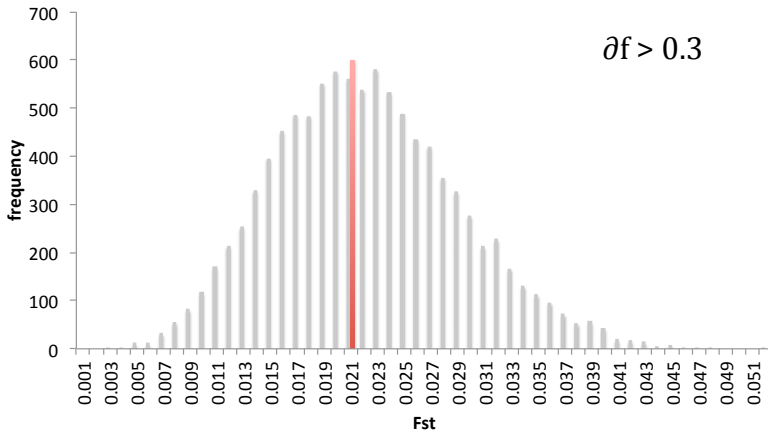
∂f (Nw – Sw)

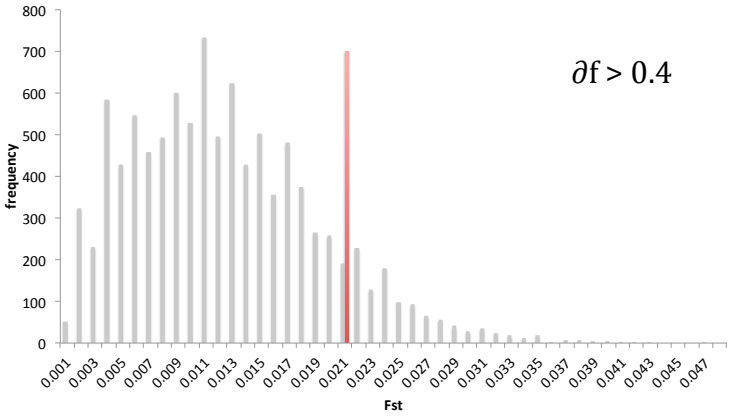

∂f > 0.3
frequency
Fst

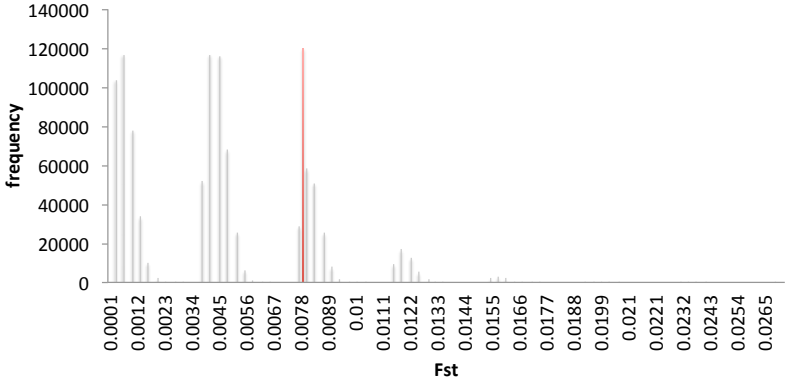

∂f > 0.4
frequency
Fst


∂f > 0.5
frequency
Fst

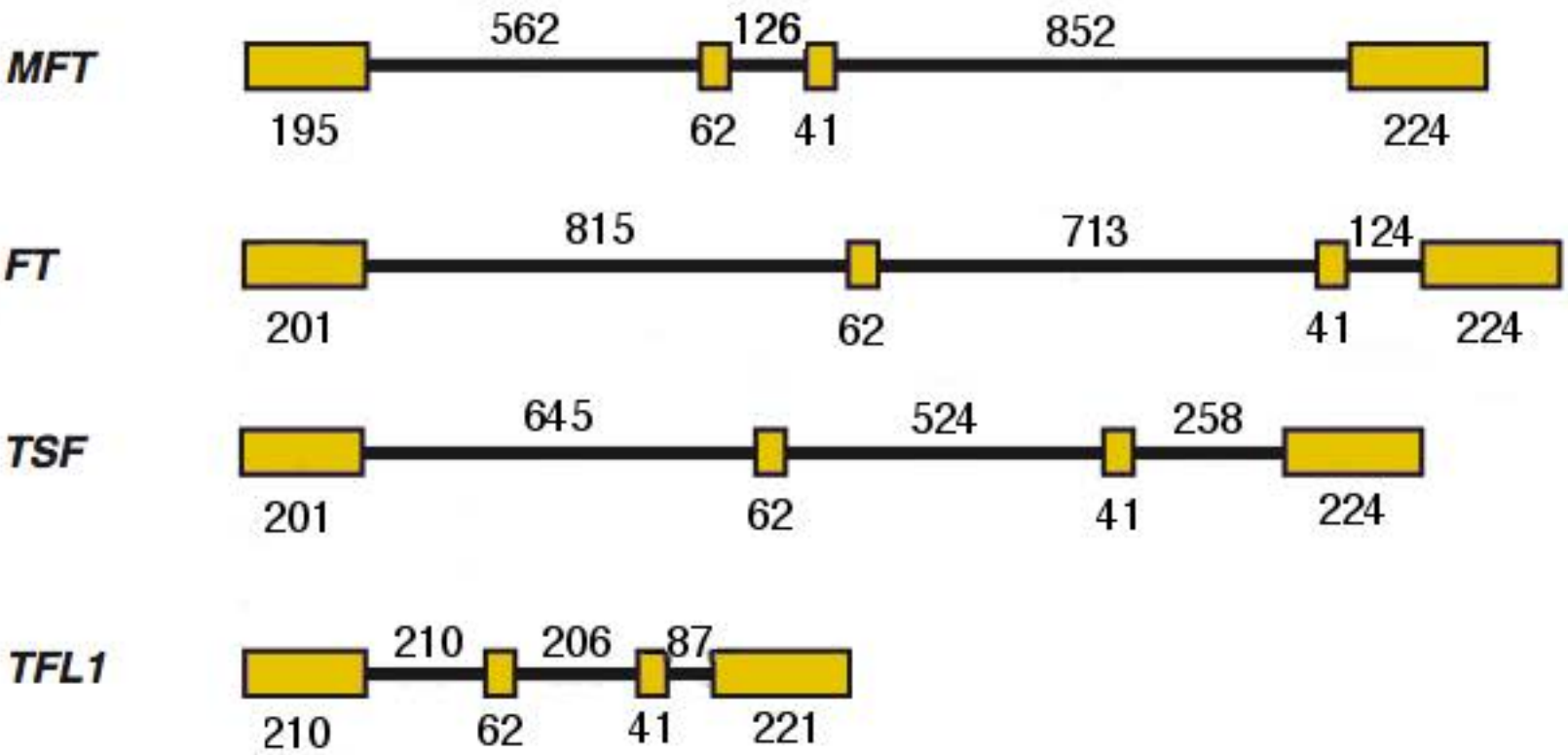
∂f (Sw – Nw)


∂f > 0.3
frequency
Fst

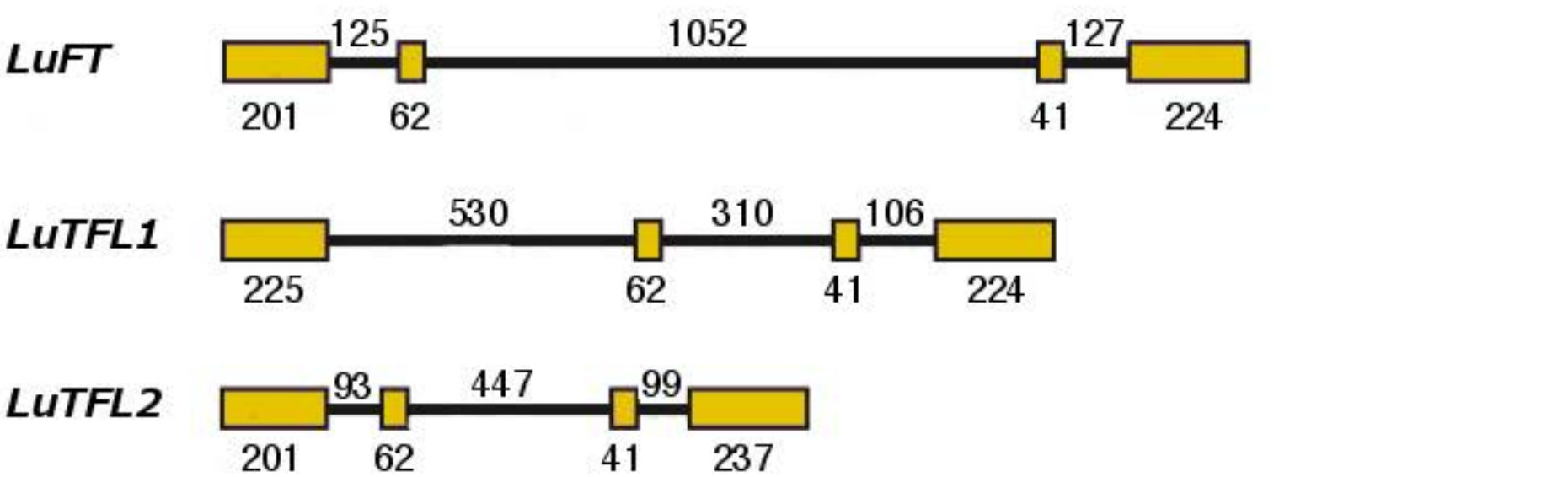

∂f > 0.4
frequency
Fst

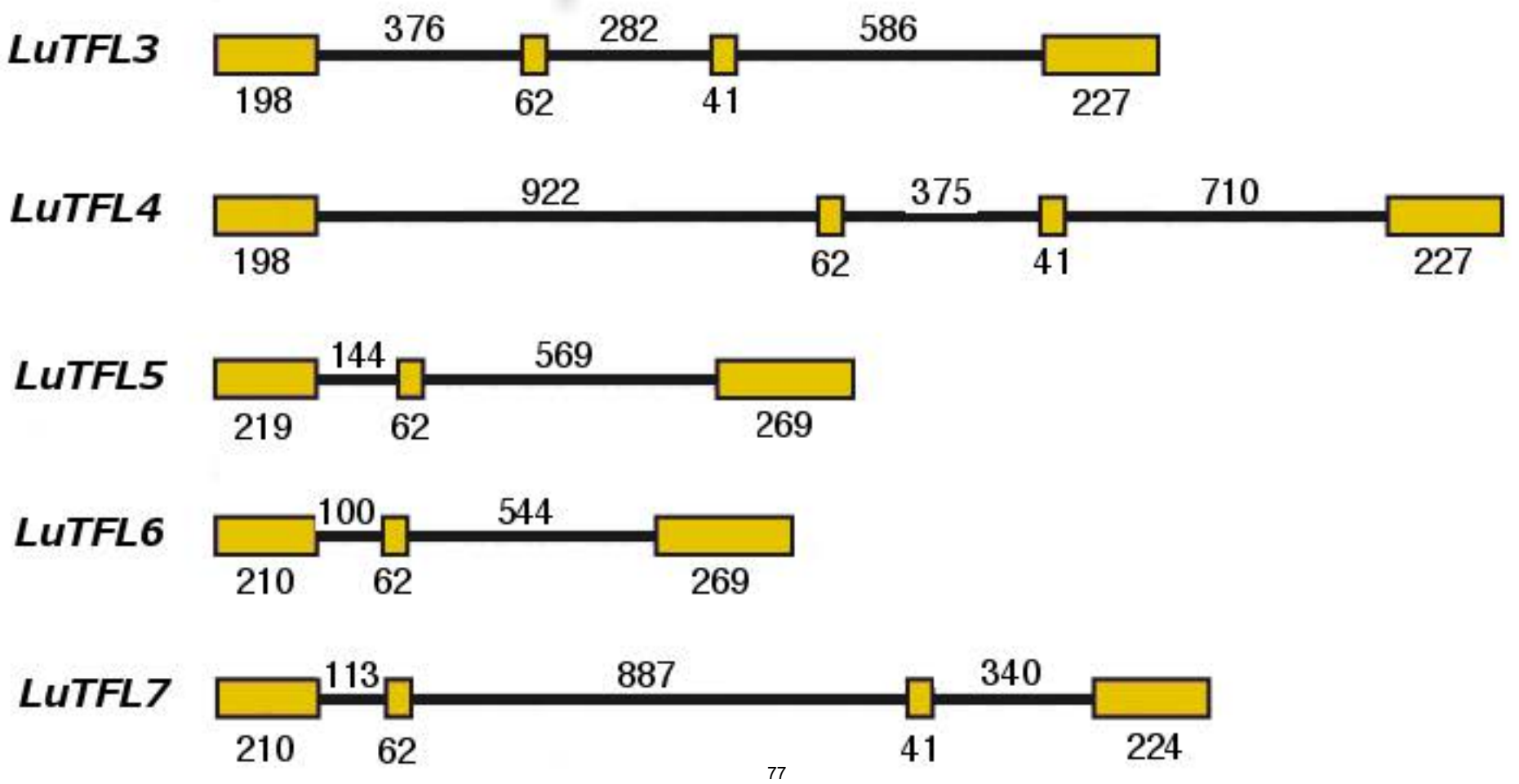

∂f > 0.5
frequency
Fst

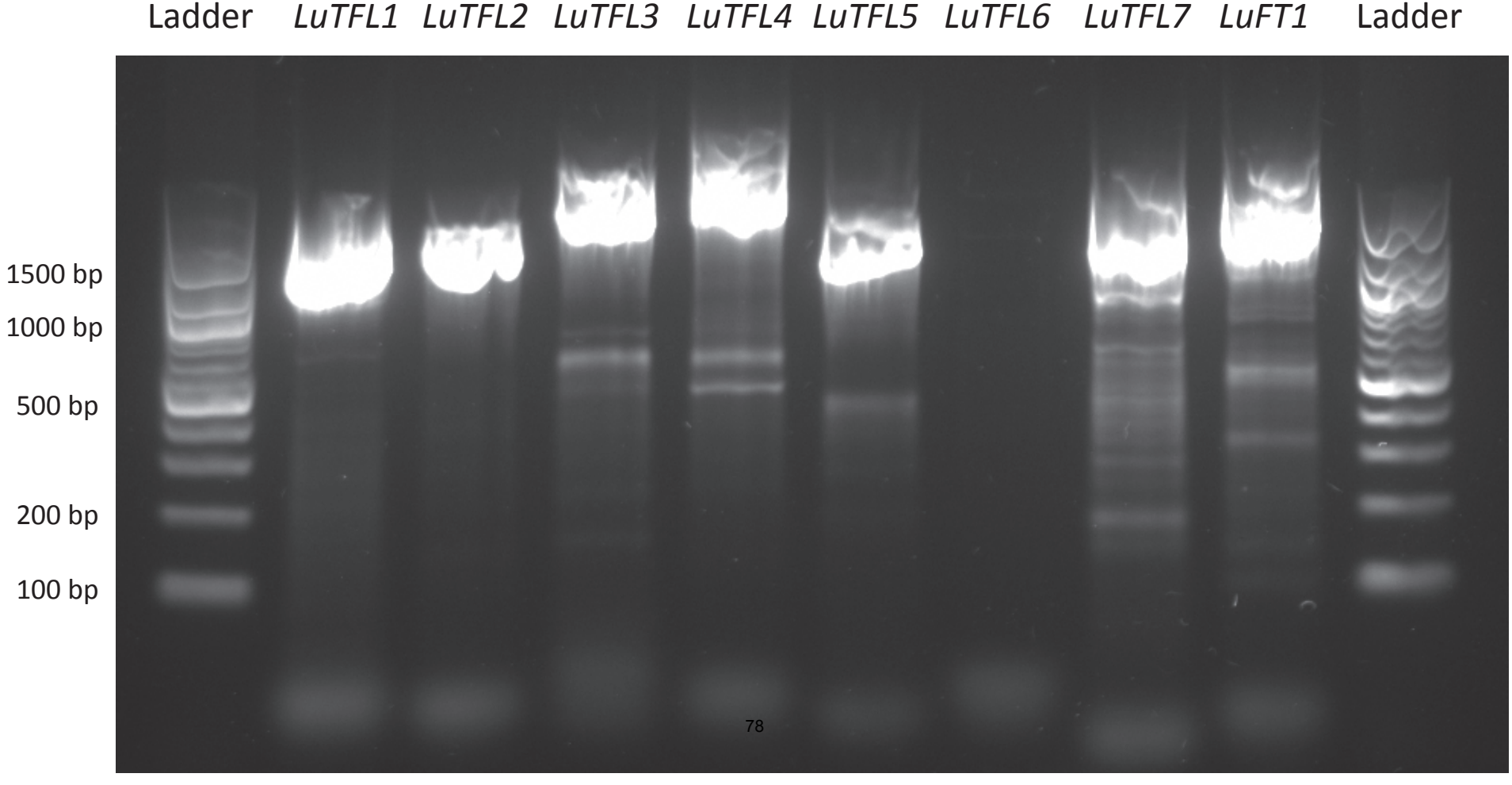
Arabidopsis thaliana

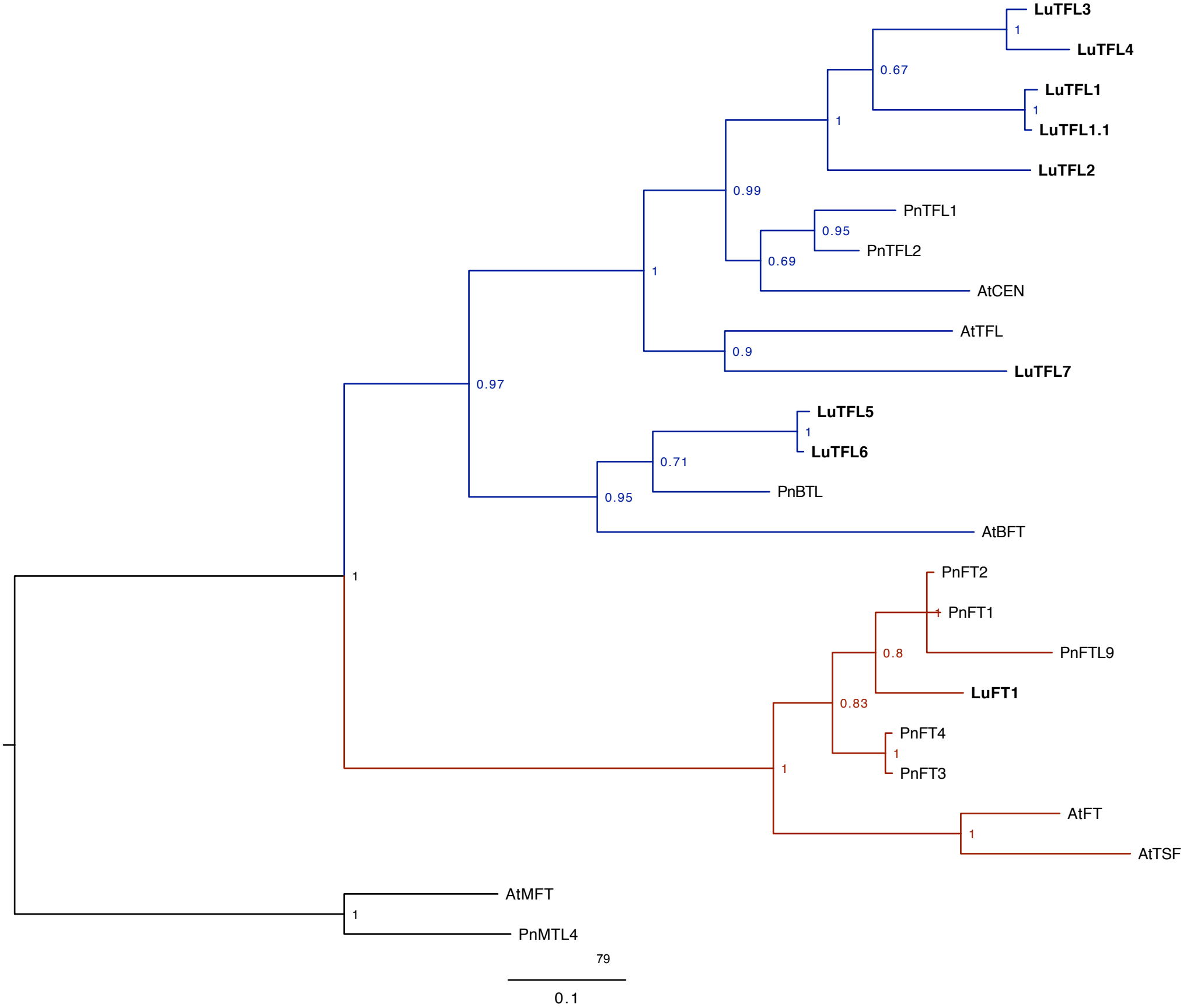


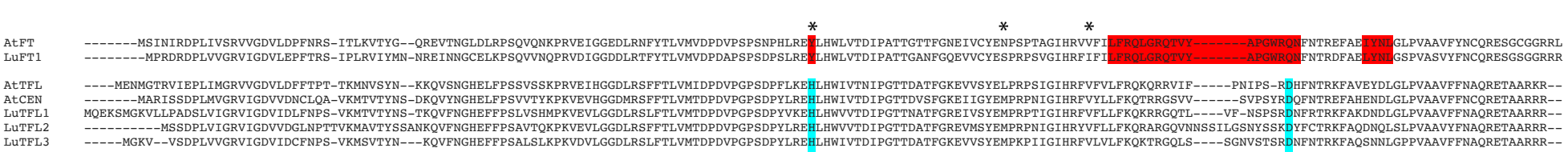
Linum usitatissimum



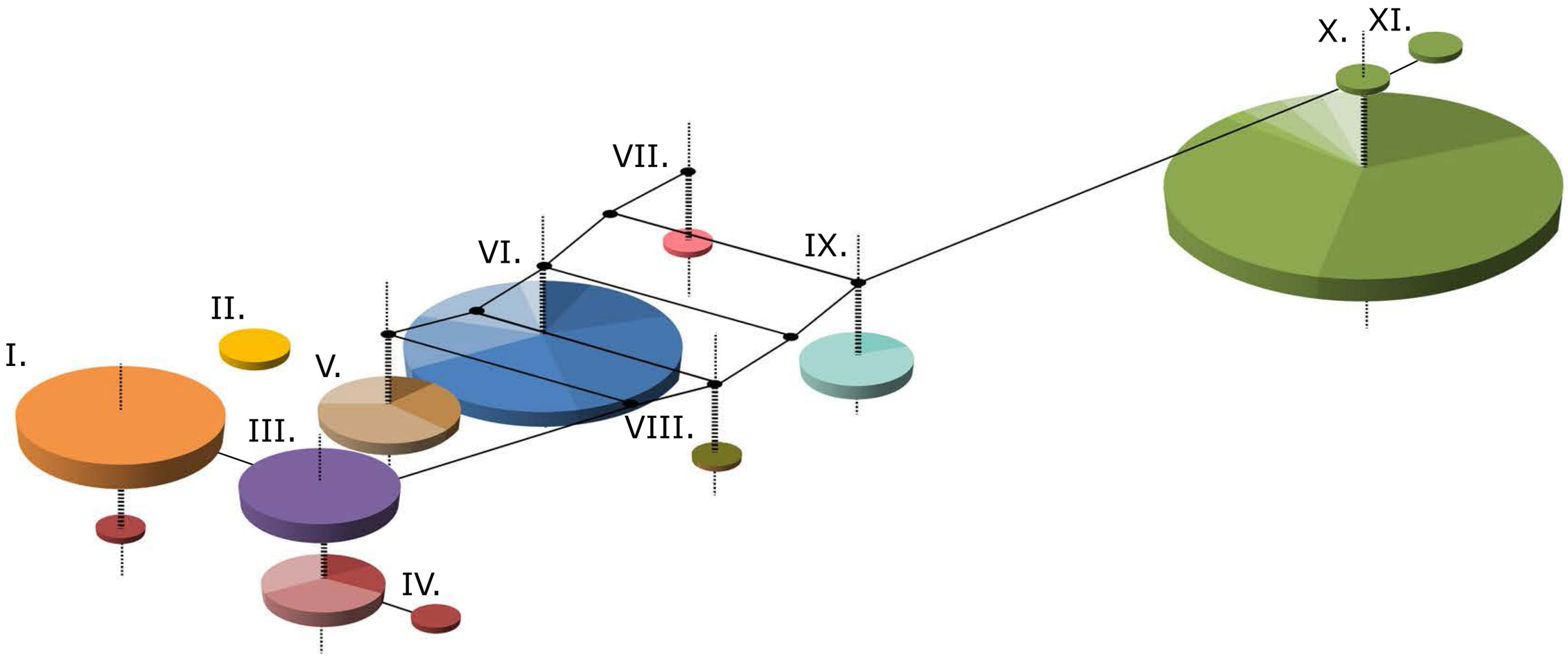



Ladder
LuTFL1
LuTFL2
LuTFL3
LuTFL4
LuTFL5
LuTFL6
LuTFL7
LuFT1
Ladder
1500 bp
1000 bp
500 bp
200 bp
100 bp

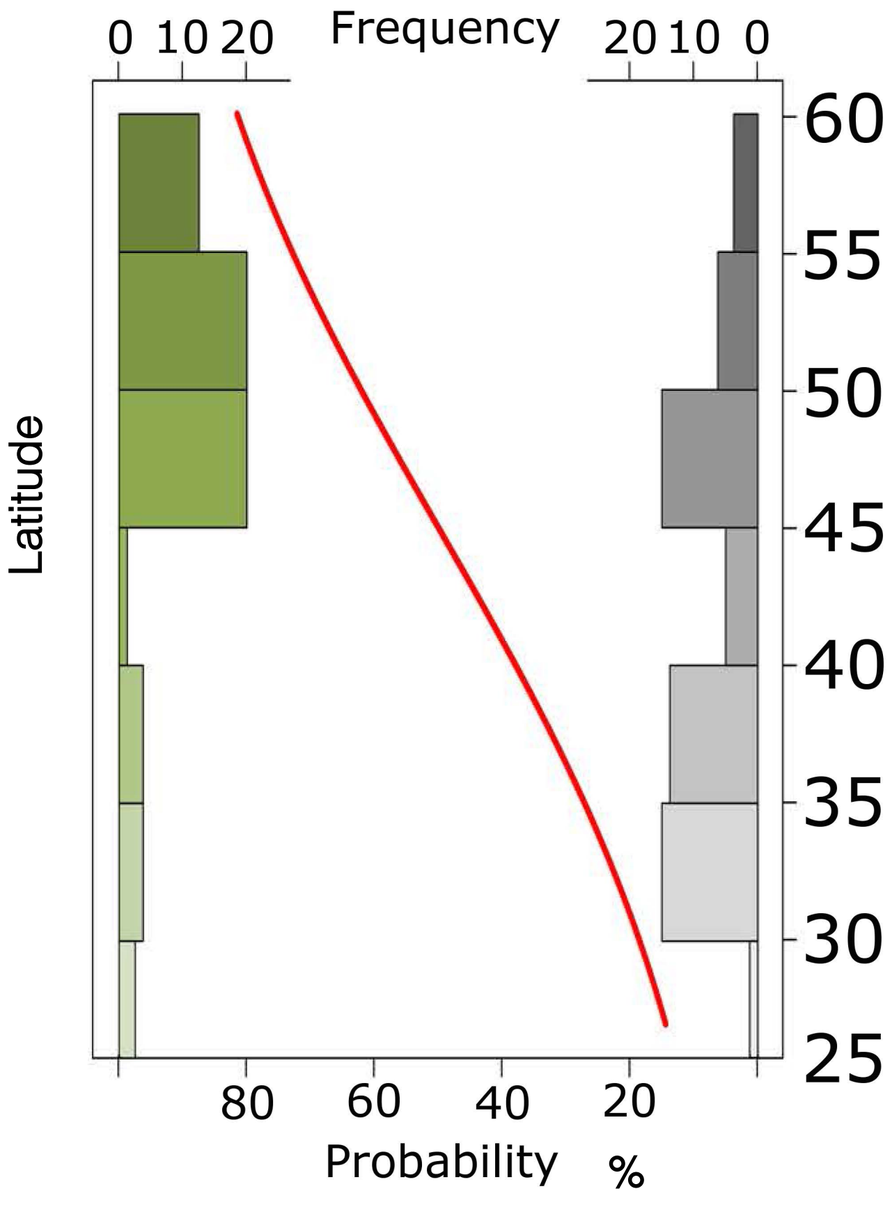

LuTFL3
1
LuTFL4
0.67
LuTFL1
1
LuTFL1.1
1
LuTFL2
0.99
PnTFL1
0.95
PnTFL2
0.69
AtCEN
1
AtTFL
0.9
LuTFL7
0.97
LuTFL5
1
LuTFL6
0.71
PnBTL
0.95
AtBFT
1
PnFT2
1
PnFT1
0.8
PnFTL9
LuFT1
0.83
PnFT4
1
PnFT3
1
AtFT
1
AtTSF
AtMFT
1
PnMTL4
79
0.1


*
*
*
AtFT -------MSINIRDPLIVSRVVGDVLDPFNRS-ITLKVTYG--QREVTNGLDLRPSQVQNKPRVEIGGEDLRNFYTLVMVDPDVPSPSNPHLREHLHWLVTDIPATTGTTFGNEIVCYENPSPTAGIHRVVFILFRQLGRQTVY-------APGWRQNFNTREFAEIYNLGLPVAAVFYNCQRESGCGGRRL
LuFT1 --------MPRDRDPLVVGRVIGDVLEPFTRS-IPLRVIYMN-NREINNGCELKPSQVVNQPRVDIGGDDLRTFYTLVMVDPDAPSPSDPSLREHLHWLVTDIPATTGANFGQEVVCYESPRPSVGIHRFIFILFRQLGRQTVY-------APGWRQNFNTRDFAELYNLGSPVASVYFNCQRESGSGGRRR
AtTFL ----MENMGTRVIEPLIMGRVVGDVLDFFTPT-TKMNVSYN--KKQVSNGHELFPSSVSSKPRVEIHGGDLRSFFTLVMIDPDVPGPSDPFLKEHLHWIVTNIPGTTDATFGKEVVSYELPRPSIGIHRFVFVLFRQKQRRVIF-----PNIPS-RDHFNTRKFAVEYDLGLPVAAVFFNAQRETAARKR--
AtCEN -------MARISSDPLMVGRVIGDVVDNCLQA-VKMTVTYNS-DKQVYNGHELFPSVVTYKPKVEVHGGDMRSFFTLVMTDPDVPGPSDPYLREHLHWIVTDIPGTTDVSFGKEIIGYEMPRPNIGIHRFVYLLFKQTRRGSVV------SVPSYRDQFNTREFAHENDLGLPVAAVFFNCQRETAARRR--
LuTFL1 MQEKSMGKVLLPADSLVIGRVIGDVIDLFNPS-VKMTVTYNS-TKQVFNGHEFFPSLVSHMPKVEVLGGDLRSLFTLVMTDPDVPGPSDPVVKEHLHWVVTDIPGTTNATFGREIVSYEMPRPTIGIHRFVFLLFKQKRRGQTL----VF-NSPSRDNFRTRKFAKDNDLGLPVAAVFFNAQRETAARRR--
LuTFL2 ----------MSSDPLVIGRVIGDVVDGLNPTTVKMAVTYSSANKQVFNGHEFFPSAVTQKPKVEVLGGDLRSFFTLVMTDPDVPGPSDPYLREHLHWVVTDIPGTTDATFGREVMSYEMPRPNIGIHRYVFLLFKQRARGQVNNSSILGSNYSSKDYFCTRKFAQDNQLSLPVAAVYFNAQRETAARRR--
LuTFL3 -----MGKV--VSDPLVVGRVIGDVIDCFNPS-VKMSVTYN---KQVFNGHEFFPSALSLKPKVDVLGGDLRSLFTLVMTDPDVPGPSDPYLREHLHWIVTDIPGTTDATFGKEVVSYEMPKPIIGIHRFVVLVLFKQKTRGQLS----SGNVSTSRDNFNTRKFAQSNNLGPPVAAVFFNAQRETAARRR--
LuTFL4 -----MGKV--VSDPLVVGRVIGDVIDGFNPS-VKMSVTYN---KQVFNGHEFFPSDLSLKPKVHVLCGDLRSVFTLVMTDPDVPGPSDPYLREHLHWIVTDIPGTTDATFGKEVVSYEMPKPIIGIHRLVLVLFKQKTRGQLS----SSNVSTSRDNFNTRKFAQCNELGPPVAAVFFNVQRETAARRR--
LuTFL5 MA--RGGGGGGGMDPLAVGRVVGEVVDVFIPS-VRMNVTYNS-NKLVCNGHELMPALISSKPRVDISGDDLRSAYTLIMTDPDYPSPSDPSLREHLHWMVTDIPGTTDASFGKEVVSYEIPRPVVGIHRYVFVLFKQARGRQTV----RAPSSSSRDCFNTRRFSEENGLGLPVAAVYFNAQRETAARKR--
LuTFL6 -----MARGGGGMDPLAVGRVVGEVVDVFIPS-VRMNVTYNS-NKLVCNGHELMPALISSKPRVDISGDDLRSAYTLIMTDPDYPSPSDPYLREHLHWMVTDIPGTTDASFGKEVVSYEIPRPVVGIHRYVFVLFKQARGRQTV----RAPSSSSRDCFNTRRFSEENGLGLPVAAVYFNAQRETAARRR--
LuTFL7 -----MAAGARALEPLIVGRVIGDVTDQFMPS-TNMTVVYAN-NKVVSNGHEFFPSAVSSRPRVDVHGGDLRTFFTLVMTDPDVPGPSDPFLREHLHWIVTDIPGTTDATFGTEVVRYEMPRPNLGIHRFVFLLYRQIQRQCLV-----TNPPSTRDYFCTRDFALLNDLGLPVAAVYFNAQRETAARRR--

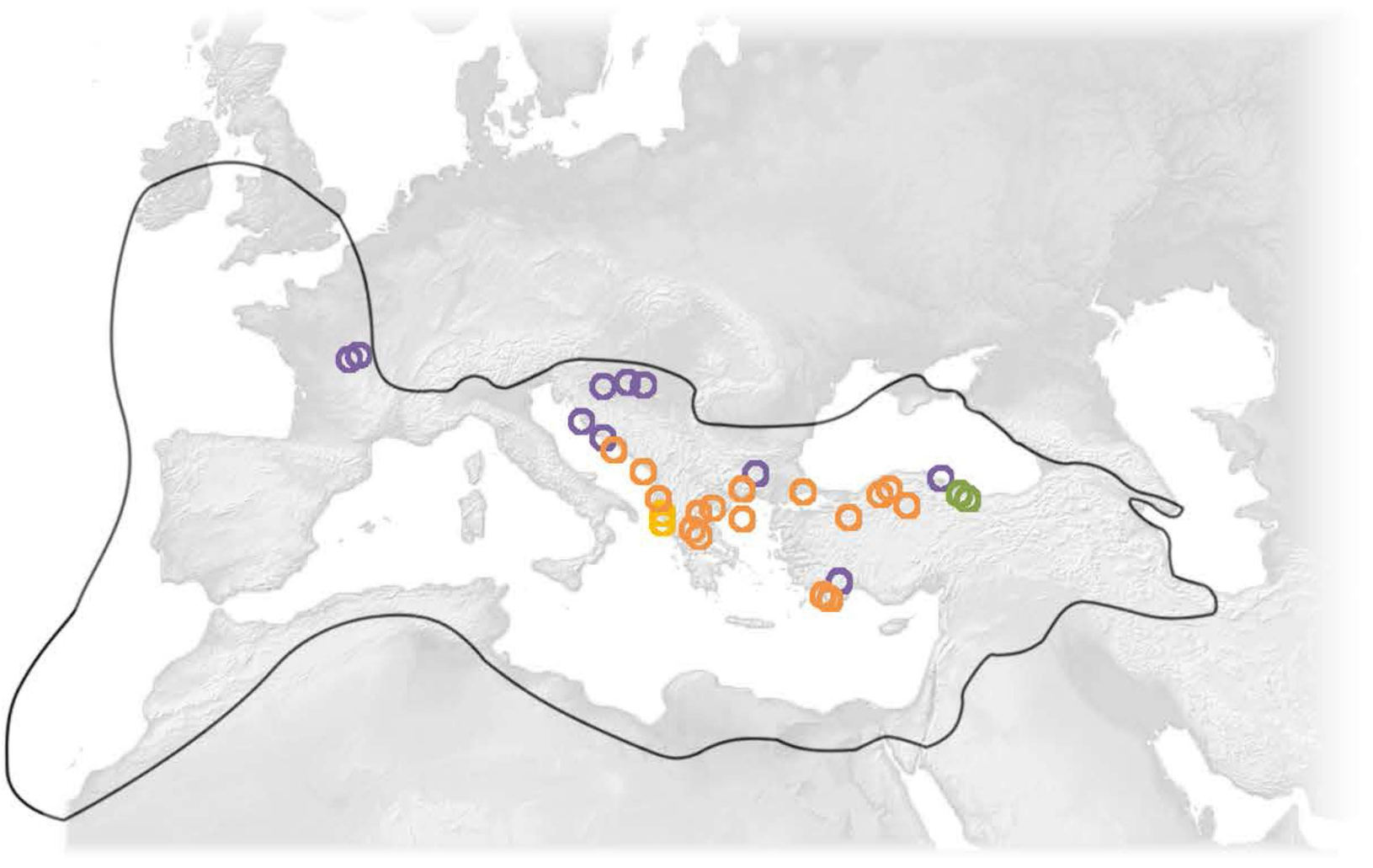
a



b

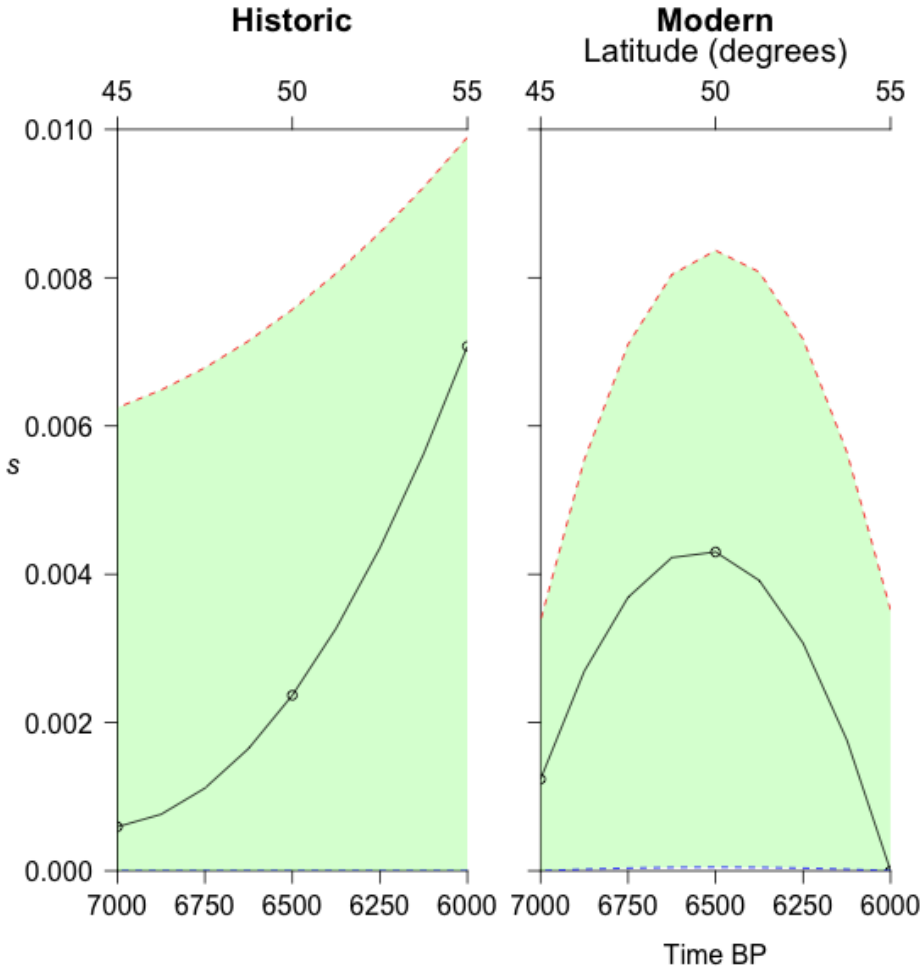


c

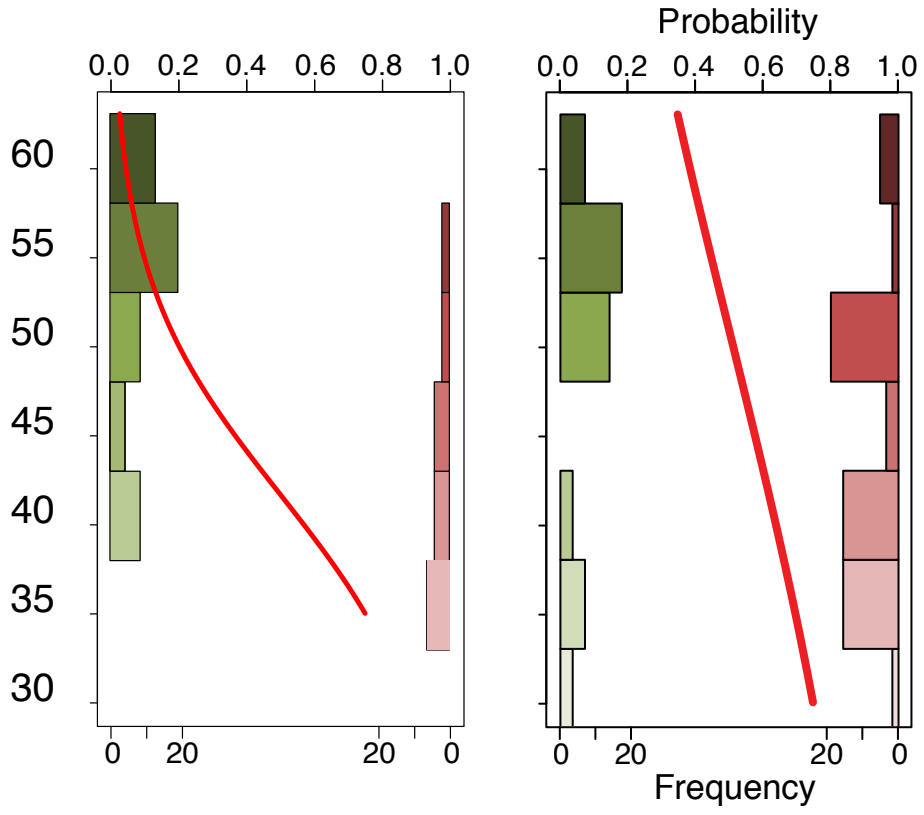


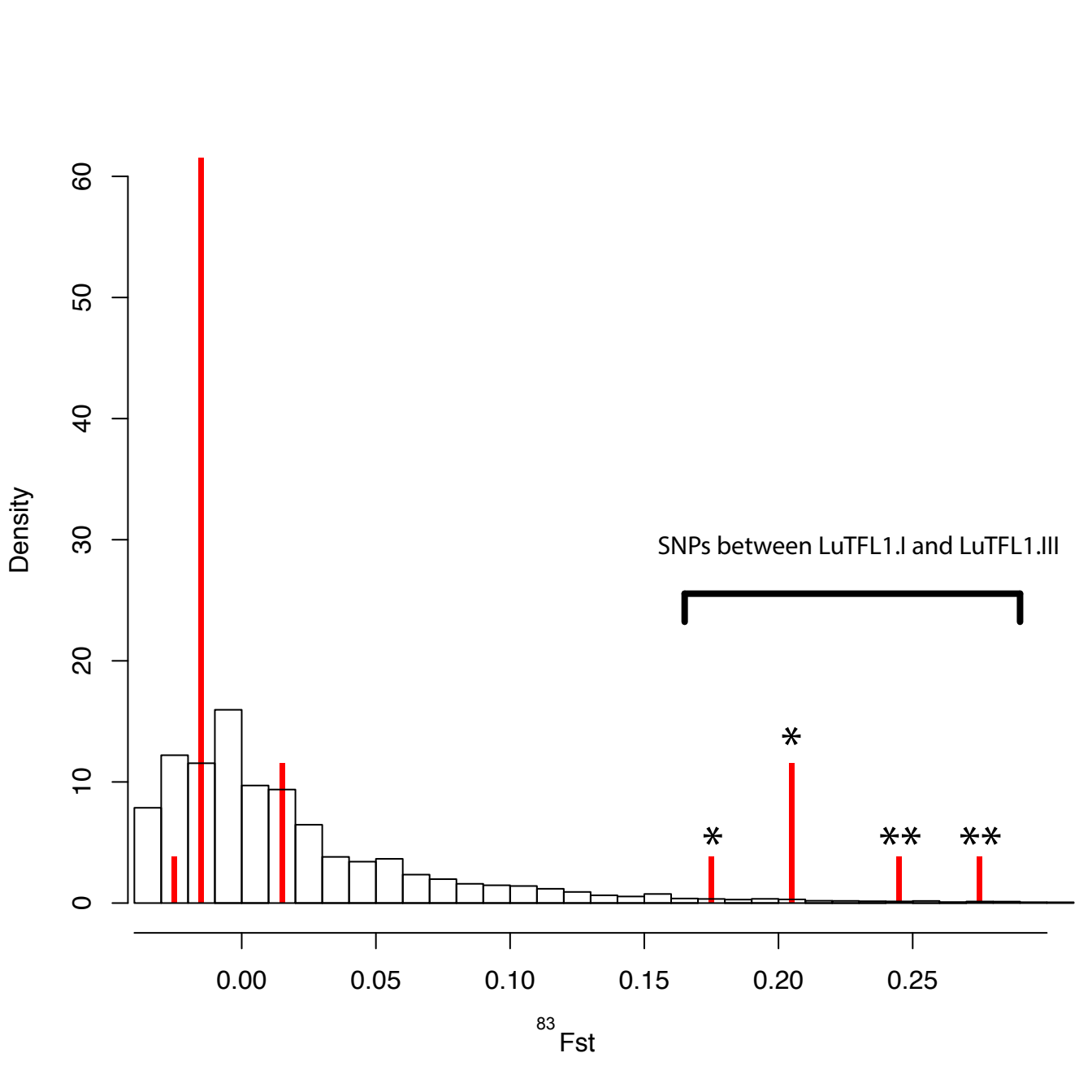
A

!

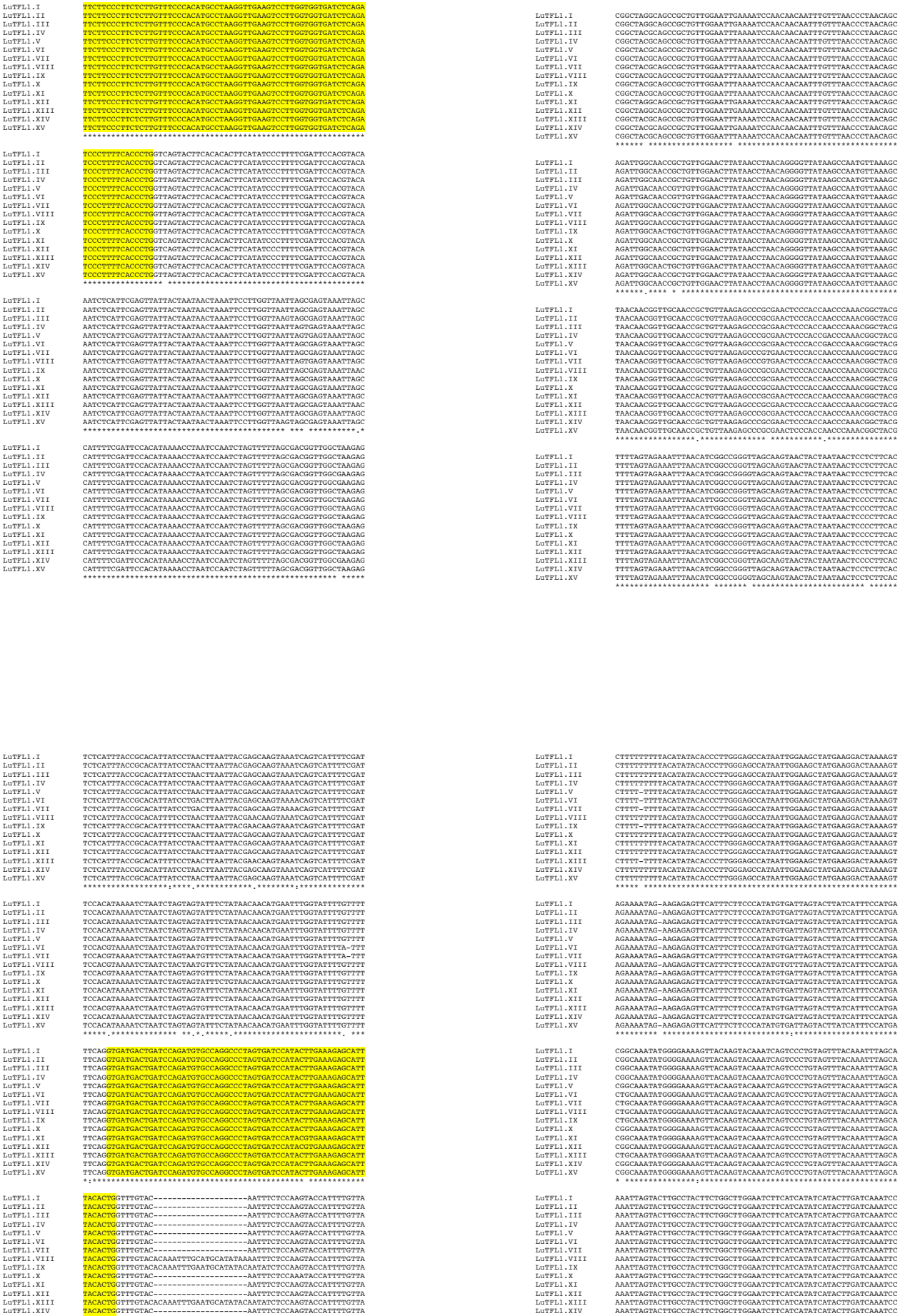

Historic
Modern
Latitude (degrees)
45
50
55
0.010
0.008
0.006
0.004
0.002
0.000
s
7000
6750
6500
6250
6000
Time BP

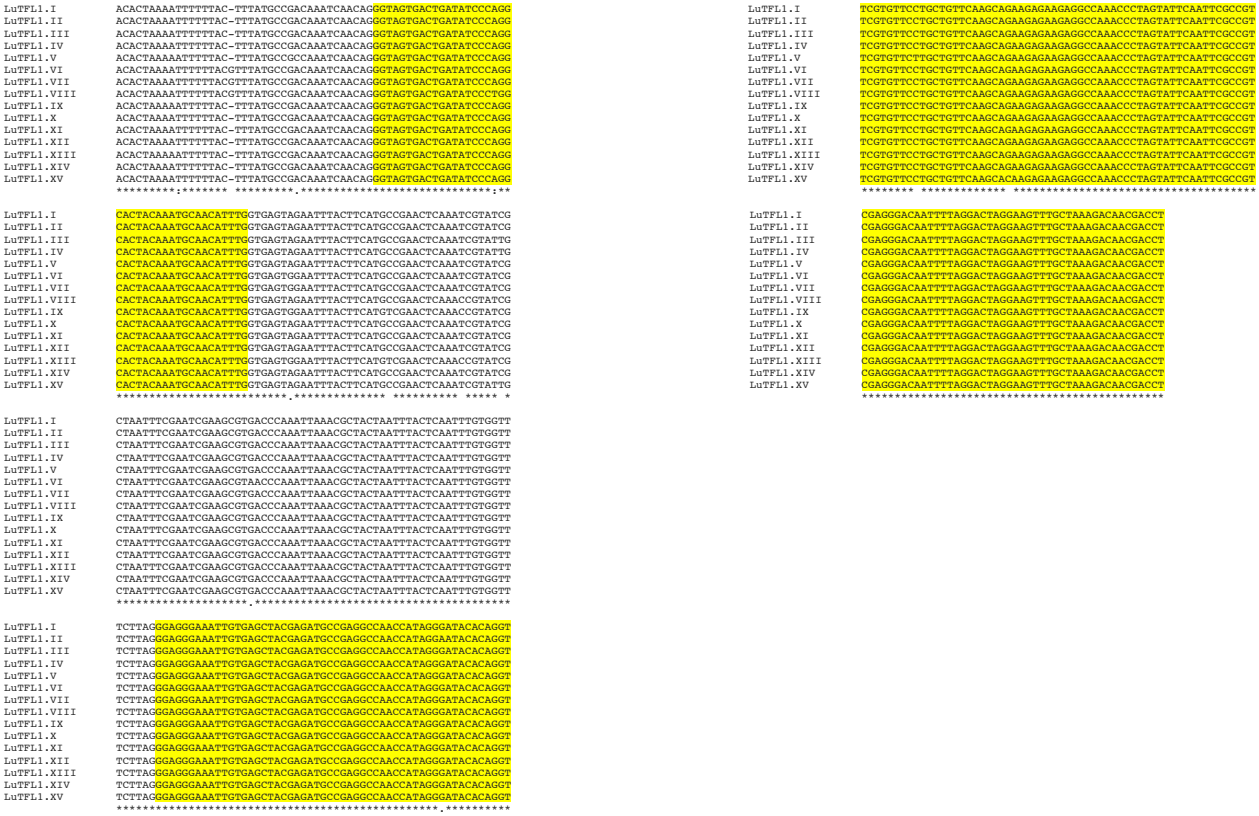
B


Probability
0.0
0.2
0.4
0.6
0.8
1.0
60
55
50
45
40
35
30
0
20
Frequency

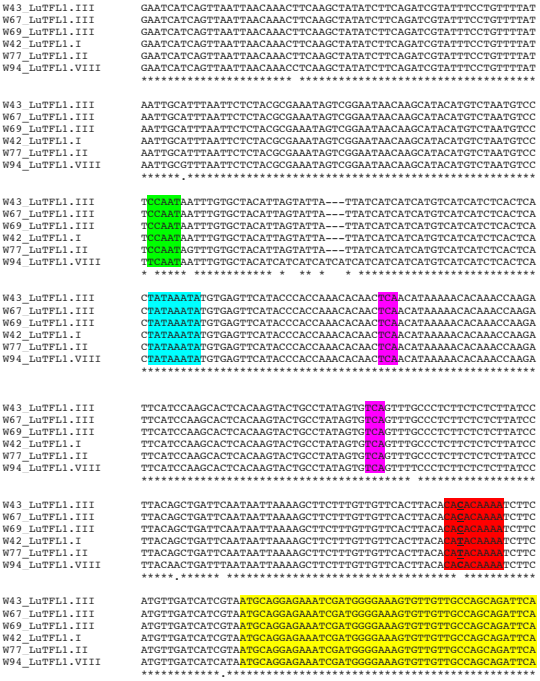

Density
0
10
20
30
40
50
60
SNPs between LuTFL1.I and LuTFL1.III
*
*
**
**
0.00
0.05
0.10
0.15
0.20
0.25
Fst

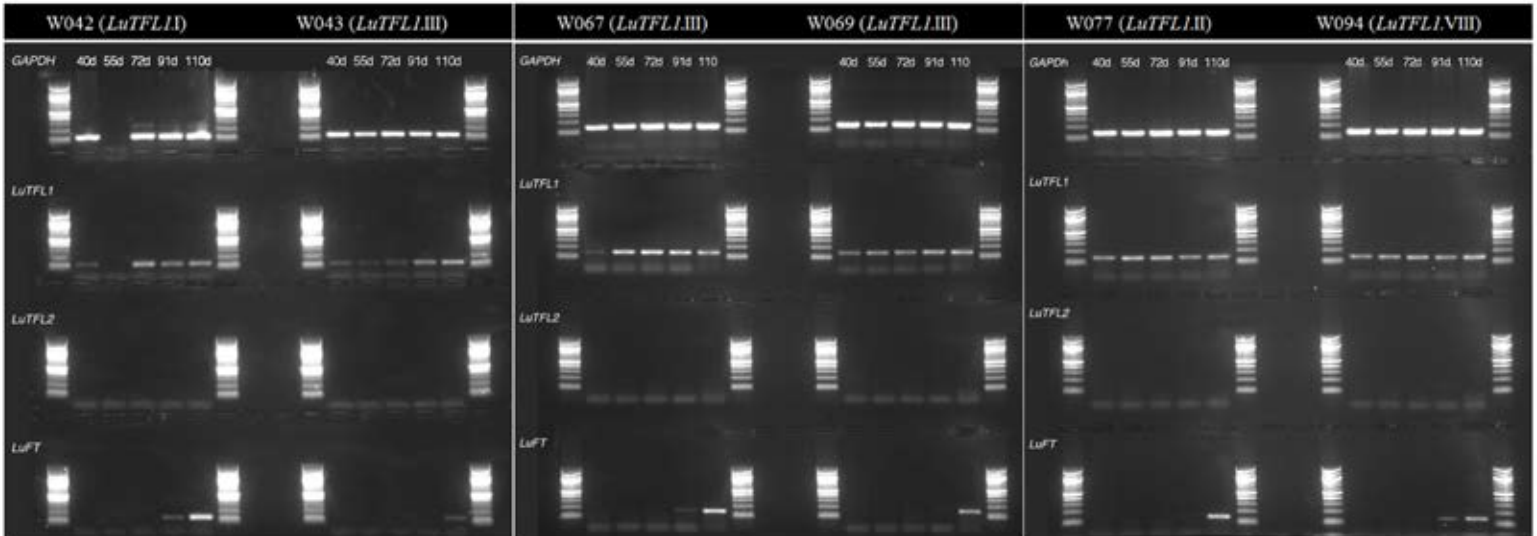
A.

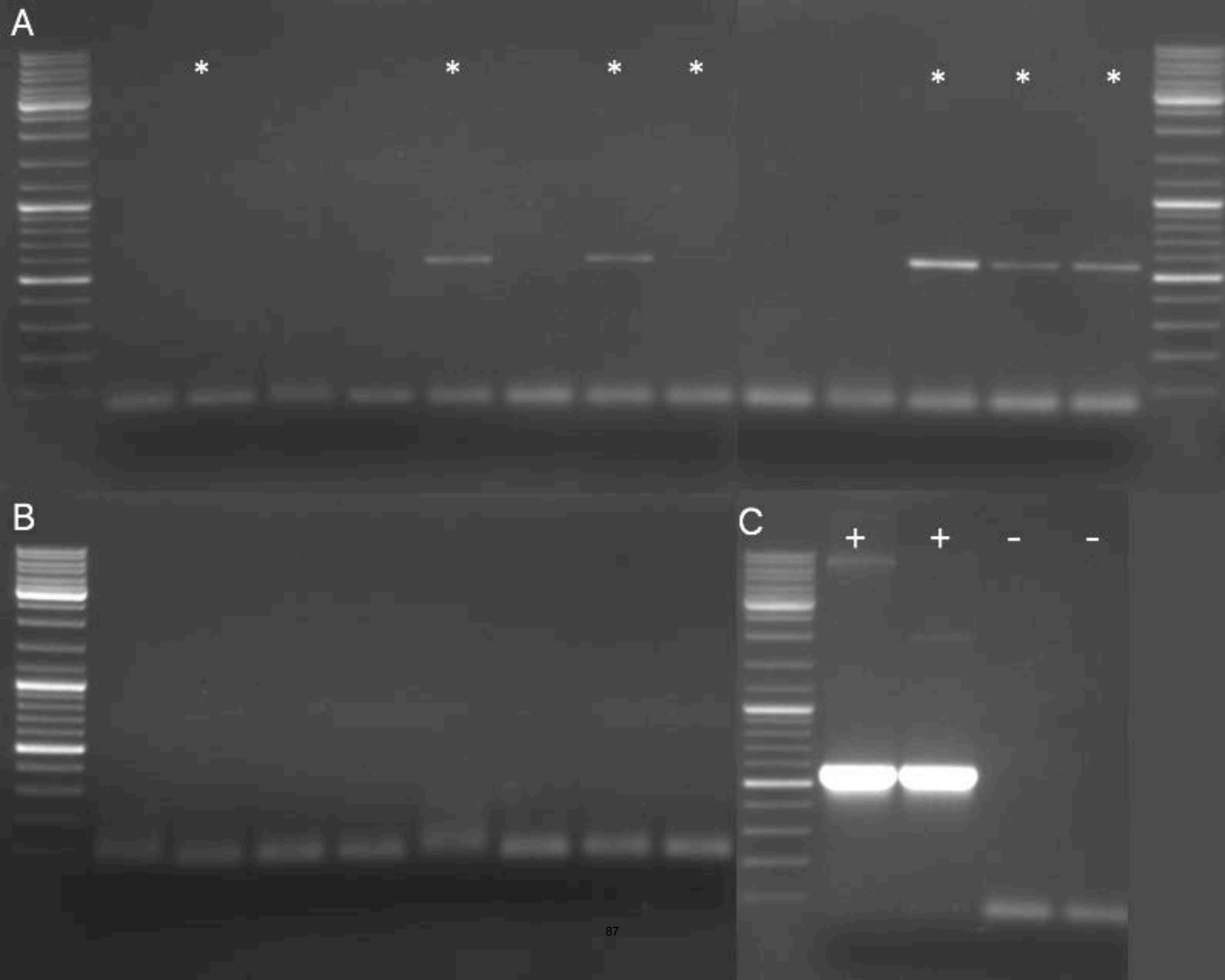


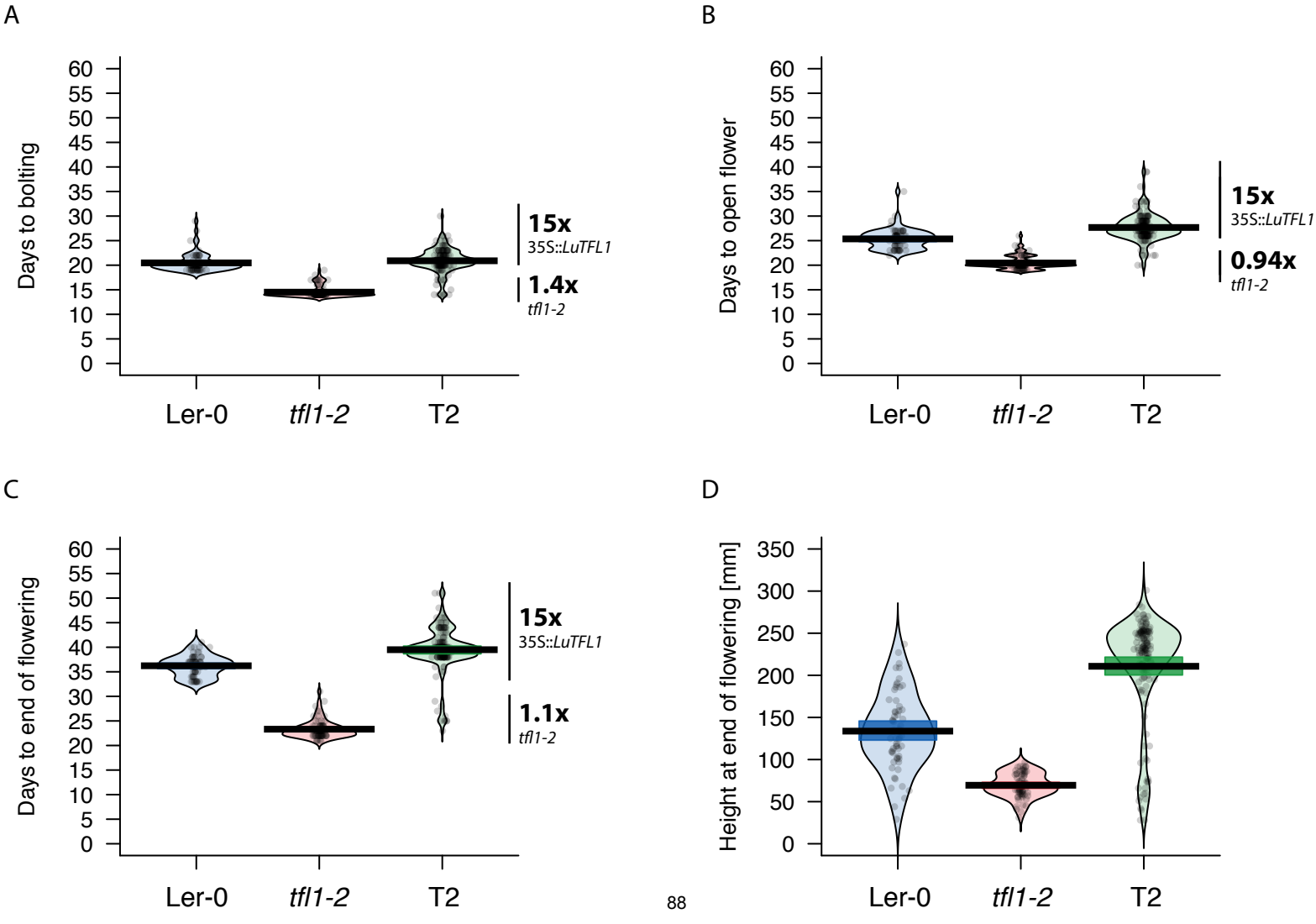


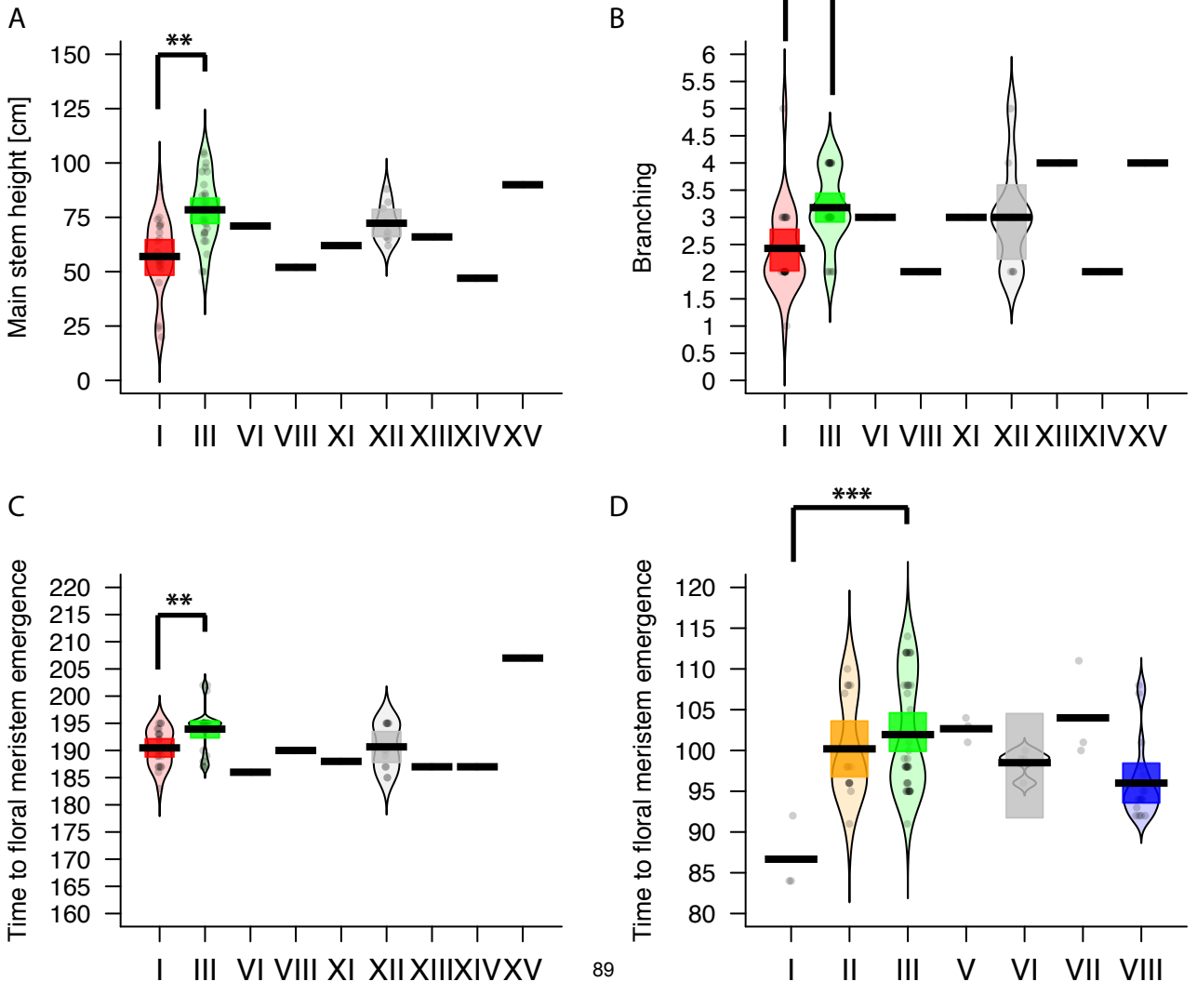
B.

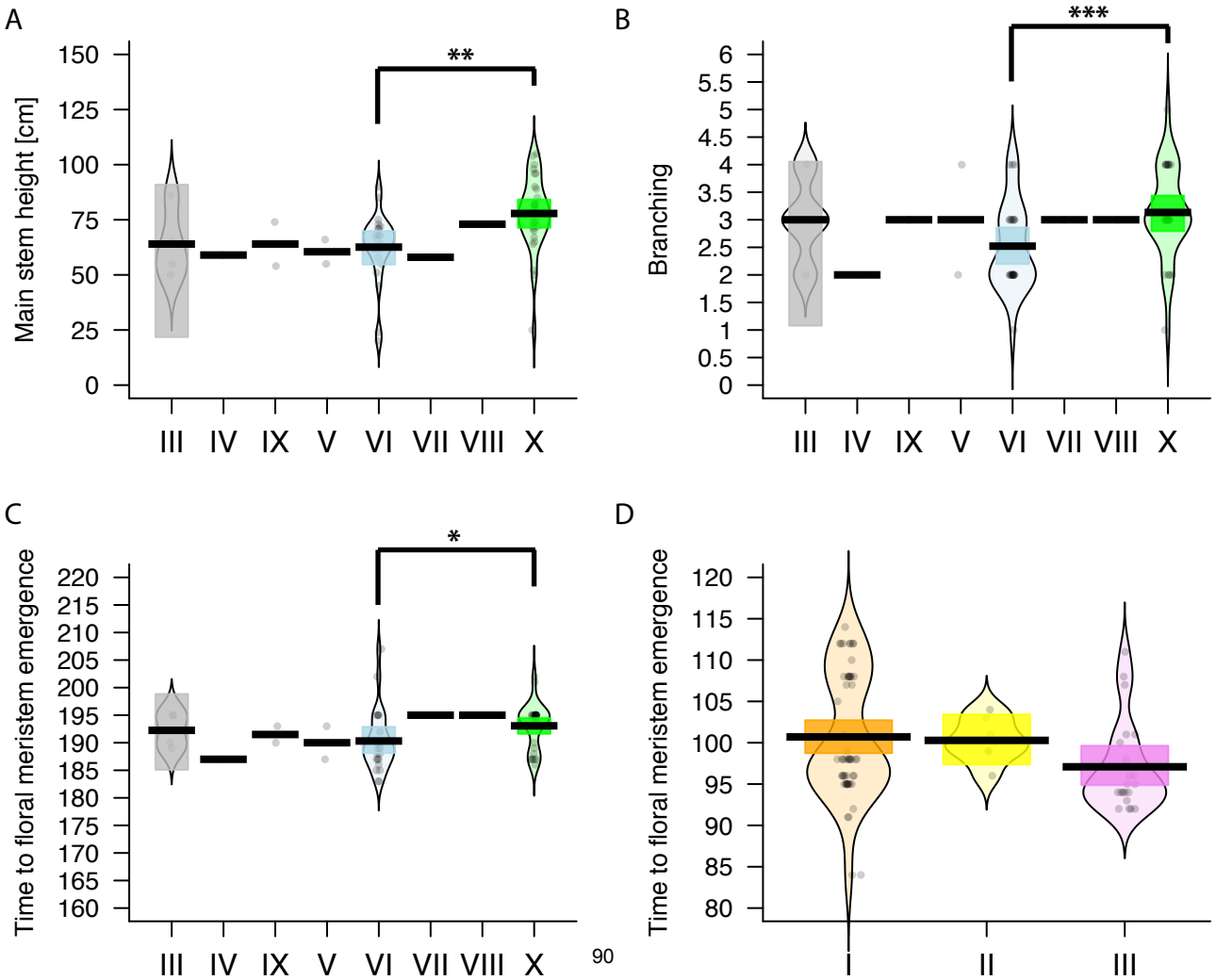


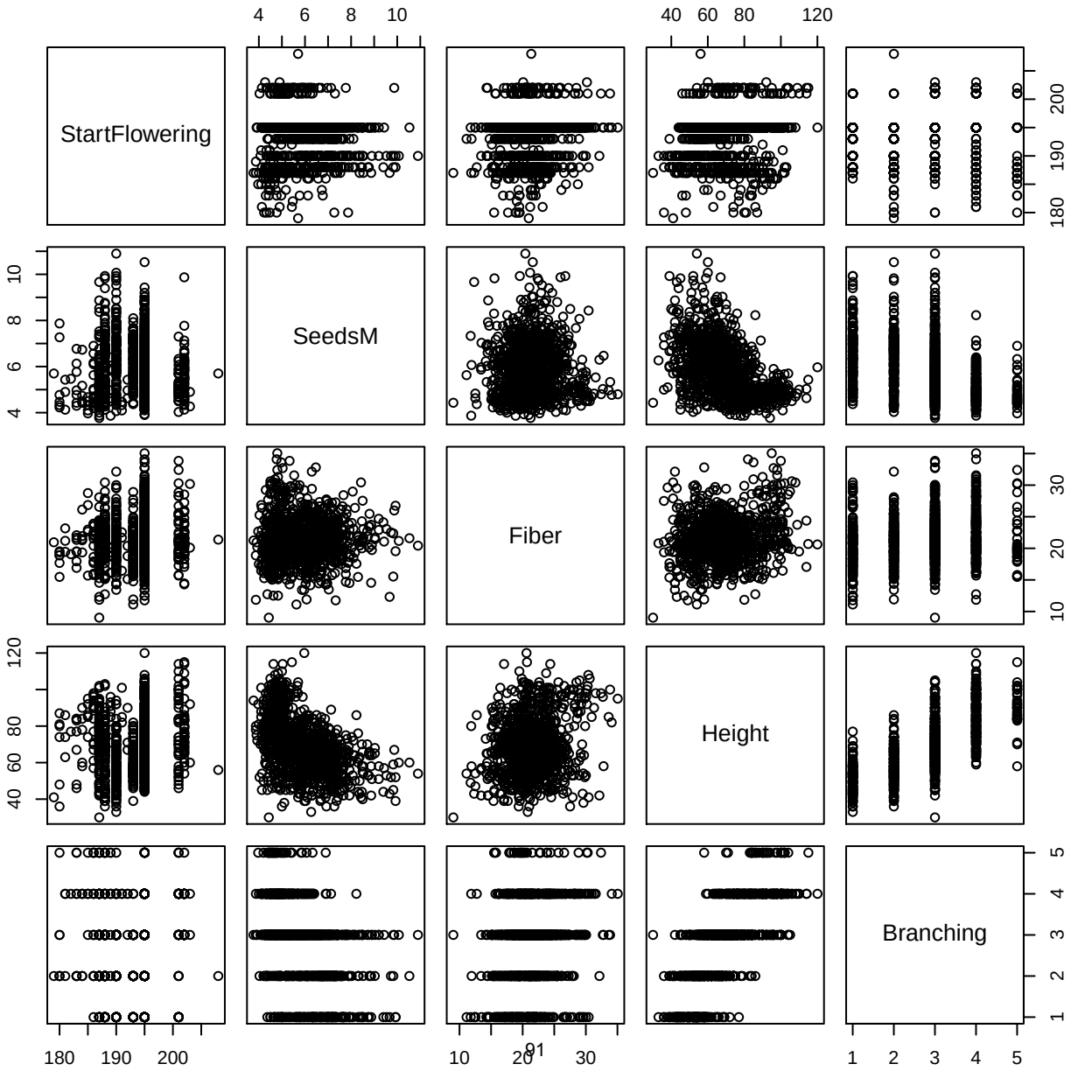

W042 (LuTFL1.I)
W043 (LuTFL1.III)
W067 (LuTFL1.III)
W069 (LuTFL1.III)
W077 (LuTFL1.II)
W094 (LuTFL1.VIII)
GAPDH
40d 55d 72d 91d 110d
LuTFL1
LuTFL2
LuFT


A
*
*
*
*
*
*
*
B
C
+
+
-
-

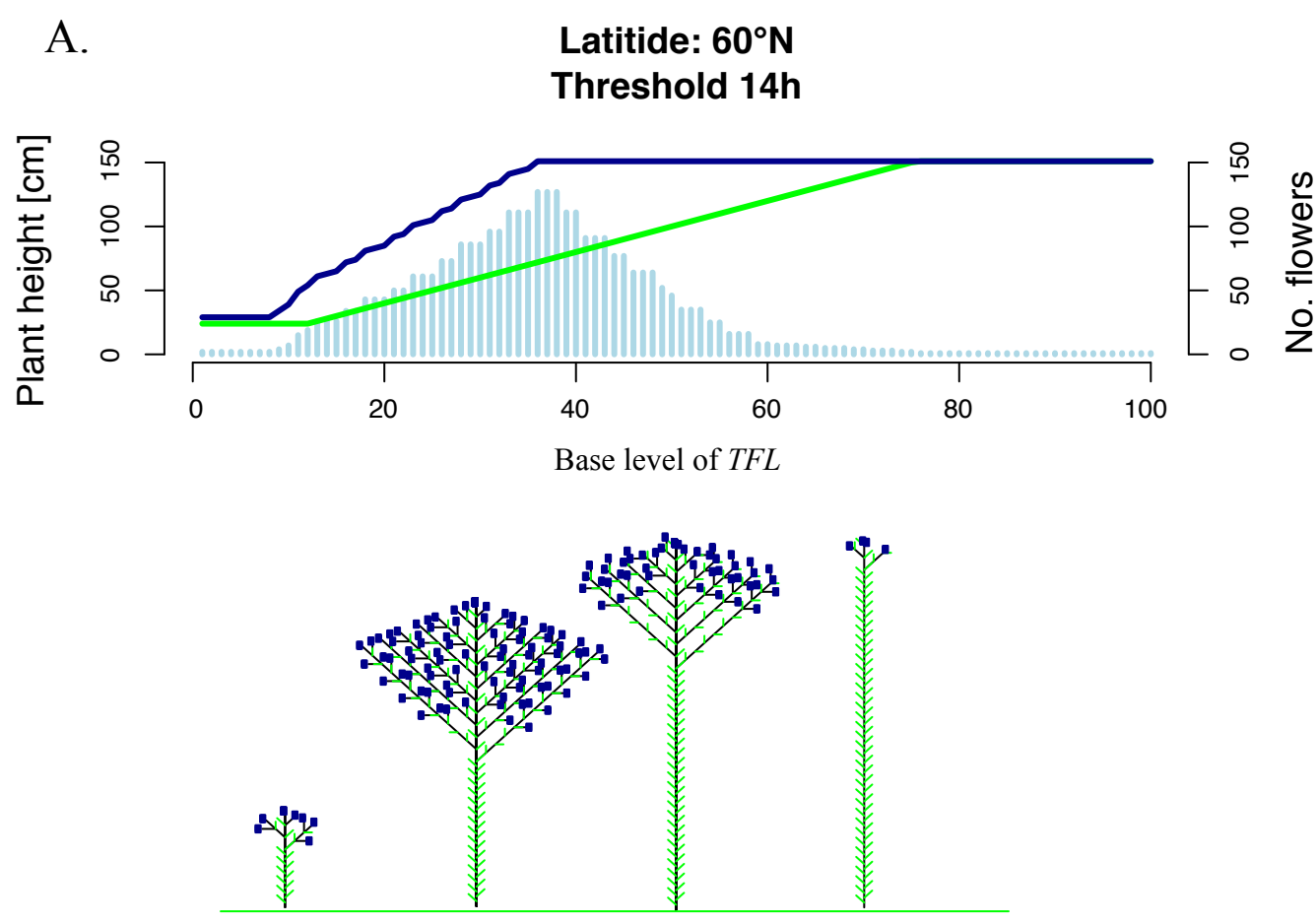

A
Days to bolting
60
55
50
45
40
35
30
25
20
15
10
5
0
15x
35S::LuTFL1
1.4x
tfl1-2
Ler-0
tfl1-2
T2
B
Days to open flower
60
55
50
45
40
35
30
25
20
15
10
5
0
15x
35S::LuTFL1
0.94x
tfl1-2
Ler-0
tfl1-2
T2
C
Days to end of flowering
60
55
50
45
40
35
30
25
20
15
10
5
0
15x
35S::LuTFL1
1.1x
tfl1-2
Ler-0
tfl1-2
T2
D
Height at end of flowering [mm]
350
300
250
200
150
100
50
0
Ler-0
tfl1-2
T2


A
Main stem height [cm]
150
125
100
75
50
25
0
**
I III VI VIII XI XII XIII XIV XV
B
Branching
6
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
**
I III VI VIII XI XII XIII XIV XV
C
Time to floral meristem emergence
220
215
210
205
200
195
190
185
180
175
170
165
160
**
I III VI VIII XI XII XIII XIV XV
D
Time to floral meristem emergence
120
115
110
105
100
95
90
85
80
***
I II III V VI VII VIII

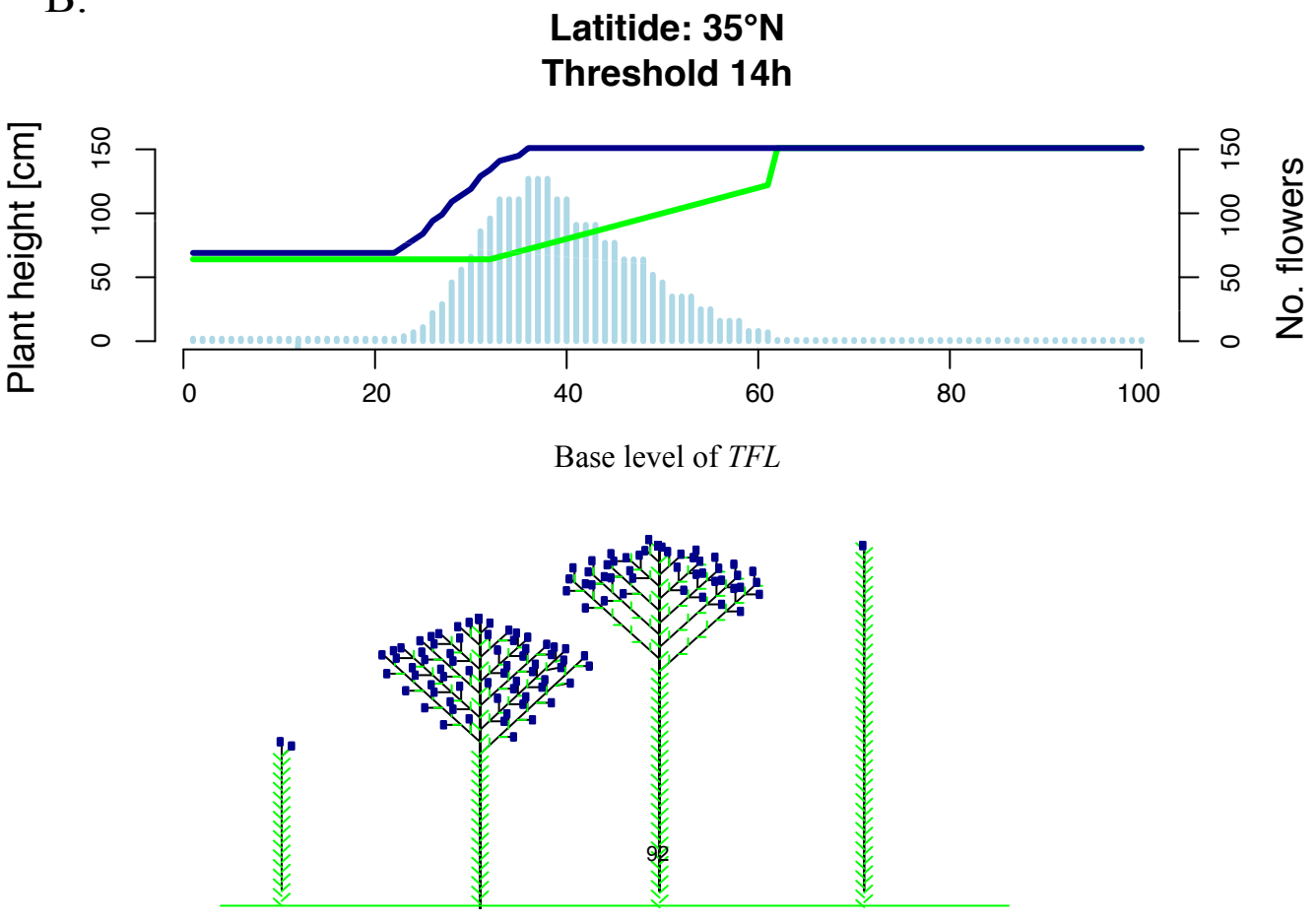

A
Main stem height [cm]
150
125
100
75
50
25
0
**
III IV IX V VI VII VIII X
B
Branching
6
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
***
III IV IX V VI VII VIII X
C
Time to floral meristem emergence
220
215
210
205
200
195
190
185
180
175
170
165
160
*
III IV IX V VI VII VIII X
D
Time to floral meristem emergence
120
115
110
105
100
95
90
85
80
I II III

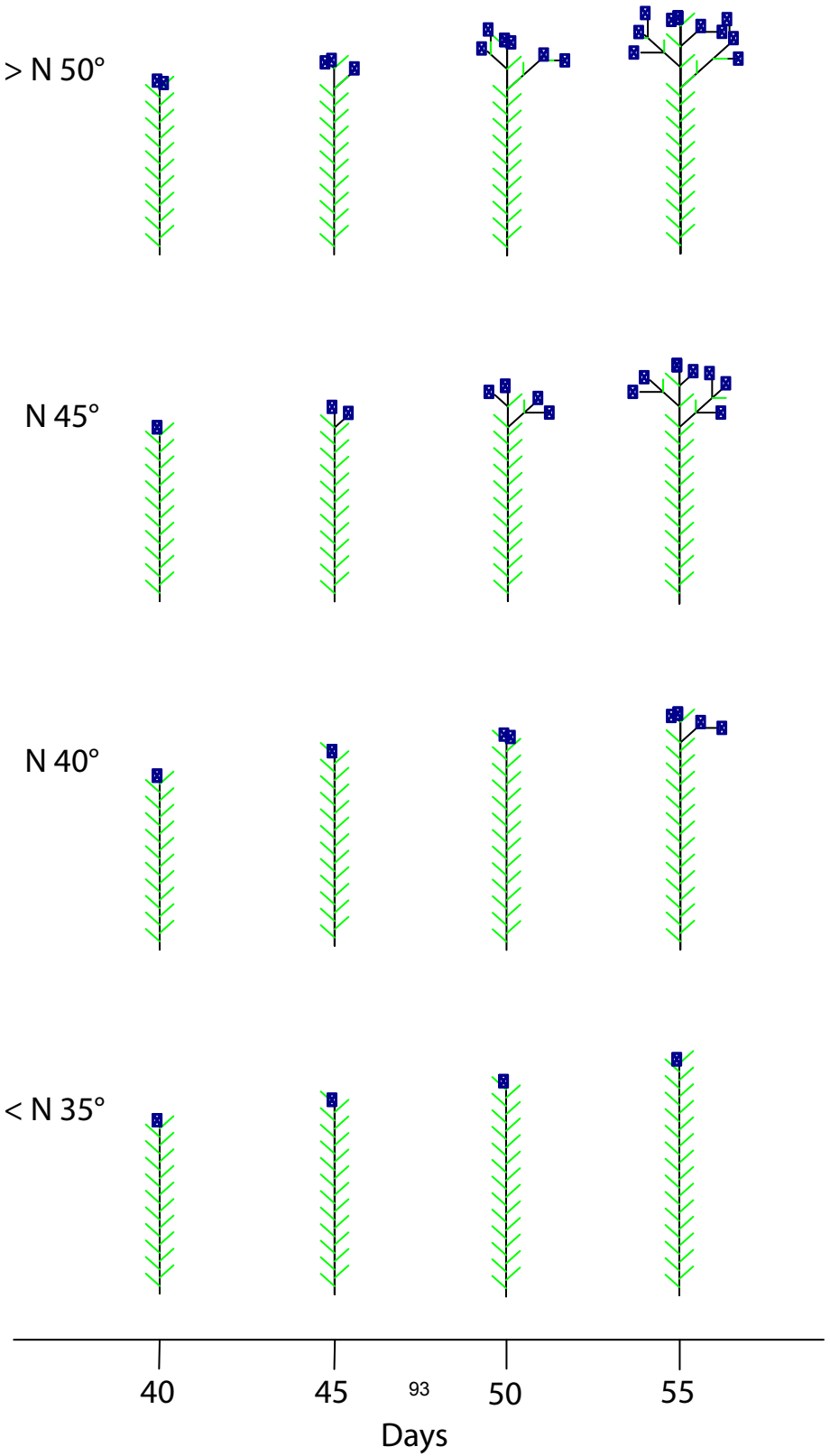

StartFlowering
SeedsM
Fiber
Height
Branching

A.


Latitide: 60°N
Threshold 14h
Plant height [cm]
0
50
100
150
No. flowers
0
20
40
60
80
100
Base level of *TFL*

B.


Latitide: 35°N
Threshold 14h
Plant height [cm]
0
50
100
150
No. flowers
0
20
40
60
80
100
Base level of *TFL*


> N 50°
N 45°
N 40°
< N 35°
40
45
93
50
55
Days