1 Flax latitudinal adaptation at *LuTFL1* altered architecture and promoted fiber production.

2 Rafal M Gutaker[1,2], Maricris Zaidem[2], Yong-Bi Fu[3], Axel Diederichsen[3], Oliver Smith[1,4] ,

3 Roselyn Ware[1], Robin G Allaby[1]*

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

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

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

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

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

Abstract

After domestication in the Near East around 10,000 years ago several founder crops, flax included, spread to European latitudes. On reaching northerly latitudes the architecture of domesticated flax became more suitable to fiber production over oil, with longer stems, smaller seeds and fewer axillary branches. Latitudinal adaptations in crops typically result in changes in flowering time, often involving the PEBP family of genes that also have the potential to influence plant architecture. Two PEBP family genes in the flax genome, *LuTFL1* and *LuTFL2*, vary in wild and cultivated flax over latitudinal range with cultivated flax receiving *LuTFL1* alleles from northerly wild flax populations. Compared to a background of population structure of flaxes over latitude, the *LuTFL1* alleles display a level of differentiation that is consistent with selection for an allele III in the north. We demonstrate through heterologous expression in *Arabidopsis thaliana* that *LuTFL1* is a functional homolog of *TFL1* in *A. thaliana* capable of changing both flowering time and plant architecture. We conclude that specialized fiber flax types could have formed as a consequence of a natural adaptation of cultivated flax to higher latitudes.

24 Crop plants were domesticated in the late Pleistocene to early Holocene at multiple centers
25 around the world and many of them subsequently expanded northwards after glacial retreat[1]. The
26 southwest Asian 'Neolithic Package'[2] including wheat, barley, lentils, chickpeas and flax was
27 domesticated in the Near East and spread out into Europe[3]. The environmental conditions
28 associated with the higher latitudes of Europe presented the plants of Near Eastern origin with
29 major challenges that have been associated with repeated agricultural collapse[4,5], possibly due to
30 the limited rates at which adaptation could take place due to the associated substitution load[6].
31 Plants of Neolithic Package, adapted to Near Eastern conditions, are often long-day and
32 vernalization dependent for flowering. Those crops would have struggled in short growing
33 season in Europe. Typically, wheat and barley adapted to these conditions through photoperiod
34 insensitivity which affects their flowering time[7–12]. Some northern flax types also show similar
35 loss of photoperiodical sensitivity phenotypes[13]. However, in contrast to the Neolithic cereals,
36 flax had a potential advantage in that it could have assimilated adaptations from its wild
37 progenitor species, which extends in range to northwestern Europe and shows patterns of
38 extensive local adaptation[14,15].
39 Cultivated flax (*Linum usitatissimum* L.) is a predominantly self-pollinating[16], diploid
40 crop plant that was domesticated from pale flax (*L. bienne* Mill.)[17] in the Near East[2]. Originally
41 domesticated for its oily seeds[18], flax is currently grown as either an oil or fiber crop. Specialized
42 fiber varieties likely emerged during the Neolithic 'Textile Revolution'[19–21], which started in
43 Central Europe about 6,000 years Before Present (BP). The two flax forms are different in their
44 seed size and distinct plant architecture[22]. In particular, specialized fiber varieties are
45 characterized with lower seed weight, fewer axillary branches and higher technical stem length,
46 relative to oil varieties[23,24]. It has been shown that those traits are under divergent selection in

47 both varieties[25].

48 In *Arabidopsis thaliana*, plant architecture can be modified simultaneously with both
49 flowering time by changes in regulation of the *FT* and *TFL1* genes[26], which belong to the
50 phosphatidyl ethanolamine-binding protein (PEBP) family[27] and control plant development
51 through regulation of floral meristem initiation. An allelic variant of a PEBP gene in barley,
52 *HvCET* has been involved in adapting this crop to the European climate through later flowering,
53 suggesting the potential importance of this family in latitudinal adaptations of crops during the
54 spread of agriculture[10]. Furthermore, *FT* and *TFL1* can control plant stature through promotion
55 or delaying floral initiation[26] and are highly conserved in eudicots[28]. The orthologs of *FT* and
56 *TFL1* in tomato have a strong effect on tomato fruit weight and number[29], plant architecture and
57 height, and flowering time[30,31]. Such differences in fruit weight and plant stature underlie the
58 distinction between fiber and oil specialized varieties in flax, raising the possibility that the
59 PEBP gene family in flax could be associated with both architectural changes and adaptation to
60 northerly latitudes through modifying flowering time.

61 In this study we screened PEBP gene orthologs in flax for signatures of selection along
62 increasing latitudes in European wild and cultivated flax. We additionally genotyped restriction
63 site-associated regions of flax genome to understand the underlying population structure in flax.
64 Finally, we validated the effect of a gene candidate for selection, *LuFTL1* on plant architecture
65 and flowering time through heterologous expression in *Arabidopsis thaliana*.

66 **Results**

67 **PEBP orthologs in flax and the diversity at *LuTFL1* locus.** In order to investigate the role of
68 PEBP genes in latitudinal adaptation and architecture control in cultivated flax, we surveyed

69 orthologs of this gene family. A scan for homologs of *A. thaliana* PEBP genes in the flax
70 assembly[32] revealed eight loci of interest with architectures similar to *TFL1* gene (Supplementary
71 Figure S1; Supplementary Table S1). Out of those, seven loci were amplified and sequenced
72 from flax genomic DNA (Supplementary Figure S2), four could be easily re-sequenced without
73 cloning and showed conserved four-exon structure (Supplementary Figure S1). Phylogenetic
74 analyses comparing the flax, *A. thaliana* and Lombardy poplar orthologs revealed that three
75 putative PEBP members in flax are closely related to floral suppressors TFL1 and ATC, while
76 one has similarity to their antagonist, FT (Supplementary Figure S3). Resequencing of those
77 genes in six accessions of flax indicated high levels of polymorphism in pale flax at the *LuTFL1*
78 locus, suggestive of broad standing variation.

79 We re-sequenced the *LuTFL1* locus in 32 wild and 113 cultivated flax accessions
80 (Supplementary Figure S4A, Supplementary Table S2). For comparison, we also re-sequenced
81 the *LuTFL2* locus in the same population of flaxes. We discovered wide molecular diversity at
82 both loci, with 16 and 11 alleles in *LuTFL1* and *LuTFL2* respectively (Supplementary Table S3).
83 A phylogenetic network of *LuTFL1* shows that allelic diversity is greater in pale flax (top plane,
84 Figure 1A) when compared to cultivated flax (bottom plane, Figure 1A). The majority (100 out
85 of 103) of cultivated flaxes were associated with just two of the groups (I and III) also found in
86 the wild. There were several groups apomorphic to groups I and III that appeared only in our
87 cultivated samples that likely represent post-domestication diversification (alleles XI, XII, XIII,
88 XIV, XV and XVI). Interestingly, the *LuTFL2* network shows greater diversity in cultivated flax
89 allele compared to pale flax (Supplementary Figure S5A, Supplementary Table S3).

90 There are multiple alleles in both *LuTFL1* and *LuTFL2* that are shared by pale and
91 cultivated flaxes. Such a pattern could be a result of multiple alleles being included during

92 domestication or through post-domestication gene flow from wild to domestic species. The
93 phylogeographic distribution of *LuTFL1* alleles in pale flax (Figure 1C), supports the notion that
94 allele I was passed from pale flax to cultivated flax during the initial domestication process in the
95 Near East. Conversely, in pale flax allele III is associated with the Bosphorus strait and the
96 Balkans beyond the domestication area but not in Eastern or Southern Turkey. We conclude that
97 *LuTFL1* III likely entered the cultivated flax gene pool through gene flow from pale flax during
98 the agricultural expansion into Europe, and reached higher frequencies in more northerly
99 latitudes (Figure 1B). The high diversity of *LuTFL2* alleles present in the cultivated gene pool
100 but absent from the wild gene pool suggests that they may have originated from wild populations
101 outside our sample set, or have become extinct in the wild. Despite not being closely located in
102 the genome, allele X of *LuTFL2* occurs in a strong disequilibrium with allele III of *LuTFL1*
103 (Supplementary Table S4), shows a similar latitudinal gradient (Supplementary Figure S5B), and
104 is absent from cultivated flaxes carrying the *LuTFL1* allele I.

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

114 We tested if *LuTFL1* and *LuTFL2* alleles are distributed in the frequencies that match the

115 expectation under neutral evolution using Tajima's D, R2, Fu and Li's D_2 and F methods
116 (Supplementary Table S5). Based on *LuTFL2* data we could not reject the neutrality hypothesis,
117 however, significantly negative scores in *LuTFL1* tests indicated an excess of rare alleles in
118 cultivated flax. Such a pattern of molecular diversity could be explained either by population
119 structure, demographic effects such as expansion, or selection. To distinguish between these
120 possibilities, we investigated the population structure and allele frequencies at neutral loci
121 through a restriction site-associated (RAD) sequencing approach[33,34].

122 **Population structure from genome-wide SNPs in wild and pale flax.** We surveyed the
123 genomes of 90 accessions of pale (28) and cultivated (62) flax (Supplementary Table S2) by
124 RAD sequencing[33]. We used *de novo* assembly approach on sequenced fragments[35], from which
125 we found 993 polymorphic RAD tags encompassing the total of 1686 SNPs present in at least
126 80% accessions. Based on those, we calculated genetic distances, which were used for
127 multidimensional scaling analysis. Three main clusters that correspond to three populations were
128 revealed: cultivated flax, dehiscent flax and pale flax (Figure 2A). Using an ADMIXTURE[36]
129 approach we identified four distinct ancestral components (lowest 5-fold cross validation error
130 for K=4). The pale flax of Eastern Anatolia represents all the admixture components that are
131 present in cultivated flaxes, including dehiscent varieties (Figure 2B), which supports an eastern
132 population involvement with the original domestication process. Two pale flaxes located in
133 eastern Greece outside the domestication center show sizeable admixture with dehiscent
134 varieties. This is congruent with previous reports that identify dehiscent varieties as genetically
135 distinct from other domesticated flax varieties, possibly reflecting an earlier independent
136 domestication[37,38]. A latitudinal gradient is apparent in the cultivated flax types, evident from a
137 tendency for southern populations to be affiliated to an orange grouping, and a cyan grouping in

138 northern populations in Figure 2B.

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

154 Given the proximity of the general differentiation of allele III between northern and
155 southern populations relative to the null distribution evident in the allele frequency spectrum, we
156 speculated that weak selection was probably involved. To investigate this further we applied a
157 method to determine the most likely selection coefficient associated with allele frequency
158 changes over time[39], using archaeological dates of the arrival of agriculture at different latitudes
159 to date the latitudinal allele frequencies, Figure 4. Our estimates confirmed weak selection with
160 values of s generally between 0.001 and 0.007 (mean 0.003), but interestingly we found an

161 increase in selection strength over time and latitude, indicating that allele III was progressively
162 more strongly selected for as flax moved to higher latitudes.

163 Finally, we investigated whether the RAD loci data harboured evidence of movement of
164 alleles between wild and cultivated flax populations beyond the area of domestication, which we
165 implied from geographic distribution of *LuTFL1* alleles in wild and cultivated flaxes (Figure 1B,
166 1C). We observed no clear signal using f_3 statistics (cultivatedN; cultivatedS, wildN), however,
167 varying rates of population differentiation through founder effects and multiple complex
168 processes of gene flow can mask signals of gene movement between populations[40]. We therefore
169 considered portions of the data where evidence of gene movement may be clearer. We identified
170 subsets of alleles that were discriminatingly higher at a range of frequencies in either northern or
171 southern populations of wild flax. We reasoned that if cultivated flax had spread northwards
172 without contact with wild populations, then the differentiation between northern and southern
173 populations at these subsets of loci should not be perturbed relative to the background
174 distribution of loci differentiation. We found that there was no significant difference between the
175 cultivated populations at loci that were higher in southern wild populations, but that there was an
176 elevated level of differentiation between cultivated populations at loci that were differentially
177 high in northern wild populations (Supplementary Figure S7). These findings are verified by an
178 increased ancestral information content (I_a)[41] at loci that were differentially high in northern but
179 not southern wild populations (Supplementary Table S7). This indicates a bias against the neutral
180 expectations of uncorrelated drift between the cultivated and wild population pairs that could
181 represent gene flow from northern wild populations into northern cultivated populations.
182 Alternatively, this signature could be an effect of a parallel selection for traits associated with
183 high frequency RAD loci alleles in the wild population. We deemed the latter unlikely since 78%

184 of loci that had elevated frequencies in northern wild populations also had elevated frequencies
185 in northern cultivated populations, which would require an explanation of selection being
186 responsible more frequently than drift for high frequency alleles in the north.
187

188 **Functional homology of *LuTFL1*.** We explored the function of *LuTFL1* and *LuTFL2*, first by
189 investigating expression of these genes in the leaf and shoot apex tissue of pale flax plants from
190 six populations over the course of the plant development to flowering using a semi-quantitative
191 PCR approach. *LuTFL1* was expressed continuously from the first point of measurement (40
192 days) to flowering (Supplementary Figure S8) consistent with *TFL1* expression in *A. thaliana*[42]
193 as opposed to *ATC*, which is expressed in the hypocotyl in *A. thaliana*[43]. We detected no
194 expression of *LuTFL2* in shoot tissue. Based on phylogenetic evidence, *LuTFL1* is more similar
195 to *ATC*, however characterization of its expression pattern led us to conclude that *LuTFL1* is
196 most likely functionally orthologous to *TFL1*.
197 We tested the hypothesis that *LuTFL1* functions as *TFL1* by complementing *TFL1* knock-
198 out mutants of *A. thaliana* with a *LuTFL1* construct under a 35S promoter. We obtained a stable
199 line after transforming a *tfl1-2* mutant in Landsberg *erecta* (Ler) genomic background[44] using an
200 *Agrobacterium* system. Over 160 plants of 35S::*LuTFL1* inbred progeny (T2) were grown in
201 long day conditions. For all the plants we measured days to bolting, days to open flowers, days to
202 end of flowering and stem height at the end of flowering. Flowering time phenotypes segregated
203 in T2 with a Mendelian 15 to 1 ratio (Supplementary Figure S9), which suggests that two copies
204 of transgene were integrated in *tfl1-2* mutant. Transgene presence/absence was validated using
205 primers specific to the transgene in 21 plants with extreme phenotypes, Supplementary Figure
206 S9. Plants that contained the 35S::*LuTFL1* transgene were compared with *tfl1-2* mutants and Ler-

207 0 wild type. The start and end of flowering was significantly lower in *tfl1-2* mutants than in Ler-
208 0 and 35S::*LuTFL1* plants (Figure 5; Supplementary Figure S9), suggesting that the flax *LuTFL1*
209 I allele can rescue non-functional *TFL1* in *A. thaliana*. Constitutive expression of this allele
210 under 35S promoter further delays flowering, when compared to Ler-0. Subsequently, the
211 amount of expressed *TFL1*/*LuTFL1* has a positive impact on plant height at the end of flowering.
212 We conclude that, similar to *TFL1* in *A. thaliana*, *LuTFL1* functions in delaying flowering,
213 promoting indeterminate growth and finally resulting with increased plant height at senescence.
214 The experimental data from transgenic 35S::*LuTFL1 A. thaliana* plants support the notion that
215 the correlation between allele type and phenotype could have a causative relationship.
216 Although *LuTFL1* is a functional homolog of *TFL1*, we still do not fully understand the
217 effect of allelic diversity in flax at this locus. We detected no non-synonymous substitutions
218 between groups I and III that could account for a potential functional basis for the difference in
219 selection histories between the alleles. However, a transcription factor-binding site was identified
220 to be present in the promoter regions of *LuTFL1* III and VIII, but absent from *LuTFL1* I
221 (Supplementary Figure S4B). We hypothesized that such a difference between the allele groups
222 could result in a change of expression pattern of *LuTFL1* and in consequence, altered flowering
223 time, which in turn could be under selection in the northerly latitudes. To help understand how
224 an adaptation to latitude through the *TFL/FT* system could produce the observed flowering and
225 architectural changes in flax we applied a plant development model. The model was used to
226 explore how selection at the *TFL1* locus and latitudinal movement could affect stem height,
227 which correlates with increased fiber content[45]. The model predicts that increased *TFL1*
228 expression is expected to produce longer stemmed plants, but also that at higher latitudes such
229 plants would be more fit through a greater rate of flower production relative to long stemmed

230 forms in the south (Figure 6). This model outcome predicts that adaptation to variable climate in
231 northerly latitudes by indeterminate flowering through selection at a *TFL1* locus allows increase
232 in stem height, which is associated with an improvement in fiber content.

233 **Discussion**

234 Crop plants were domesticated independently around the world at the 'Domestication Centers' in
235 prehistoric times[1]. Early farmers spread out from those centers and brought with them major
236 crops as they expanded into new geographic zones. There are numerous examples of crop plants
237 that adapted to new environments through changes in flowering time, particularly during
238 northwards expansions. Maize, domesticated in Mesoamerica, would only be successfully
239 cultivated in the temperate North America after adapting through reduction of flowering time[46].
240 Similarly, after domestication of rice in Eastern China, its successful cultivation in Korea and
241 Japan required changes in flowering behaviour[47]. In many plants, *TFL1* family genes that
242 influence flowering time by acting antagonistically with *FT* to delay flowering[48–50] have been
243 identified as targets for selection in adaptation to northern latitudes and have also been shown to
244 influence plant architecture[26,30,31] and fruit yield[29]. *TFL1* homologs have played an important role
245 in adaptation to temperate climate in barley[10] and soybean[51], where it has been noted that
246 adaptive variant at this locus is strongly enriched in the northern end of crop distribution.

247 Cultivated flax was domesticated in the Near East alongside wheat and barley[2] and
248 likewise adapted to temperate climate of prehistoric Europe. In this paper, we show that flax
249 genome encodes multiple PEBP orthologs. Of those, we show that *LuTFL1* has latitudinally
250 structured allelic diversity and molecular signatures consistent with selection. This has been
251 validated against null distribution of neutrally evolving loci, which we generated using RADseq
252 approach. To further reinforce evidence of selection on *LuTFL1*.III in the northern Europe, we

253 anticipate the availability of full genome information for flax populations. Similarly to other crop
254 plants, adaptation to Northerly latitudes in flax was achieved through modifications in flowering
255 behaviour. Through heterologous overexpression of *LuTFL1*, we show that this gene is in fact a
256 functional homolog of *TFL1* and is capable of changing flowering time[26,30,52]. Similarly, *LuTFL1*
257 is shown to be a decisive factor in establishing plant architecture and height. Based on that we
258 conclude that the increase in frequency of *LuTFL1*.III in the north and signatures consistent with
259 natural selection are results of flax adaptation in Europe through changes in flowering time.
260 Interestingly, this change had an impact on flax architecture and could have predisposed flax to
261 fibre production. This evolutionary transition may have been key to the rise of the textile
262 revolution in central Europe[20], which is evidenced in the reduction of flax seed size[19] and
263 improved tools for fiber production[21] in Neolithic contexts. Our conclusions are consistent with
264 phenotypic associations in cultivated flax, in which fibre content is correlated with plant height
265 and flowering time at the same time[45].

266 Flax is shown to have adapted to Northerly latitudes in a similar fashion to other crop
267 plants. However, this study additionally provides an unusual example on how the influence of
268 natural adaptations within a crop may ultimately influence the use of that crop. Normally, the
269 evolution of domesticated plants is associated with an initial rise of a suite of traits collectively
270 termed the domestication syndrome[53], later followed by diversification or improvement traits[54].
271 Such traits are considered for the most part to be agronomically advantageous. Mounting
272 evidence shows that plant adaptations within the human environment are not restricted to
273 crops[55–57] and within crop plants not all adaptations are of obvious benefit to agronomic
274 productivity and yield, but related to the wider ecology[58]. Examples include a tendency to
275 increase rates of water loss in domesticate forms[59]. Such adaptations highlight a tension between

276 the effective exploitation of plants as a resource by endowment of useful traits and ongoing
277 natural adaptations that occur in plants that may compromise productivity. This study shows that
278 the adaptation of flax to the European variable climate during Neolithic likely involved the
279 *LuFTL1* locus and resulted in a change of cultivation purpose through substantial modification of
280 the stem height, which is correlated with the technical fiber quality.

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

290 **Materials and Methods**

291 **Plant Materials**

292 Samples from 16 populations were collected from natural habitats in the summer of 2011 from
293 Croatia, Montenegro, Albania, Greece and Bulgaria. These were combined with a further 16
294 populations of pale flax supplied by the Plant Genetic Resources of Canada, Agriculture and
295 Agri-Food Canada, Saskatoon Research Centre. Seed material for 58 modern cultivars of flax
296 and 18 landraces were obtained from the Plant Genetic Resources of Canada, Agriculture and
297 Agri-Food Canada, Saskatoon Research Centre. A further 28 historic landrace accessions were

298 obtained from Plant Genetic Resources of the N. I. Vavilov Research Institute of Plant Industry,
299 St. Petersburg, Russia. Samples ranging in age up to 194 years were sampled from Herbaria
300 variously from the Natural History Museum, University of Oxford and the University of
301 Warsaw.

302 **Identification and resequencing of *LuTFL1* locus**

303 Based on multiple alignments that contain sequences from different eudicots, degenerate primers
304 were designed in exonic regions to cover between 400 bp of *TFL1* region. The putative
305 sequences of *TFL1* in flax were subject to BLAST searches against the cultivated flax genome
306 scaffolds v1.0 and CDs v1.0 databases[62] in order to identify full sequences of genes of interest
307 and their flanking regions. Phylogenetic trees with *TFL1* homologs from *A. thaliana* and *P. nigra*
308 were estimated using Bayesian Inference in MRBAYES v3.2[63].

309 A total of 148 samples were used in this study, including 58 cultivars, 18 landraces, 38
310 historic landraces of cultivated flax, 32 pale flax samples and two *L. decumbens* accessions. The
311 total DNA from 20 mg of seeds was isolated using modified DNEasy® Plant Mini Kit (Qiagen)
312 protocol. In case of herbarium specimens – mixed plant tissue was used for extraction using
313 CTAB protocol[64]. A 1367 bp target region was amplified from *LuTFL1* spanning exons 1-4
314 using primers (5-TTACAACTCCACCAAGCAAGTC, and 5- TGTCTCGCGCTGAGCATT).
315 Haplotype networks was made using uncorrected_P character transformation and Rooted Equal
316 Angle splits transformation in *SplitsTree4*[65]. The size of the network nodes was made
317 proportional to the number of samples that share a particular haplotype. The relationship between
318 latitude and *LuTFL1* and *LuTFL2* haplotypes separately was modelled using logistic regression
319 in *R* programme and plotted using *popbio* package.

320 **Selection signatures at *LuTFL1* locus**

321 In order to test if *LuTFL1* and *LuTFL2* were inherited neutrally or were under selection,
322 neutrality tests with variety of estimators were carried out. Tajima's D[66], Fu and Li's D, Fu and
323 Li's F[67] statistics were calculated in *INTRAPOP* and the R2 statistic[68] was calculated in *R* using the
324 *PEGAS* package. The extent of linkage disequilibrium was examined between the alleles of the
325 *LuTFL1* and *LuTFL2* loci. Expected frequencies of genotypes were calculated from the observed
326 individual allele frequencies. The correlation coefficient of *D*, *r*, was calculated from observed
327 genotyping data where:

328 $$r = \frac{[p1p2]\times[q1q2]-[p1q2]\times[q1p2]}{\sqrt{p1p2q1q2}}$$

329 Square brackets indicate the observed proportion of the sample of genotype combinations, *p1,*
330 *p2, q1 and q2* refer to the allele frequencies of *p* and *q* type alleles at the first and second locus
331 respectively, and the numerator equates to the statistic *D*.

332 Each latitude in Europe is associated with a different time of arrival of agriculture based
333 on archaeological data. The program SELECTION_TIME.pl[39] estimates selection coefficients
334 from dated frequencies. The model takes as input a series of dated frequencies, mating strategy
335 of the organism, whether the data is phenotype or allele frequency and whether the trait under
336 selection is dominant or recessive. The program takes a pair of input observed dated allele
337 frequencies as start and stop frequencies. The inbreeding coefficient *F* is used to determine the
338 initial genotype frequencies. The selection coefficient *s* is then described by equation:

339 $$s = 1 - \frac{1}{(1-k)}$$

340

341 Where *k* is the selection differential. To account for sample error, we generated upper and lower
342 bounds of *s* with Beta distributions assuming a binomial process of *x* observations in *n* samples,
343 where x equates to the number of samples carrying the *LuTFL1* III allele, and *n* equates to the

344 total number of samples at a specific latitude. For details see Supplementary Methods.

345 **RAD sequencing and processing**

346 From the total collection, 90 plants were chosen for the RADseq experiment (Supplementary
347 Table S1). They represent 28 pale flax and 62 cultivated flax accessions. Plants were grown in
348 glasshouse conditions and harvested after seedlings reached 10 cm of height. Snap-freezed
349 material was ground with 3 glass beads (3 mm diameter) in TissueLyser machine (Qiagen) and
350 then used for DNA isolation with DNeasy® Plant Mini Kit (Qiagen) following the
351 manufacturer's manual. DNA was digested with 2U of SbfI HF restriction enzyme (New
352 England Biolabs) for one hour and then the reaction was heat inactivated. For RADseq library
353 preparation, 40 µl of genomic DNA at concentration of 25 ng/µl was prepared following
354 previous methods[34]. Barcoded libraries of 10 different DNA isolates were pooled together into
355 the total of 2 µg of DNA in 300 µl of solution. DNA was sheared ten times on ice in Bandelin
356 Sonoplus HD 2070 sonicator. Sheared DNA was purified using AMPure XP beads (Agencourt)
357 in proportion 1 to 1 with buffer. Approximately 20 ng of library was amplified with Phusion HF
358 Master Mix (NEB) in total of 14 cycles. The first 50 libraries were sequenced on HiSeq 2000
359 Illumina platform at Oxford Genomics Centre using the TruSeq reagents. Another 40 samples
360 were submitted for sequencing on Genome Analyzer II Illumina platform in Genomic Centre at
361 University of Warwick.

362 Raw reads from Illumina sequence runs were assembled into FastQ format and assessed
363 in FAST QC software v0.10.1 to check for standard quality measures: quality scores,
364 overrepresented sequences, GC content and N content[69]. We generated SNPs from *de novo*
365 assembled RADtags. Scripts from the STACKS pipeline v1.05 was employed to de- multiplex
366 and discover RAD sequence markers[35]. Low quality sequences and those characterized by

367 erroneous barcodes were discarded. Remaining sequences were sorted according to their
368 barcodes. SNPs were called with the following settings: 5 identical, raw reads required to create
369 marker stack, 5 mismatches were allowed between alleles in single individual and calling
370 haplotypes from secondary reads was disabled.

371 **Population structure and differentiation**

372 Distances calculated from identity by sequence between all individuals were measured using
373 plink (v1.07)[70]. This distance matrix was used for multi-dimensional scaling (MDS) approach in
374 R with cmdscale function. Population ancestry of each individual was modelled using
375 ADMIXTURE (v1.23)[36]. We assumed existence of two to seven ancestral populations and chose
376 the model, which was characterised with the lowest five-fold cross-validation error.

377 We utilized the ancestral information content concept[41] to employ a high variant test to
378 explore possible gene movement between populations when there may be complex confounding
379 movements masking a general signal. We identified three subsets of loci in which the difference
380 in frequency (∂f) between northern and southern wild populations was more than 0.5, 0.4 or 0.3,
381 with the higher allele in the northern population. We examined the ancestry information content
382 of the ∂f loci subsets using the I_a ($I_{a\partial f}$) statistic relative to the information content of all loci (I_{at}).
383 We then compared the ∂f subset mean F_{ST} with the total null background of markers 100000
384 randomly sampled subsets, for details see Supplementary Methods.

385 Cultivated flax individuals were segregated into two groups based on their latitude;
386 individuals from above N40° were included in the northern subpopulation while individuals from
387 below N40° in southern. We calculated size-corrected and uncorrected F_{ST} values between
388 northern and southern cultivated flax populations for each RADtag using functions in
389 OutFLANK R package. We applied trimming approach to infer the distribution of F_{ST} for neutral

390 markers and used likelihood estimation for outliers[71]. Allele frequency spectrums were generated
391 from in house scripts from STRUCTURE format files acquired from *de novo* assembly RAD
392 tags. P value heat maps were generated from the allele frequency spectrums using in house
393 scripts that calculated probability density functions of Gaussian distributions from mean and
394 standard deviations of frequency distributions perpendicular to the diagonal.

395 **Expression and functional homology of *LuTFL1***

396 Plants of six populations (W042, W043, W067, W069, W077 and W094) were sown in 100
397 replicates each. Seeds were stratified for 3 days in the dark at 4°C and grown at room
398 temperature with a 16-hour photoperiod for 10 days followed by vernalization in cold room at
399 4°C for a period of 40 days. Subsequently plant were moved to growth chambers with 16h
400 daylight at 24°C until they flowered. Samples of three plants were taken at 0, 15, 17, 19 and 21
401 days. Samples were snap-frozen and the total RNA extraction was carried out from 20 mg of
402 ground tissue using mirVana™ miRNA isolation kit (Invitrogen) following the standard
403 protocol. DNA contaminants were digested in reaction with 2 units of DNase I (Invitrogen) and 2
404 µg of total RNA in DEPC-treated water. cDNA was synthesized from 1µg of DNase-treated
405 RNA with use of SuperScript® II reverse transcriptase and unspecific Oligo dT primer following
406 the manufacturer instructions. Specific primers were designed in exonic regions of *LuTFL1*,
407 *LuTFL2* and *LuFT* to cover approximately 150 bp long fragment. The *GAPDH* housekeeping
408 gene was chosen as an expression control. Semi-quantitative PCR was carried out for each
409 sample with 30 cycles.

410 We have used Landsberg erecta (Ler-0) ecotype with functional *TFL1* gene as a reference
411 and its tfl mutant, *tfl1-2*[44] for complementation experiment with *LuTFL1*.III coding sequence. In
412 these experiments plants of *Arabidopsis thaliana* were grown on soil in long days (16 h light/8

413 hours dark) under a mixture of cool and warm white fluorescent light at 23°C and 65% humidity.

414 Coding sequence of *LuTFL1*.III allele was amplified and introduced into a Sma1-digested
415 Gateway entry vector, pJLBlue[rev], as a blunt-end fragment. The resulting entry vector,
416 propagated in *Escherichia coli* DH5α strain, was then used in a recombination reaction with a
417 modified Gateway recombination cassette pGREEN-IIS[72]. This recombination reaction
418 effectively placed the *LuTFL1*.III allele behind the constitutive CaMV 35S promoter in the
419 pGREEN vector conferring resistance to BASTA. Sequence of PCR products and subsequent
420 plasmid vectors were checked by Sanger sequencing and compared to the expected input
421 sequence from the cDNA and vector backbone. The expression construct was introduced into *A.*
422 *thaliana tfl1-2* mutants by *Agrobacterium tumefaciens*-mediated transformation[73]. Transformed
423 seeds were stratified and selected in soil with BASTA herbicide. Seeds from a single line with
424 resistance to BASTA were collected.

425 Second generation of 35S::*LuTFL1*.III tfl1-2 transformants was sown together with wild
426 type Ler-0 and *tfl1-2* mutant plants for phenotyping. We have measured flowering time as days
427 to emergence of floral meristem and days to opening of first flower. Additionally, we measured
428 days to end of flowering and plant height at the end of flowering. We have selected random 32
429 individuals for transgene genotyping. The difference in mean phenotypes for 35S::*LuTFL1*.III,
430 Ler-0 and tfl1-2 was tested using t-test in R.

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

433 **Acknowledgments**

434 We would like to acknowledge Sabine Karg, for her help in clearing the archaeological
435 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).

Author contributions.

RMG, MZ, OS and RW carried out the research. RMG, YBF and RGA designed the project. RMG and RGA analyzed the data. RMG, YBF, AD and RGA wrote the manuscript.

Additional Information

Competing interests: The authors declare no competing financial or non-financial interests.

454 **References**

455 1. Purugganan, M. D. & Fuller, D. Q. The nature of selection during plant domestication. *Nature* **457,**
456 843–848 (2009).
457 2. Zohary, D., Hopf, M. & Weiss, E. *Domestication of Plants in the Old World: The Origin and Spread*
458 *of Domesticated Plants in South-West Asia, Europe, and the Mediterranean Basin, 4th Edition*.
459 (2012).
460 3. Colledge, S., Conolly, J. & Shennan, S. The Evolution of Neolithic Farming from SW Asian Origins
461 to NW European Limits. *European Journal of Archaeology* (2016). doi:10.1177/1461957105066937
462 4. Shennan, S. *et al.* Regional population collapse followed initial agriculture booms in mid-Holocene
463 Europe. *Nat. Commun.* **4,** 2486 (2013).
464 5. Timpson, A. *et al.* Reconstructing regional population fluctuations in the European Neolithic using
465 radiocarbon dates: a new case-study using an improved method. *J. Archaeol. Sci.* **52,** 549–557
466 (2014).
467 6. Allaby, R. G., Kitchen, J. L. & Fuller, D. Q. Surprisingly Low Limits of Selection in Plant
468 Domestication. *Evol. Bioinform. Online* **11,** 41–51 (2015).
469 7. Worland, A. J. *et al.* The influence of photoperiod genes on the adaptability of European winter
470 wheats. *Euphytica* **100,** 385–394 (1998).
471 8. Turner, A., Beales, J., Faure, S., Dunford, R. P. & Laurie, D. A. The pseudo-response regulator Ppd-
472 H1 provides adaptation to photoperiod in barley. *Science* **310,** 1031–1034 (2005).
473 9. Jones, H. *et al.* Population-based resequencing reveals that the flowering time adaptation of
474 cultivated barley originated east of the fertile crescent. *Mol. Biol. Evol.* **25,** 2211–2219 (2008).
475 10. Comadran, J. *et al.* Natural variation in a homolog of Antirrhinum CENTRORADIALIS contributed
476 to spring growth habit and environmental adaptation in cultivated barley. *Nat. Genet.* **44,** 1388–1392
477 (2012).
478 11. Diaz, A., Zikhali, M., Turner, A. S., Isaac, P. & Laurie, D. A. Copy number variation affecting the

479 photoperiod-B1 and vernalization-A1 genes is associated with altered flowering time in wheat
480 (Triticum aestivum). *PLoS One* **7,** (2012).
481 12. Takenaka, S. & Kawahara, T. Evolution and dispersal of emmer wheat (Triticum sp.) from novel
482 haplotypes of Ppd-1 (photoperiod response) genes and their surrounding DNA sequences. *Theor.*
483 *Appl. Genet.* **125,** 999–1014 (2012).
484 13. Darapuneni, M. K., Morgan, G. D., Ibrahim, A. M. H. & Duncan, R. W. Effect of vernalization and
485 photoperiod on flax flowering time. *Euphytica* **195,** 279–285 (2014).
486 14. Uysal, H., Kurt, O., Fu, Y.-B., Diederichsen, A. & Kusters, P. Variation in phenotypic characters of
487 pale flax (Linum bienne Mill.) from Turkey. *Genet. Resour. Crop Evol.* **59,** 19–30 (2012).
488 15. Uysal, H. *et al.* Genetic diversity of cultivated flax (Linum usitatissimum L.) and its wild progenitor
489 pale flax (Linum bienne Mill.) as revealed by ISSR markers. *Genet. Resour. Crop Evol.* **57,** 1109–
490 1119 (2010).
491 16. Dillman, A. C. Natural crossing in flax. *Journal of the American Society of Agronomy* **30,** (1938).
492 17. Diederichsen, A. & Hammer, K. Variation of cultivated flax (Linum usitatissimum L. subsp.
493 usitatissimum) and its wild progenitor pale flax (subsp. angustifolium (Huds) Thell). *Genet. Resour.*
494 *Crop Evol.* **42,** 263–272 (1995).
495 18. Fu, Y.-B. & Allaby, R. G. Phylogenetic network of Linum species as revealed by non-coding
496 chloroplast DNA sequences. *Genet. Resour. Crop Evol.* **57,** 667–677 (2010).
497 19. Herbig, C. & Maier, U. Flax for oil or fibre? Morphometric analysis of flax seeds and new aspects of
498 flax cultivation in Late Neolithic wetland settlements in southwest Germany. *Veg. Hist. Archaeobot.*
499 **20,** 527–533 (2011).
500 20. Leuzinger, U. & Rast-Eicher, A. Flax processing in the Neolithic and Bronze Age pile-dwelling
501 settlements of eastern Switzerland. *Veg. Hist. Archaeobot.* **20,** 535–542 (2011).
502 21. Maier, U. & Schlichtherle, H. Flax cultivation and textile production in Neolithic wetland
503 settlements on Lake Constance and in Upper Swabia (south-west Germany). *Veg. Hist. Archaeobot.*
504 **20,** 567–578 (2011).

505 22. Kulpa, W. & Danert, S. Zur systematik von Linum usitatissimum L. *Kulturpflanze* **3,** 341–388
506 (1962).
507 23. Diederichsen, A. & Raney, J. P. Seed colour, seed weight and seed oil content in Linum
508 usitatissimum accessions held by Plant Gene Resources of Canada. *Plant Breed.* **125,** 372–377
509 (2006).
510 24. Diederichsen, A. & Fu, Y. B. Phenotypic and molecular (RAPD) differentiation of four infraspecific
511 groups of cultivated flax (Linum usitatissimum L. subsp. usitatissimum). *Genet. Resour. Crop Evol.*
512 **53,** 77–90 (2006).
513 25. Soto-Cerda, B. J., Diederichsen, A., Ragupathy, R. & Cloutier, S. Genetic characterization of a core
514 collection of flax (Linum usitatissimum L.) suitable for association mapping studies and evidence of
515 divergent selection between fiber and linseed types. *BMC Plant Biol.* **13,** 78 (2013).
516 26. Prusinkiewicz, P., Erasmus, Y., Lane, B., Harder, L. D. & Coen, E. Evolution and development of
517 inflorescence architectures. *Science* **316,** 1452–1456 (2007).
518 27. Karlgren, A. *et al.* Evolution of the PEBP Gene Family in Plants: Functional Diversification in Seed
519 Plant Evolution. *Plant Physiol.* **156,** 1967–1977 (2011).
520 28. Igasaki, T., Watanabe, Y., Nishiguchi, M. & Kotoda, N. The FLOWERING LOCUS T/TERMINAL
521 FLOWER 1 family in Lombardy poplar. *Plant Cell Physiol.* **49,** 291–300 (2008).
522 29. Krieger, U., Lippman, Z. B. & Zamir, D. The flowering gene SINGLE FLOWER TRUSS drives
523 heterosis for yield in tomato. *Nat. Genet.* **42,** 459–463 (2010).
524 30. Shalit, A. *et al.* The flowering hormone florigen functions as a general systemic regulator of growth
525 and termination. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 8392–8397 (2009).
526 31. Lifschitz, E. *et al.* The tomato FT ortholog triggers systemic signals that regulate growth and
527 flowering and substitute for diverse environmental stimuli. *Proc. Natl. Acad. Sci. U. S. A.* **103,**
528 6398–6403 (2006).
529 32. Wang, Z. *et al.* The genome of flax (Linum usitatissimum) assembled de novo from short shotgun
530 sequence reads. *Plant J.* **72,** 461–473 (2012).

531 33. Baird, N. A. *et al.* Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS*
532 *One* **3,** (2008).
533 34. Etter, P. D., Bassham, S., Hohenlohe, P. lA, Johnson, E. A. & Cresko, W. A. SNP discovery and
534 genotyping for evolutionary genetics using RAD sequencing. in *Molecular Methods for*
535 *Evolutionary Genetics* (ed. Rockman, M. V.) 157–178 (Humana Press, 2011).
536 35. Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W. & Postlethwait, J. H. Stacks: building and
537 genotyping loci de novo from short-read sequences. *G3-Genes Genomes Genetics* **1,** 171–182
538 (2011).
539 36. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated
540 individuals. *Genome Res.* **19,** 1655–1664 (2009).
541 37. Fu, Y.-B., Diederichsen, A. & Allaby, R. G. Locus-specific view of flax domestication history. *Ecol.*
542 *Evol.* **2,** 139–152 (2012).
543 38. Fu, Y.-B. Genetic evidence for early flax domestication with capsular dehiscence. *Genet. Resour.*
544 *Crop Evol.* **58,** 1119–1128 (2011).
545 39. Allaby, R. G., Stevens, C., Lucas, L., Maeda, O. & Fuller, D. Q. Geographic mosaics and changing
546 rates of cereal domestication. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **372,** (2017).
547 40. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192,** 1065–1093 (2012).
548 41. Rosenberg, N. A., Li, L. M., Ward, R. & Pritchard, J. K. Informativeness of genetic markers for
549 inference of ancestry. *Am. J. Hum. Genet.* **73,** 1402–1422 (2003).
550 42. Yoo, S. J. *et al.* BROTHER OF FT AND TFL1 (BFT) has TFL1-like activity and functions
551 redundantly with TFL1 in inflorescence meristem development in Arabidopsis. *Plant J.* **63,** 241–253
552 (2010).
553 43. Mimida, N. *et al.* Functional divergence of the TFL1-like gene family in Arabidopsis revealed by
554 characterization of a novel homologue. *Genes Cells* **6,** 327–336 (2001).
555 44. Alvarez, J., Guli, C. L., Yu, X.-H. & Smyth, D. R. terminal flower: a gene affecting inflorescence
556 development in Arabidopsis thaliana. *Plant J.* **2,** 103–116 (1992).

557 45. Diederichsen, A. & Ulrich, A. Variability in stem fibre content and its association with other
558 characteristics in 1177 flax (Linum usitatissimum L.) genebank accessions. *Ind. Crops Prod.* **30,** 33–
559 39 (2009).
560 46. Swarts, K. *et al.* Genomic estimation of complex traits reveals ancient maize adaptation to temperate
561 North America. *Science* **357,** 512–515 (2017).
562 47. Xue, W. *et al.* Natural variation in Ghd7 is an important regulator of heading date and yield potential
563 in rice. *Nat. Genet.* **40,** 761 (2008).
564 48. Shannon, S. & Meekswagner, D. R. A mutation in the Arabidopsis TFL1 gene affects inflorescence
565 mersitem development. *Plant Cell* **3,** 877–892 (1991).
566 49. Bradley, D., Ratcliffe, O., Vincent, C., Carpenter, R. & Coen, E. Inflorescence commitment and
567 architecture in Arabidopsis. *Science* **275,** 80–83 (1997).
568 50. Bradley, D. *et al.* Control of inflorescence architecture in Antirrhinum. *Nature* **379,** 791–797 (1996).
569 51. Tian, Z. *et al.* Artificial selection for determinate growth habit in soybean. *Proc. Natl. Acad. Sci. U.*
570 *S. A.* **107,** 8563–8568 (2010).
571 52. McGarry, R. C. & Ayre, B. G. Manipulating plant architecture with members of the CETS gene
572 family. *Plant Sci.* **188,** 71–81 (2012).
573 53. Harlan, J. R., de Wet, J. M. J. & Price, E. G. Comparative Evolution of Cereals. *Evolution* **27,** 311–
574 325 (1973).
575 54. Meyer, R. S. & Purugganan, M. D. Evolution of crop species: genetics of domestication and
576 diversification. *Nat. Rev. Genet.* **14,** 840–852 (2013).
577 55. Spahillari, M., Hammer, K., Gladis, T. & Diederichsen, A. Weeds as Part of Agrobiodiversity.
578 *Outlook Agric.* **28,** 227–232 (1999).
579 56. Senda, T., Hiraoka, Y. & Tominaga, T. Inheritance of Seed Shattering in Lolium temulentum and L.
580 persicum Hybrids. *Genet. Resour. Crop Evol.* **53,** 449–451 (2006).
581 57. Thomas, H., Archer, J. E. & Turley, R. M. Evolution, Physiology and Phytochemistry of the
582 Psychotoxic Arable Mimic Weed Darnel (Lolium temulentum L.). in *Progress in Botany 72* (eds.

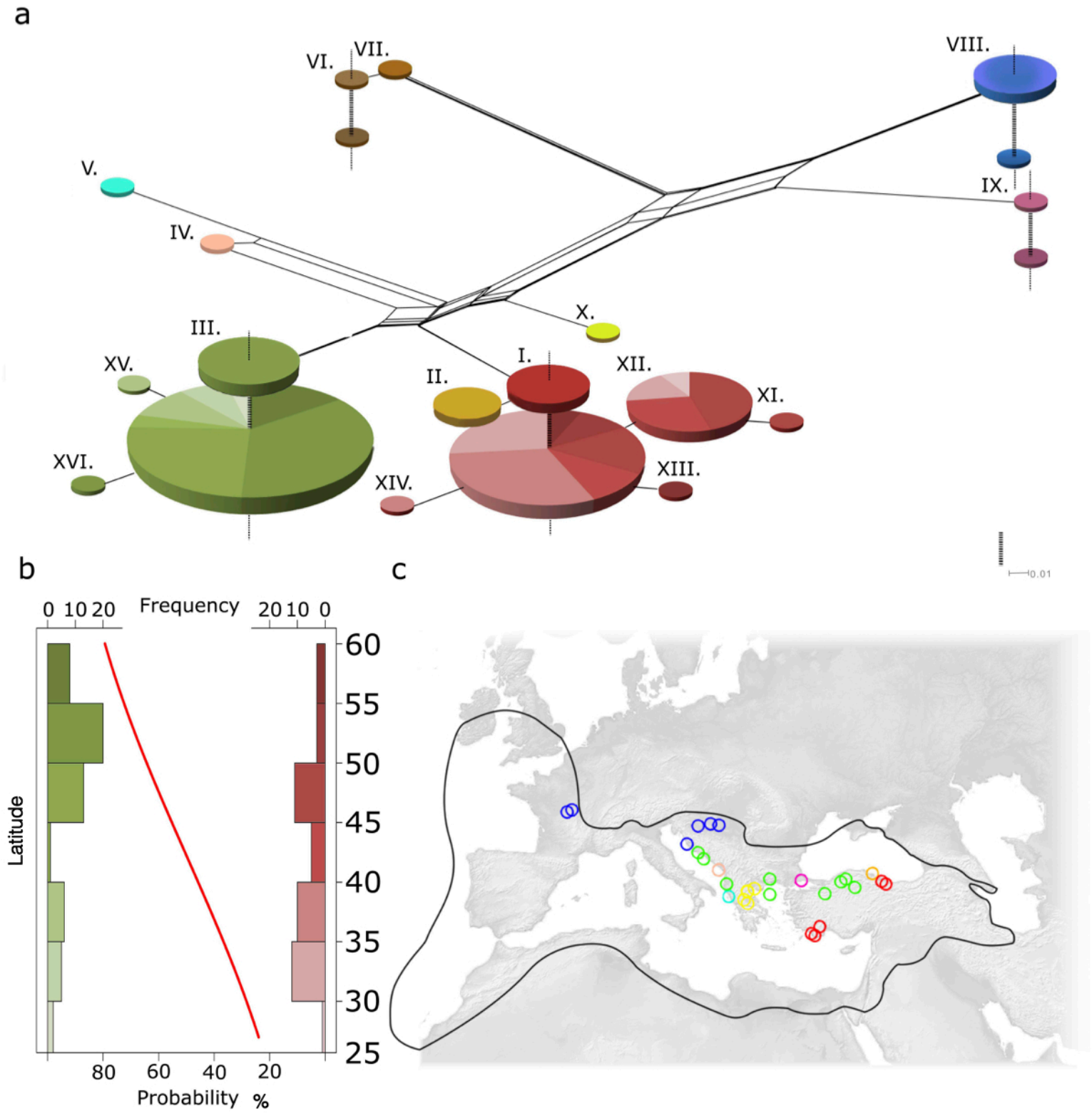
583 Lüttge, U. E., Beyschlag, W., Büdel, B. & Francis, D.) 73–104 (Springer Berlin Heidelberg, 2011).
584 58. Milla, R., Osborne, C. P., Turcotte, M. M. & Violle, C. Plant domestication through an ecological
585 lens. *Trends Ecol. Evol.* **30,** 463–469 (2015).
586 59. Milla, R., de Diego-Vico, N. & Martín-Robles, N. Shifts in stomatal traits following the
587 domestication of plant species. *J. Exp. Bot.* **64,** 3137–3146 (2013).
588 60. Takuno, S. *et al.* Independent Molecular Basis of Convergent Highland Adaptation in Maize.
589 *Genetics* **200,** 1297–1312 (2015).
590 61. Roffet-Salque, M. *et al.* Widespread exploitation of the honeybee by early Neolithic farmers. *Nature*
591 **527,** 226–230 (2015).
592 62. Rowland, G. & Cloutier, S. *Linum.ca database*. (2012).
593 63. Ronquist, F. *et al.* MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a
594 large model space. *Syst. Biol.* **61,** 539–542 (2012).
595 64. Kistler, L. Ancient DNA extraction from plants. *Methods Mol. Biol.* **840,** 71–79 (2012).
596 65. Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol.*
597 *Evol.* **23,** 254–267 (2006).
598 66. Tajima, F. Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism.
599 *Genetics* **123,** 585–595 (1989).
600 67. Fu, Y. X. & Li, W. H. Statistical test of neutrality of mutations. *Genetics* **133,** 693–709 (1993).
601 68. Ramos-Onsins, S. E. & Rozas, J. Statistical properties of new neutrality tests against population
602 growth. *Mol. Biol. Evol.* **19,** 2092–2100 (2002).
603 69. Andrews, S. FastQC A Quality Control tool for High Throughput Sequence Data. (2012).
604 70. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage
605 analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).
606 71. Whitlock, M. C. & Lotterhos, K. E. Reliable Detection of Loci Responsible for Local Adaptation:
607 Inference of a Null Model through Trimming the Distribution of FST*. *Am. Nat.* (2015).
608 doi:10.1086/682949

609 72. Hellens, R. P., Edwards, E. A., Leyland, N. R., Bean, S. & Mullineaux, P. M. pGreen: a versatile and
610 flexible binary Ti vector for Agrobacterium-mediated plant transformation. *Plant Mol. Biol.* **42,** 819–
611 832 (2000).
612 73. Weigel, D. & Glazebrook, J. *Arabidopsis: a laboratory manual*. (CSHL Press, 2002).

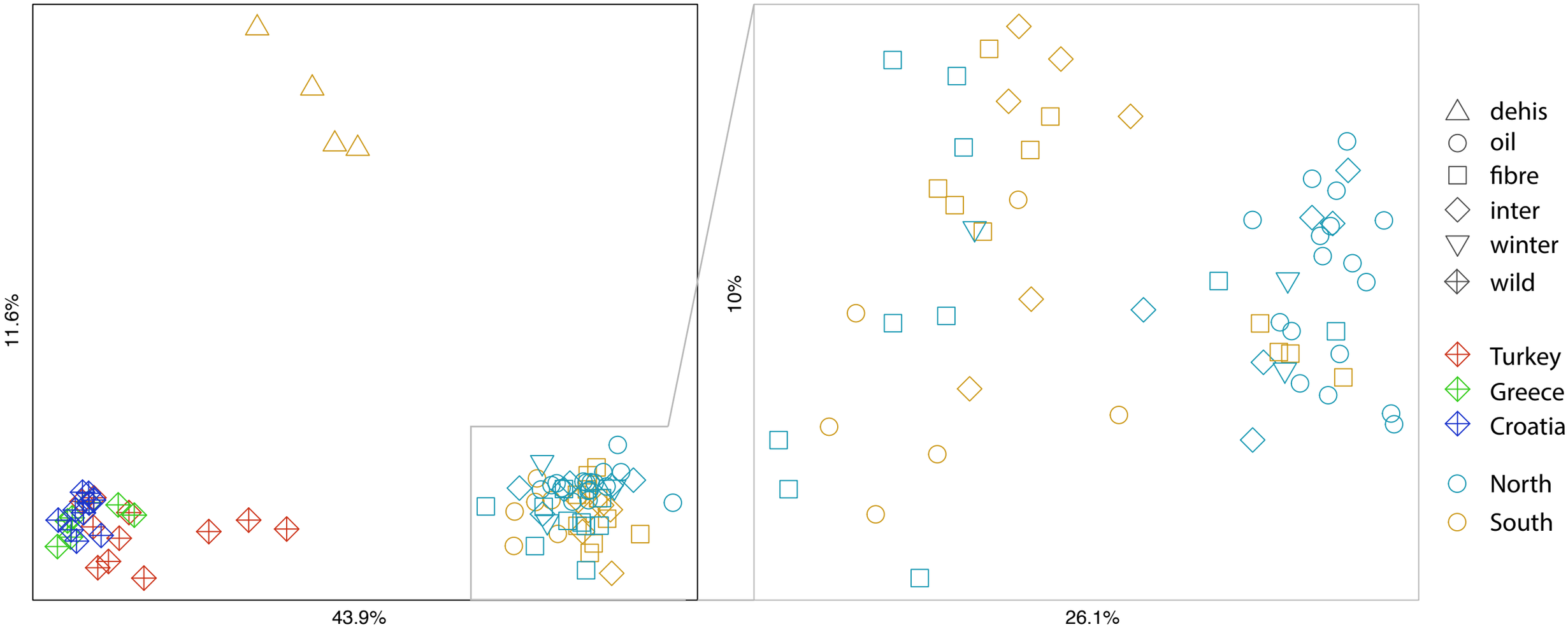
613 **Figure Legends**

614 **Figure 1**. Haplotype network of *LuTFL1* locus in cultivated and pale flax: A. Splits Tree network of pale
615 (top) and cultivated (bottom) flax, size of nodes is proportional to number of samples with the same
616 haplotype, continuous branches denote molecular distance between haplotypes, vertical dotted lines link
617 different species within same haplotype. B. Histogram showing latitudinal gradient of *LuTFL1* alleles in
618 cultivated flax, in green, frequency of northern haplotype cluster (*LuTFL1*.III), in red, frequency of
619 southern haplotype cluster (*LuTFL1*.I) with fitted logistic regression curve (p-value of 0.00144) reflecting
620 occurrence probability of northern haplotype in latitude gradient. C. Map of Europe marked with wild
621 distribution of pale flax (black line) and pale flax sampling locations (colours correspond to haplotypes in
622 splits network).
623
624 **Figure 2**. Population structure of pale flax and varieties of cultivated flax based on RADseq data: A.
625 Multidimensional Scaling Plot of pairwise nucleotide distances between flax
626 accessions separates three main flax populations. B. Admixture plot of flax four ancestral components.
627 Pale flax samples are arranged from east (Eastern Anatolia) to northwest (Central Europe), while samples
628 of cultivated flax were arranged based on admixture proportions. Above presented is the latitude at which
629 the individual plant was sampled. Below, individuals are annotated with allele of *LuTFL1* they carry;
630 colours are the same as in Figure 1.
631
632 **Figure 3**. Outliers to neutral allele frequencies between northern and southern cultivated flax populations:

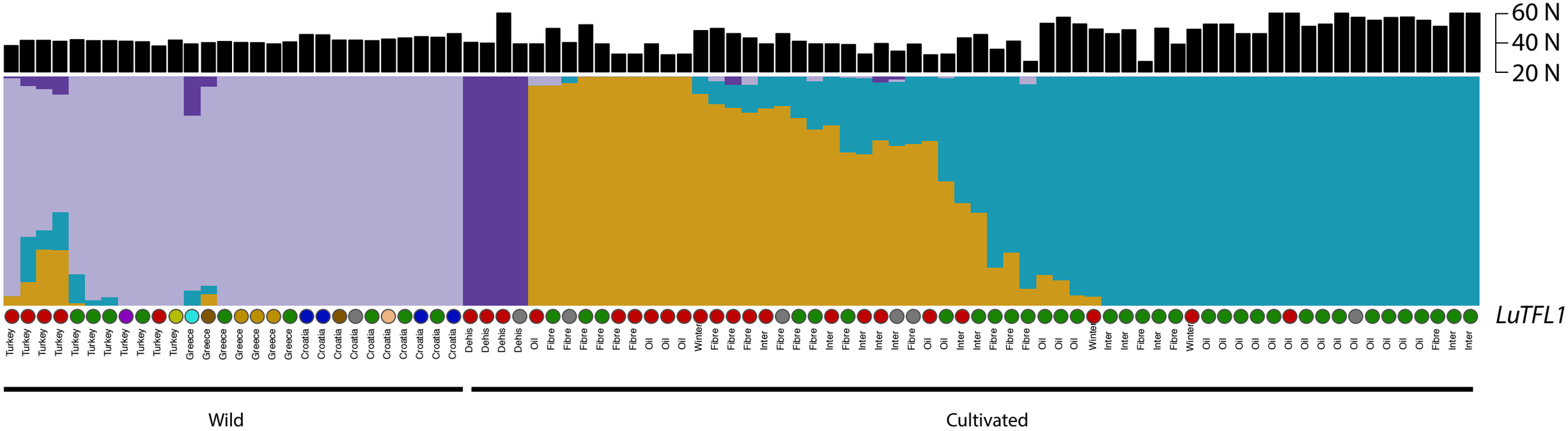
633 A. F_{ST} values for RAD neutral SNPs plotted as a function of heterozygosity. Outliers with elevated F_{ST}
634 and heterozygosity are marked with red circles. Outlier SNPs distinguishing alleles I and III in *LuTFL1*
635 locus are marked with black. B. Two dimensional allele frequency spectrum of RAD loci alleles between
636 northern and southern cultivated flaxes. Frequency of *LuTFL1*.III SNPs are marked as circles. To the
637 right: probability heat maps of sampling frequency combinations given the allele frequency spectrum
638 based on distance from the diagonal.
639
640 **Figure 4.** Strength of selection on *LuTFL1* III in modern and historic flaxes. Estimates of selection
641 coefficient over latitude and time, where latitude correlates to time of arrival of agriculture.
642
643 **Figure 5**. Flax *LuTFL1* gene complements *TFL1* mutant in *Arabidopsis thaliana*: A. Images of plants in
644 2x magnification at the median time to first open flower in *tfl1-2* mutant (20 days). Both Ler-0 wild type
645 and tfl1-2 mutant complemented with 35S::*LuTFL1* are not yet flowering at that time. B. Plants at their
646 median time to end of flowering in respective cohorts. Mutant *tfl1-2* ended flowering after 23 days and
647 reached average height of 69 mm; by contrast Ler-0 wild type and 35S::*LuTFL1* plants flowered until
648 36th and 40th day reaching 134 and 211 mm height respectively.
649
650 **Figure 6**. Plots summarizing parameters of flax inflorescence simulated with pgrowth model over
651 different base levels of *TFL1* expression. Shown below charts are representative flax architectures
652 associated with base levels of *TFL1*. A. Chart and architecture diagrams for plants grown at N 60° latitude
653 with assumed floral initiation threshold set to 14 hours. B. Chart and architecture diagrams for plants
654 grown at N 35° latitude with assumed flowering threshold set to 14 hours.

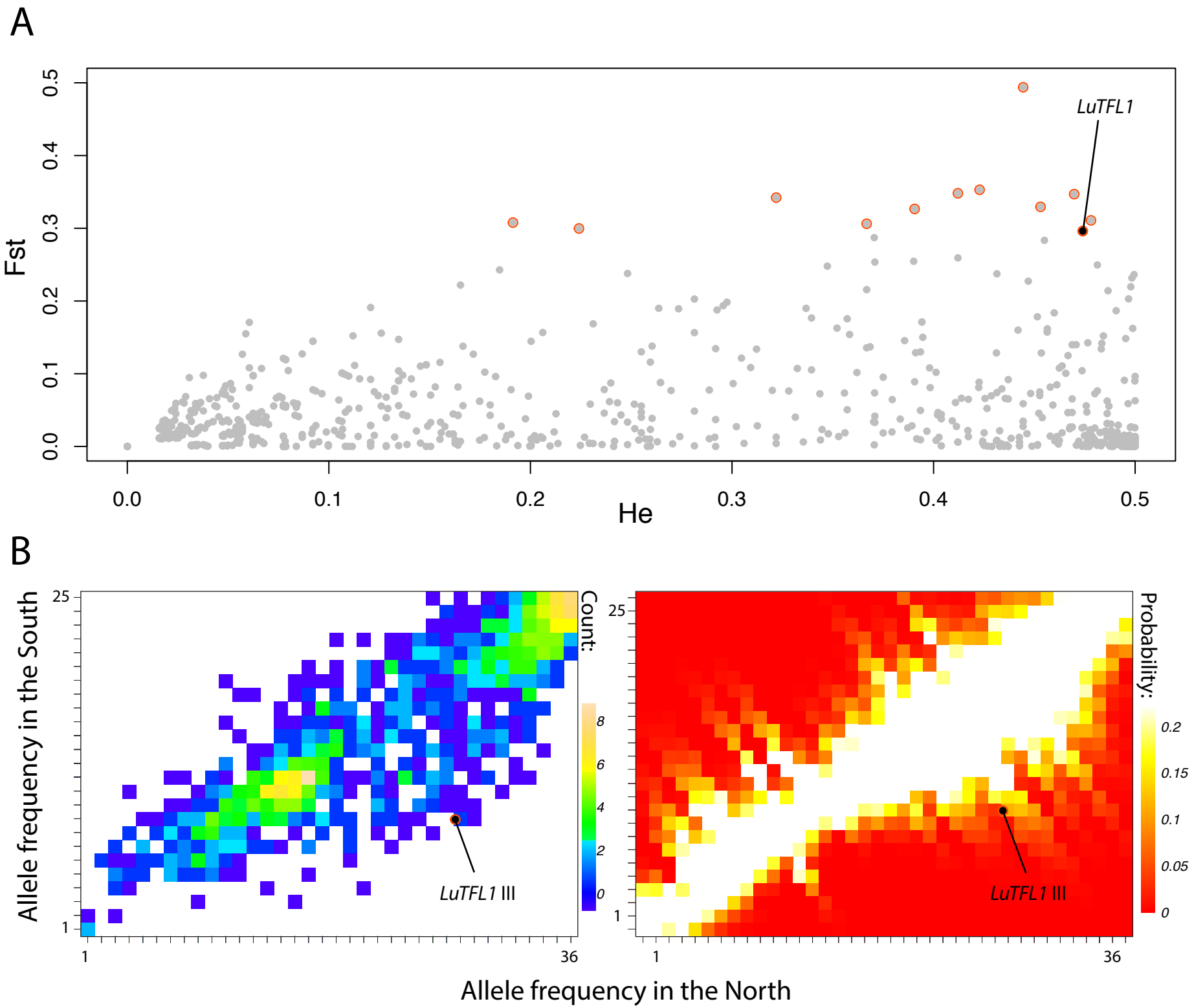

a
VI.
VII.
VIII.
V.
IX.
IV.
X.
III.
XV.
I.
XII.
II.
XI.
XVI.
XIII.
XIV.
0.01
b
0 10 20
Frequency
20 10 0
60
55
50
45
40
35
30
25
Latitude
80
60
40
20
Probability %
c

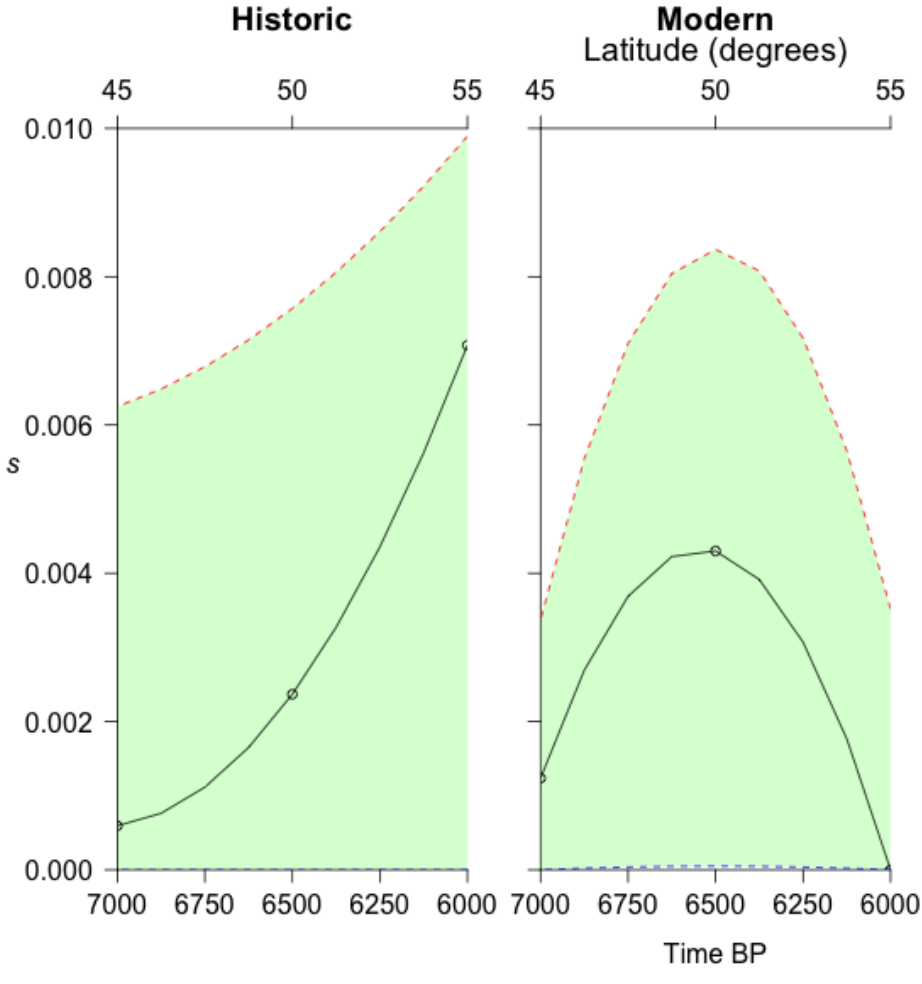
A


11.6%
43.9%
10%
26.1%
dehis
oil
fibre
inter
winter
wild
Turkey
Greece
Croatia
North
South

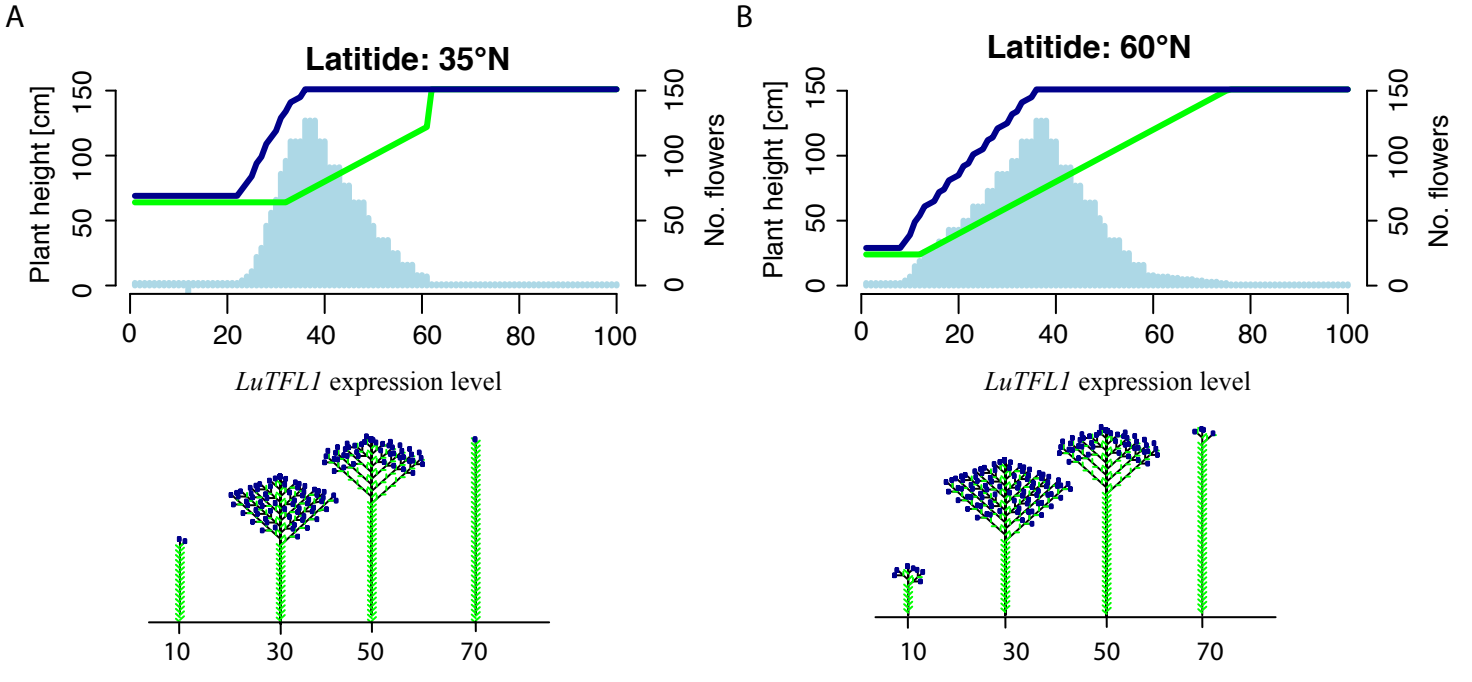
B


60 N
40 N
20 N
LuTFL1
Wild
Cultivated


A
Fst
He
LuTFL1
0.0
0.1
0.2
0.3
0.4
0.5
B
Allele frequency in the South
Allele frequency in the North
Count:
Probability:
LuTFL1 III
1
25
36
0
2
4
6
8
0.05
0.1
0.15
0.2


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

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


A
Latitide: 35°N
Plant height [cm]
No. flowers
0
50
100
150
0
20
40
60
80
100
LuTFL1 expression level
10
30
50
70
B
Latitide: 60°N
Plant height [cm]
No. flowers
0
50
100
150
0
20
40
60
80
100
LuTFL1 expression level
10
30
50
70