Original Research

Genome-wide signatures in flax pinpoint to adaptive evolution along its ecological gradient

Demissew Sertse[1,2,3], Frank M. You[2], Sylvie Cloutier[1,2,*]

[1]*Department of Biology, University of Ottawa, Ottawa, ON K1N 6N5, Canada,* [2]*Department of Agriculture and Agri-Food, Ottawa Research and Development Center, Ottawa, ON K1A 0C6, Canada,* [3]*Aquatic and Crop Resource Development Research Centre, National Research Council of Canada, Saskatoon, SK S7N 5C2, Canada*

TABLE OF CONTENTS

1. Abstract

Background: Flax is one of the eight founder crops of agriculture. It is believed to have been domesticated as a long-day plant that has since spread to survive in a wide range of eco-geographic regions extending from the warm Indian subcontinent to the low latitude east African highlands and to the cool and high-latitude Eurasia. Understanding the genetic basis underlying its adaptation and selection events throughout its dispersion is essential to develop cultivars adapted to local environmental conditions. **Methods**: Here we detected genetic signatures of local adaptation and selection events of flax based on 385 accessions from all major flax growing regions of the world using genome scan methods and three genomic datasets: (1) a genome-wide dataset of more than 275K single nucleotide polymorphisms (SNPs), (2) a filtered dataset of 23K SNPs with minor allele frequency >10% and, (3) a 34K exon-derived SNP dataset. **Results**: Principal component (PC) and fixation index (F_{ST})-based genome scans yielded consistent outlier SNP loci on chromosomes 1, 8, 9 and 12. Additional loci on chromosomes 3, 7, 8, 10, 11, 13 and 14 were detected using both the PC and F_{ST} methods in two of the three datasets. A genome-environment association (GEA) analysis using the 23K dataset and the first PC of cropping season temperature, day-length and latitude identified significant SNPs on chromosomes 3, 7, 9 and 13. **Conclusions**: Most of the loci detected by the three methods harbored relevant genes for local adaptation, including some that play roles in day-length, light and other biotic and abiotic stresses responses. Such genetic signatures may help to

Submitted: 19 August 2021 Revised: 11 November 2021 Accepted: 22 November 2021 Published: 30 December 2021

select pre-breeding materials potentially adapted to specific growing niches prior to field performance trials. Given the current low genotyping cost and freely available environmental data, the genome scans along with GEA can readily provide opportunity to sort out materials suitable to various environmental conditions from large set of germplasm in gene banks and/or *in situ*, thereby assisting the breeding and genetic conservation efforts.

2. Introduction

Adaptation of a species to a gradient of environmental conditions is attributed to the phenotypic plasticity and genetic variation within the gene pool of the species [1, 2]. While phenotypic plasticity to a wide range of environmental conditions maintains genetic homogeneity [3], variants permit differential adaptation to local environments via selection and prompts genetic divergence of populations [4]. Climate factors act as major forces in the selection of variants increasing fitness from a gene pool and consequently drive local adaptation and genetic divergence [5]. Early domesticated crops have spread to extensive eco-geographic ranges, far from where their wild ancestors originated. Their success along latitudinal gradients is usually governed by their phenological behavior in response to spatial variations in climate and related factors, especially day-length and temperature [6]. The agricultural founder crops that are believed to have been domesticated in the Fertile Crescent were presumed to be adapted to vernalization and long-day flowering [6]. These crops are now well spread throughout the world and adapted to a range of eco-geographic conditions; consequently, post-domestication genetic divergence can be observed among eco-spatially separated populations [7]. Such divergence is presumably attributed to loci selected in specific environments with high coefficients of genetic differentiation between populations that specify the genetic basis underlying the adaptation [8].

Understanding the genetic signatures underlying local adaptation of landraces and/or cultivars is an essential step for developing hastened and effective breeding and conservation schemes in crops. Recent advances in sequencing technologies enable the production of large genomic datasets, and their genome-wide scans enable the detection of outlier loci that are signatures of local adaptations [9–11]. Outlier loci detection based on genetic differentiation without prior knowledge of the driving environmental forces and genome-environment association (GEA) have recently been used to detect and cross-validate these outliers as important signatures of adaptation to diverse environments [12, 13]. Alleles under selection in a specific environment experience a higher fixation rate than alleles of neutral effects for the environment [8]. In the absence of environmental condition records, such as climate data, genome scan techniques can be used to discriminate loci harboring alleles under heavy selection pressure. These techniques along with GEA based on multi-year environmental data records have become useful to study the genetic basis of adaptation to climatic and other environment-specific conditions [13, 14]. Genome scans based on outlier loci detection and GEA have been used to discover local adaptation signatures in several plant species like barrel clover [15], sorghum [16], barley [17], maize [18], oat [19], common bean [20], crop wild relatives [21–23] and *Arabidopsis* [24].

Flax (*Linum usitatissimum* L.) is one of the eight founding crops of agriculture in the fertile crescent [25]. It is believed to have first been domesticated in present-day Syria [26] and spread to nearly all of its current eco-geographic distribution in the Old-World millennia ago. As archeological and paleontological evidences show, flax was cultivated in Mesopotamian and Egyptian irrigated fields ca. 7000 BC [25] and Europe ca. 6000 BC [27]. It started being used for its fiber and seeds in Western Europe ca. 5500 BC [28] and reached as far as China by 3000 BC [29]. Some of these post-domestication distributions eventually became secondary centers of diversification [30]. Today, the crop is grown from the warm Indian subcontinent to the temperate zones of Europe and America and the low latitude North-East of African Highlands. Genetic variation and population structure of flax are majorly attributed to environmental and anthropogenic selection pressures [31]. The genetic variation of flax correlates with latitudinal gradient-related variables such as the day-length during the cropping months [31, 32]. With the assumption that such genetic variations are attributed to loci for environmental adaptation, we performed two genome scans, namely principal component analysis (PCA) and Wright's fixation index (F_{ST}), and a GEA analysis to detect outlier loci for the first principal component (PC) of cropping month temperature, day-length, and latitude (1) to identify loci contributing to strong variations among the populations, (2) to assess the genetic bases for local adaptation, and (3) to identify genomic regions underlying adaptation to any of these eco-geographic parameters.

3. Materials and methods

3.1 Plant materials

The plant materials used in this project include 385 accessions that were collected from more than 35 flax-growing countries. Approximately 60% of the germplasm originated from the Old-World, i.e., where flax has been cultivated for millennia, including the postulated centers of origin and secondary centers of diversity.

3.2 Genotyping and data quality control

Genome-wide SNP datasets were extracted from a 1.7M SNP dataset originally generated by resequencing the flax core collection ($n = 407$) using the Illumina HiSeq 2000

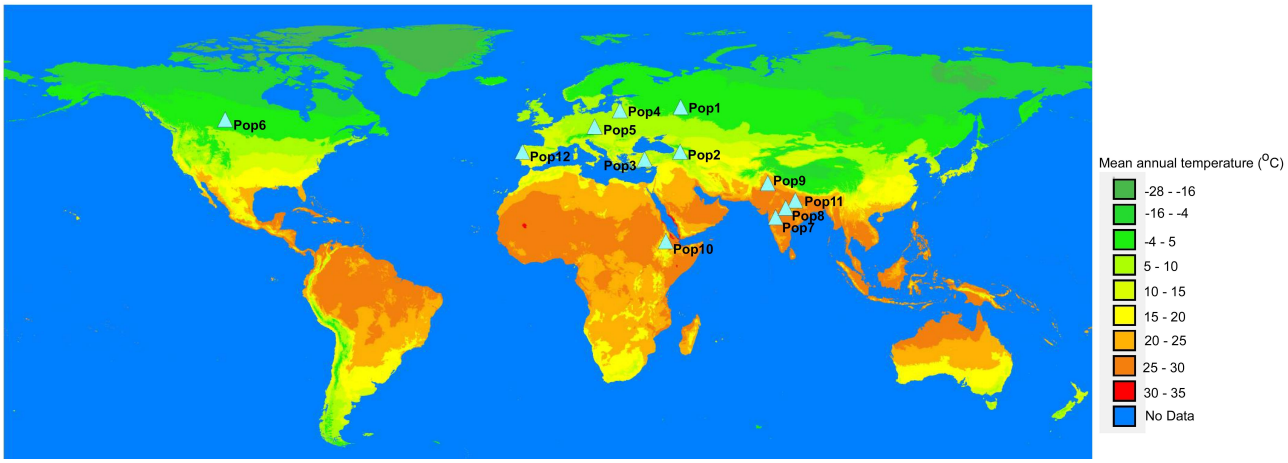

Fig. 1. World map overlaid with mean annual temperatures and showing the geographic locations assigned to the populations (Pop1-12).

platform [31, 33]. SNPs were filtered with a 90% call rate and a minimum minor allele frequency (MAF) of 5%. To reduce the number of missing SNPs, imputation was performed using LinkImpute [34] implemented in TASSEL v5 [35] with default parameters, but with the maximum distance between sites set to 100 kb. Individuals with >10% missing SNPs after imputation were omitted. The resulting SNP dataset is referred to as the 277K dataset. A second dataset was prepared by extracting from the 277K dataset only the SNPs without missing data and a MAF >10%. The resulting dataset is later defined as the 23K dataset. A third dataset containing only SNPs located in exons was filtered from the original SNP dataset using an 80% call rate and a MAF >5%. This dataset was also imputed. The resulting exon dataset was further filtered to retain only SNPs with a 90% call rate after imputation. This set is referred to as the 34K exon dataset.

3.3 Population structure and assignment of genotypes to clusters

In order to define the appropriate number of ancestral populations, estimates were obtained using the pcadapt tool [36]. Here eigenvalues were computed for 50 PCs and visualized into scree plots. By applying the cattle rule [37], the PC to the left of the last PC with eigenvalues that deviates from the smooth line is considered to be the most appropriate number of populations. To corroborate the estimated number of populations, a cross validation [38] was performed using ADMIXTURE [39] which uses a Bayesian clustering approach. PCA was performed based on the estimated number of populations (*K*) and a neighbor-joining (NJ) phylogenetic analysis was carried out using TASSEL [35]. The PCA based on the first three PCs and the NJ tree were visualized using ggplot2 [40] in R and interactive Tree Of Life (iTOL) [41], respectively.

Following the PCA clustering using the 23K dataset, individuals were assigned to one of 12 populations. To estimate the genetic variation between populations, pairwise genetic differentiation between populations was estimated with the fixation index (F_{ST}) based on population size estimated after 10000 permutations at α = 0.05 and 0.01 using Arlequin 3.5 [42]. To quantify the contribution of each SNP to the variation, the F_{ST} of individual SNPs was obtained for each dataset using the R package LEA which performs Landscape and Ecological Association (LEA) analyses [43]. Haplotypes were assessed using the web-based tool SNiPlay v3 [44] based on SNPs with significant ($p < 0.05/n$, where n = number of SNPs in the data set) F_{ST} values at α = 0.05 from the 23K dataset for each chromosome. Haplotypes were defined as those present in at least three individuals within the overall germplasm collection or one of its populations. To determine the distribution of private haplotypes for each population, haplotypes observed in at least three individuals were considered for each chromosome. To understand the effect of bottlenecks in haplotype, the per population gene diversity was calculated for each of the 12 populations using GENEPOP v4.7.5 [45] in R package genepop [46].

3.4 Environmental data curation

The geographic region/country that best represents the population was assigned for each population based on the passport data, which indicates the origin of dominant members of a population. The representative geographic areas were inferred based on the history of cultivation of flax in each of the regions (**Supplementary Table 1**). Therefore, geographic coordinates of a representative district or a province with a long history of flax cultivation were used as environmental data tags. To get insight into some environmental factors, mean annual temperature data downloaded from www.wroldclim.org was overlaid onto the world map using DIVA-GIS v7.5 (LizardTech Inc, Portland, OR, USA) [47]. The representative coordinates of each population were positioned onto the world map along with the annual temperature data (Fig. 1).

The monthly mean temperature and day-length data for each population were downloaded from 22 years of records available in the NASA database (https://power.larc.nasa.gov/). The average records of these periods were considered. The cropping period for the selected regions was determined based on the major crop calendars from FAO (http://www.amis-outlook.org/amis-about/calendars/soybeancal/en/) for all populations that follow similar northerly cropping patterns. Because flax is a Rabi crop in some parts of the world such as India and Pakistan, different cropping calendars (https://nfsm.gov.in/nfmis/rpt/calenderreport.aspx and http://namc.pmd.gov.pk/crop-calender.php) were used for genotypes of Indian and Pakistani origins, respectively.

3.5 Adaptive loci assessment

To detect outlier loci associated with local adaptation signatures, two genome-scan methods were used: individual SNP F_{ST} using LEA [43] and the principal component-based to detect local adaptation (PCADAPT) as implemented in the R package pcadapt [36]. Each dataset was analyzed separately using both methods and results were compared to identify robust genetic signature loci. To capture adaptive loci associated with specific environmental conditions, a GEA analysis was performed following the latent factor mixed model 2 (LFMM2) using lfmm in R [48] based on a PC of average cropping season temperature, day-length and latitude of the representative locations of each population for the 23K dataset. The use of a representative PC was chosen because the environmental factors were highly correlated (**Supplementary Fig. 1**). For each population and all three datasets, allelic frequencies were calculated and summarized for all loci detected in all the datasets by at least two methods.

Genomic regions that consistently showed a strong association with an environmental factor across the datasets for both of the first two methods (PCADAPT, and F_{ST}) were retained as potential signatures of local adaptation. For the GEA, genome association with the first PC, which explained more than 97% of the variation of the three environmental factors of latitude, cropping season average temperature and day-length among the populations, was applied (**Supplementary Fig. 2**).

3.6 Candidate gene inference

Genomic regions spanning 20 kb up and downstream of the most significant SNP loci were examined for the presence of candidate genes for local adaptation using the flax reference genome annotation [33]. Linkage disequilibrium (LD) between candidate genes and their associated marker was calculated using gpart package in R [49]. Candidate gene putative functions were further assessed through the identification of their *Arabidopsis* orthologs (www.arabidopsis.org) and through a literature search evidencing their role(s) in adaptation.

4. Results

4.1 SNPs and genetic structure

Filtering of the datasets yielded a total of 277,399, 23,592 and 34,451 SNPs for the two genome-wide (277 and 23K) and the exon-based (34K) datasets, respectively. The former two contained 385 genotypes, while the exon-based dataset included 393. The estimated number of populations [36] was 12 (**Supplementary Fig. 3**) which is also consistent with the result from cross validation technique where the lowest error was obtained at K = 13 (**Supplementary Fig. 4**). The populations tended to follow geographic gradient where Pop1-5 were dominated by Eurasian accessions (**Supplementary Table 1**. Pop6, Pop10 and Pop12 contained mostly Canadian, Abyssinian and Mediterranean accessions, respectively, whereas the majority of the South Asian accessions grouped into the remaining three populations (Fig. 2A,B). The NJ phylogenetic analysis clustered the accessions slightly differently, i.e., reflecting both geographic and historical-use patterns of variation (Fig. 2C). Accessions from Old-World flax-growing regions tended to have longer branches compared to those from the New-World regions. In addition, the majority of the fiber types clustered in the single clade Pop1_FIB (Fig. 2C).

Pairwise population differentiation was significant ($p < 0.01$) for all comparisons except between one of the temperate populations (Pop4) and a population dominated by Canadian cultivars (Pop6). Most populations dominated by Eurasian and Canadian accessions displayed low differentiations (Table 1). Populations with the strongest (F_{ST} = 0.76) differentiation were the fiber-dominated Pop1 and the South-Asian Pop11 (Table 1).

Pop1 and Pop2 harbored the most common haplotype from each of the 15 chromosomes (Table 2). All accessions considered, chromosome 1 displayed the lowest percentage of common haplotype with ~19% while chromosome 4 had the highest with 78%. Some populations, such as Pop8, 9 and 11, contained very few common haplotypes while others, such as Pop1, 2, 4 and 6, comprised a large proportion of individuals with common haplotypes across all chromosomes (Table 2).

Private haplotypes are those found in a single population. Based on this definition, 30 private haplotypes were observed in seven of the populations (**Supplementary Table 2**). The South Asian (Pop11) and Abyssinian (Pop10) populations contained 48% and 20% of the private haplotypes, respectively (Fig. 3). The most frequently observed private haplotype was Chr14: Hap1 which was present in 97% of the accessions of the South Asian population (Pop11) representing 8.6% of all accessions (**Supplementary Table 2**). The highest gene diversity was in Pop12 followed by Pop5 which putatively originated from Mediterranean Portugal and Turkey regions respectively (**Supplementary Table 1**). In contrast, Pop7 and Pop10 displayed no diversity (Table 2).

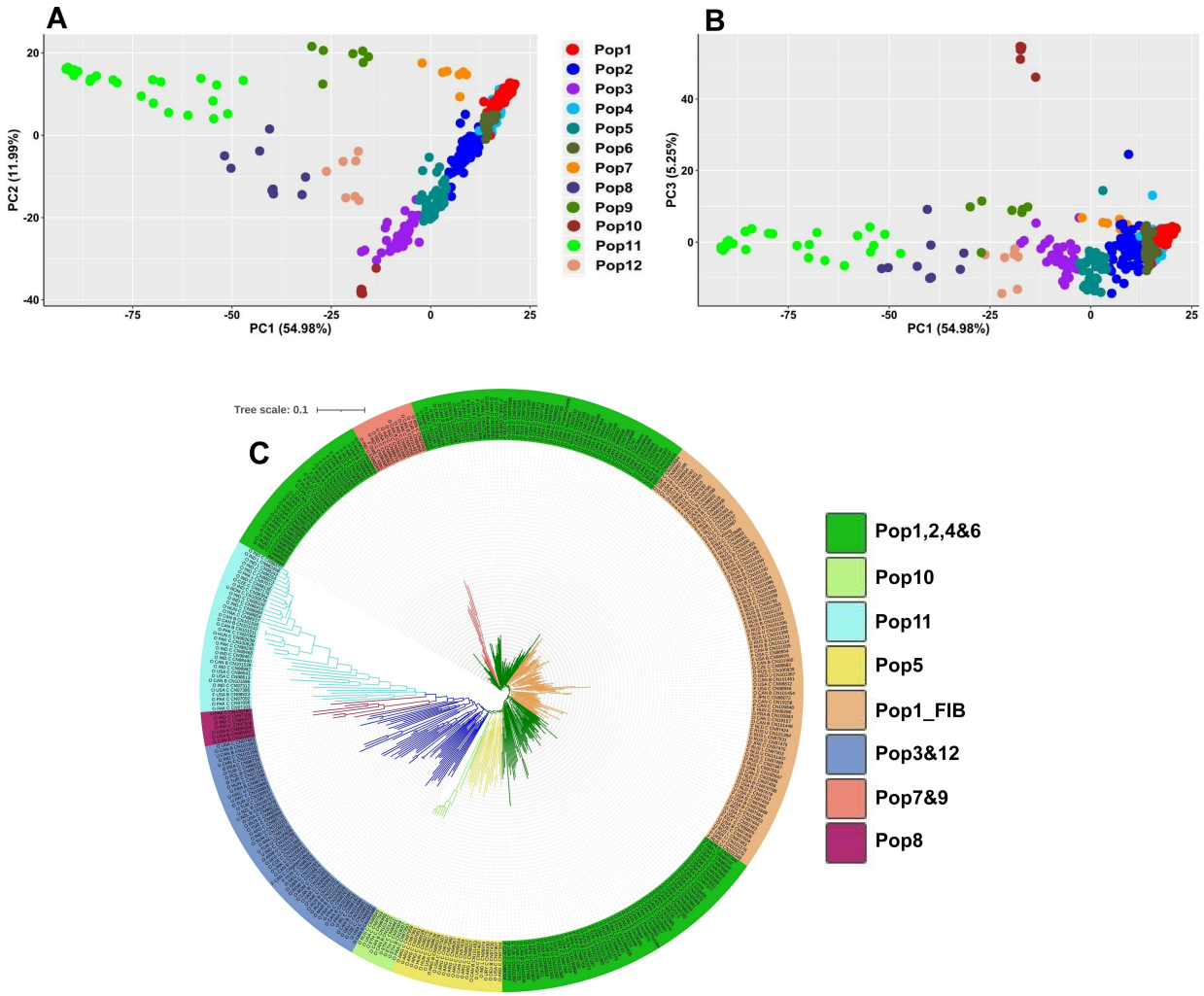

Fig. 2. Population genetic structure based on the 23K SNP dataset. (A) and (B) principal component analysis (PCA) clustering of the accessions based on the first three principal components (PC1, PC2 and PC3). (C) Neighbor-joining (NJ) phylogenetic tree where accessions from the Old-World flax growing regions tended to have longer branches. Accession names in NJ indicate the type (O, oil; F, fiber; U, unknown), the country of origin, the breeding status (C, cultivar; B, breeding material; L, landrace) followed by the accession name.

4.2 Adaptive SNP detection and their linked genes

A total of 19 outliers were detected, including six on chromosomes 1, 8, 9 and 12 that were detected in all three datasets with both PCADAPT and F_{ST} methods (Fig. 4 and **Supplementary Table 3**). The allele frequencies at outlier loci detected in all the three datasets differed among populations and where the frequencies were higher in populations from old flax growing regions for either of the two alleles at a locus (Fig. 5). Candidate gene searches revealed several genes with putative function in ecological adaptation. For example, the Chr9:9310330 locus harbors gene *Lus10006147*, an ortholog to *Arabidopsis AT4G15530*, which encodes a pyruvate orthophosphate dikinase. The locus marked by Chr1:5466653 harbors the predicted flax gene *Lus10011967*, whose ortholog *AT4G18130* encodes a phytochrome E (PHYE). The other locus on chromosome1 (Chr1:10413007) includes two genes: *Lus10022627* and *Lus10022628*, which are orthologs of *AT2G36800* and *AT2G36780*, encoding URIDINE 5'-DIPHOSPH (UDP)-GLYCOSYLTRANSFERASE 73C5 (UGT73C5) and UDP-glycosyltransferase 73C3 (UGT73C3), respectively. The Chr8:2932993 locus contained the tandemly repeated genes *Lus10012356*, *7* and *8*, orthologous to the *Arabidopsis* gene *AT3G07870* that encodes the F-BOX PROTEIN92 (FBX92). The Chr7:10407650 locus, one of the most significant SNPs of the 277K (p~3.48E-52 for PCADAPT; $p < 8.1 \times 10^{-224}$ for F_{ST}) and 23 K (p~6. 7 $\times 10^{-28}$ for PCADAPT; p~1.94×10^{-66} for F_{ST}) datasets contains the predicted flax gene *Lus10000371*, which is orthologous to the early flowering gene *AT1G17455* encoding ELF4-L4.

PC-based genome scan of the 34K dataset detected several significant SNPs on chromosome 1 between positions 3612570 and 3671655 (Fig. 4) with p-values between 2.1×10^{-16} and 6.9×10^{-12}. This locus contains six genes

Table 1. Genetic differentiation (F_{ST}) among populations.

Population	Pop1	Pop2	Pop3	Pop4	Pop5	Pop6	Pop7	Pop8	Pop9	Pop10	Pop11
Pop2	0.09										
Pop3	0.35	0.13									
Pop4	0.05	0.05	0.22								
Pop5	0.24	0.06	0.06	0.14							
Pop6	0.06	0.03	0.22	0.02ns	0.13						
Pop7	0.22	0.13	0.23	0.21	0.18	0.17					
Pop8	0.63	0.38	0.24	0.48	0.29	0.49	0.39				
Pop9	0.54	0.33	0.30	0.45	0.31	0.42	0.27	0.32			
Pop10	0.72	0.51	0.38	0.69	0.43	0.62	0.75	0.53	0.69		
Pop11	0.76	0.61	0.51	0.67	0.55	0.67	0.62	0.34	0.48	0.68	
Pop12	0.53	0.27	0.18	0.38	0.19	0.38	0.34	0.23	0.32	0.54	0.48

ns, non significant variation at $p < 0.05$ and 0.01. All other pairs are significant at $p < 0.01$.

Table 2. Frequency distribution of the most common haplotype of each chromosome in the overall population and in each of the 12 populations and population gene diversity.

Chr[1]	Overall	Pop1	Pop2	Pop3	Pop4	Pop5	Pop6	Pop7	Pop8	Pop9	Pop10	Pop11	Pop12
Chr1	18.96	40.83	15.52	0.00	11.11	0.00	25.00	28.57	0.00	0.00	0.00	0.00	0.00
Chr2	64.68	97.50	70.69	40.00	88.89	30.77	95.45	57.14	0.00	14.29	28.57	0.00	0.00
Chr3	29.09	54.17	22.41	0.00	66.67	56.41	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Chr4	78.18	99.17	94.83	71.43	100.00	92.31	90.91	71.43	11.11	0.00	0.00	0.00	28.57
Chr5	75.06	100.00	81.03	45.71	100.00	92.31	90.91	14.29	0.00	0.00	100.00	0.00	57.14
Chr6	51.43	94.17	58.62	14.29	55.56	30.77	52.27	14.29	0.00	0.00	0.00	0.00	0.00
Chr7	51.95	56.67	68.97	37.14	77.78	74.36	65.91	0.00	0.00	0.00	85.71	0.00	14.29
Chr8	64.42	55.83	74.14	97.14	77.78	100.00	70.45	0.00	44.44	14.29	100.00	5.88	85.71
Chr9	49.61	56.67	65.52	25.71	88.89	64.10	75.00	0.00	0.00	0.00	0.00	0.00	28.57
Chr10	59.22	56.67	81.03	48.57	88.89	97.44	75.00	0.00	44.44	0.00	0.00	0.00	57.14
Chr11	35.32	54.17	43.10	0.00	83.33	10.26	61.36	0.00	0.00	0.00	0.00	0.00	0.00
Chr12	32.47	51.67	37.93	0.00	77.78	7.69	50.00	0.00	0.00	14.29	0.00	0.00	0.00
Chr13	60.00	100.00	58.62	14.29	88.89	28.21	97.73	14.29	0.00	0.00	0.00	0.00	14.29
Chr14	65.19	57.50	84.48	97.14	88.89	89.74	75.00	0.00	33.33	14.29	100.00	0.00	57.14
Chr15	62.86	62.50	84.48	77.14	55.56	92.31	72.73	28.57	0.00	0.00	100.00	2.94	42.86
Div		0.096	0.189	0.444	0.209	0.3914	0.206	0.000	0.222	0.286	0.000	0.337	0.571
No.Ind	385	120	58	35	18	39	44	7	9	7	7	34	7

[1]Chr, Chromosome; Div, gene diversity; No.Ind, number of individual.

of which five were orthologous to *AT2G30140* encoding UDP-GLUCOSYL TRANSFERASE 87A2 (UGT87A2), and one was orthologous to *AT2G30150* encoding UDP-GLUCOSYL TRANSFERASE 87A1 (UGT87A1). All *UGT* genes at this locus were in strong LD with their associated SNPs (**Supplementary Fig. 5**).

GEA based on PC1 associated with latitude, average day-length and temperature during the cropping season performed using the assigned coordinates of the population resulted in significant SNPs on chromosomes 3, 7, 9 and 13 (Fig. 6). The significant locus at Chr3:16799360 harbored multiple candidate genes for flowering time regulation, including the *AT2G29950* ortholog *Lus10040667* that is predicted to be an *EARLY FLOWERING LOCUS-LIKE1* (*ELF4-L1*) gene (Table 3). The locus Chr7:16799360 contains *AT1G67170* ortholog *Lus10015495* that is predicted to encode a *FLOWERING LOCUS C EXPRESSOR-LIKE 2* (*FLL2*) gene (Table 3).

5. Discussion

Understanding the adaptation of genotypes to environmental gradients is important for breeding and conservation [50]. Natural selection in a wide-range of environmental gradients leads to genetic divergence and selection of adapted variants [51]. Environmental variations along the latitudinal gradient are major forces of selection that lead to genetic divergence in plants [52]. Phenological variations are some of the better known patterns in plants along the latitudinal biosphere [53]. Flax is a species that spread through nearly the full span of crops' latitudinal range, being grown from the low latitude of the East African highlands to the high latitude of temperate regions, as well as from the warm South Asian to the cool Eurasian climates. Apart from natural selection, anthropogenic influences, such as selection and germ introduction into new niches, also play major roles in the success and spread of

Table 3. Outlier loci associated with the first principal component (PC) of latitude, temperature and day-length of the cropping seasons and their candidate genes proposed based on the annotation of their *Arabidopsis* orthologs.

Outlier SNP	*p* value	Candidate gene	Gene position	*Arabidopsis* ortholog	Gene annotation
Chr3:6169770	5.4×10^{-11}	*Lus10040667*	6206833-6207246	*AT2G29950*	*ELF4-L1*
Chr3:9175844	1.0×10^{-10}				
Chr7:16799360	4.0×10^{-10}	*Lus10015495*	16806888-16808239	*AT1G67170*	*FLL2*
Chr7:16798937	4.1×10^{-10}				
Chr9:11962263	4.4×10^{-12}	*Lus10028960*	1178447-1180169	*AT5G57280*	*RID2*
Chr9:11798402	5.3×10^{-11}				
Chr13:878058	5.1×10^{-8}				
Chr13:10377859	4.1×10^{-7}	*Lus10010694*	1019336-1020295	*AT2G02540*	*HB21/ZFHD4*
		Lus10010693	1025094-1025489	*AT2G02540*	*HB21/ZFHD5*

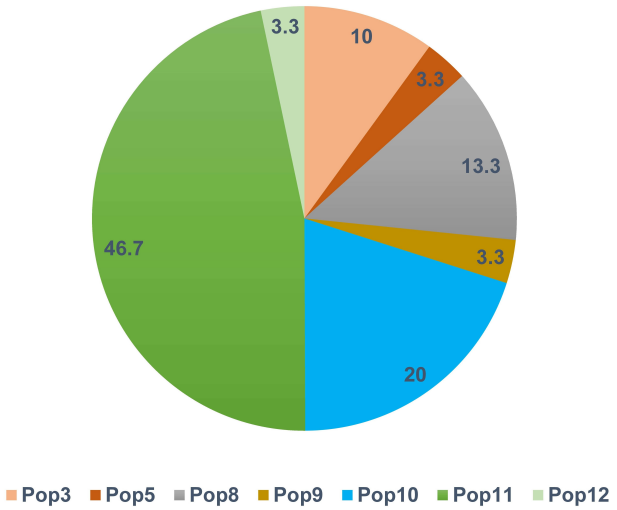

Fig. 3. Frequency distribution of private haplotypes by population. Only seven of the 12 populations had private haplotypes.

crop plants, including flax [31]. This study provides insights into genetic signatures of global scale adaptation of flax across its wide range of habitats.

5.1 Genetic structure and differentiation

The genetic structure observed in this experiment is consistent with what was establish in previous works, which demonstrated the clustering of accessions attributed to their eco-geographic origin [31, 54]. The relatively high differentiation and high concentration of private haplotypes in populations dominated by South Asian accessions may suggest the adaptation of the crop to warm regions [55, 56]. This is unlike the cool and/or temperate habitats where flax is widely grown [57]. Flax in this warm South Asian region might also be adapted to short-day photoperiod given that flax accessions from this region differ from the common flax adapted to grow under the long-day seasons of most cooler regions [56, 58]. In a similar way, the higher rate of private haplotypes in the Abyssinian population may be attributable to adaptation of the crop to the equatorial region, with equinox effects, elevation-induced cool temperate climate and windstorms [31, 59]. Despite the high private haplotype concentration, the low gene diversity in the Abyssinian population may suggest effect of genetic drift [60]. The lack of private haplotypes in some populations such as Pop1, 2, 4, and 6, dominated by Eurasian and Canadian accessions, can be due to introduction of materials of different origins, selection of dual flax types (both fiber and oil) or through hybridization with wider germplasm in breeding programs [31, 61].

5.2 Adaptive loci and their linked genes

The outlier loci harbored genes of known roles for local adaptation. Most loci detected using the PCADAPT and F_{ST} methods are linked to genes that mediate responses to stresses. The flax genes *Lus10011967*, *Lus10022627* and *Lus10022628* that are predicted to encode UGT73Cs might be involved in *Fusarium oxysporum* wilt tolerance in flax [62–64] and other plant species [65–67]. Dmitriev *et al.* [68] demonstrated the up regulation of UGT73C3 in flax in response to *Fusarium* infection. By and large, UGTs play an important role in *Fusarium* wilt resistance [69] including in flax [62, 64]. *Fusarium* fungi are the most common diseases in many crop plants and can cause devastating losses. *Fusarium* wilt in flax is one of the most severe biotic stresses and may lead to a complete loss of flax production [70]. Hence, *Fusarium* diseases can be one of the natural selection forces that result in genetic divergence in many crops [71], including flax [72]. The consistent outlier SNP Chr1:10413007, that marks the locus harboring both UGT73C3 and UGT73C5, is likely an important genetic signature for local adaptation [73, 74]. The other locus marked by Chr1:5466653 included a gene predicted to encode PHYE. This gene was hypothesized to be involved in regulating responses to light quality and temperature [75]. As such, this locus may be an essential genetic signature of divergent adaptation to different eco-geographic regions.

The *ELF4-L4* gene at Chr7:10407650 locus plays a significant role in circadian clocking [76] and flowering time [77], which are both affected by day-length, and con-

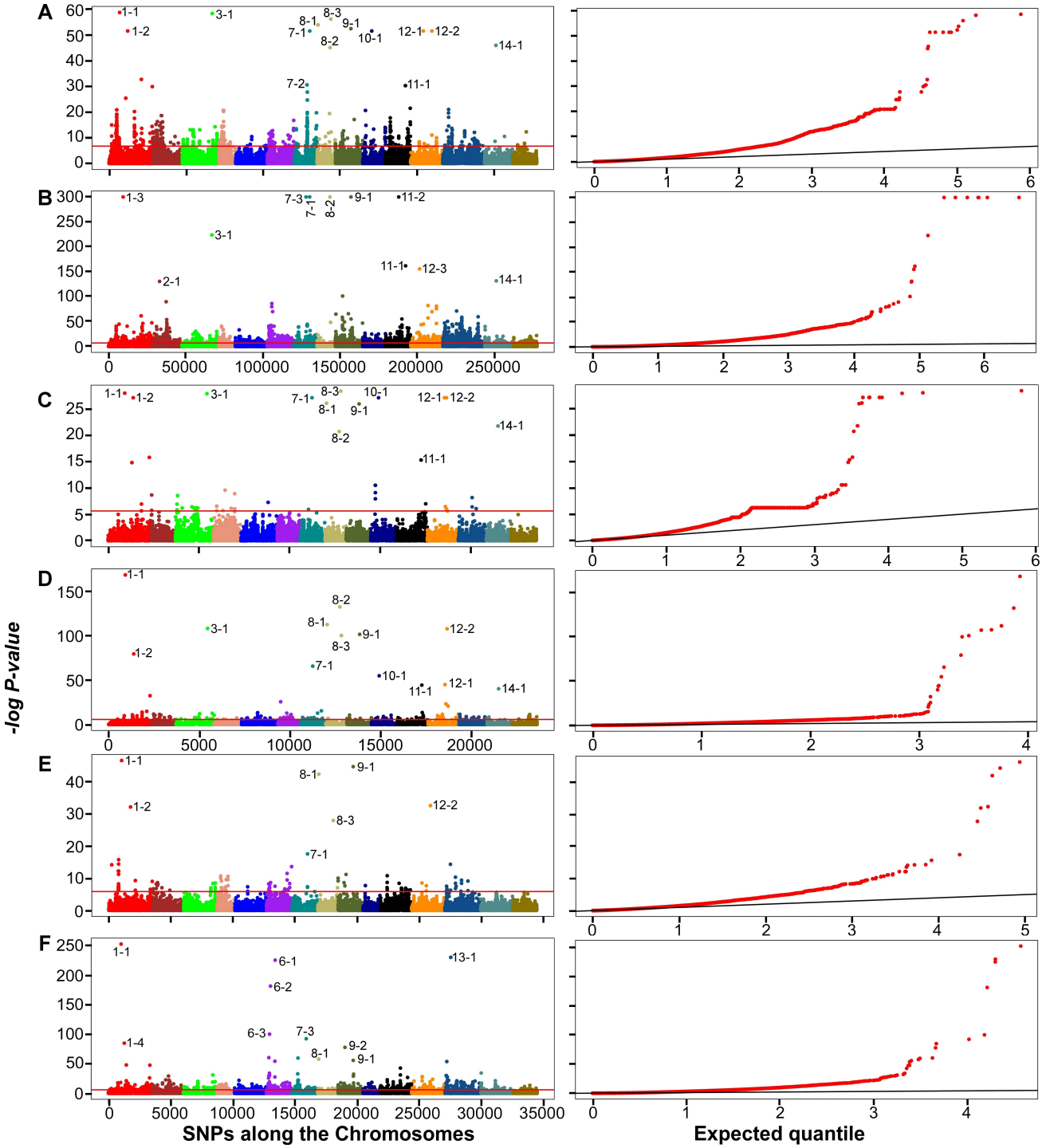

Fig. 4. Manhattan plots (left panels) and quantile-quantile (Q-Q) plots (right panels) of genome-wide associations using PCADAPT (A,C,E) and F_{ST} (B,D,F) of the 277K (A,B), 23K (C,D) and exon-34K (E,F) datasets. Outliers with the same number indicate the same SNP. The numbers before and after the hyphen indicate the chromosome number and the outlier SNP as listed in **Supplementary Table 3**, respectively.

sequently, latitude. The tandemly repeated *FBX92* genes at Chr8:2932993 locus can also be important signatures of adaptation of flax to varying environmental factors along its latitudinal gradient. FBX92 mediates responses to different abiotic stresses including light [78]. The FBX92 protein affects leaf sizes in *Arabidopsis* [79], which may contribute to adaptation to latitude-induced abiotic stresses such as temperature [80]. The multiple copies of the *UGT87A* gene at locus Chr1:3612570 and Chr1:3671655 might also play an important role in flowering time regulation and abiotic stress response. In *Arabidopsis*, *UGT87A2* mutants overexpressing the flowering repressor *FLOWERING LOCUS C* (*FLC*) had delayed flowering times regardless of the duration of the day-length [81]. *UGT87A2* is also in-

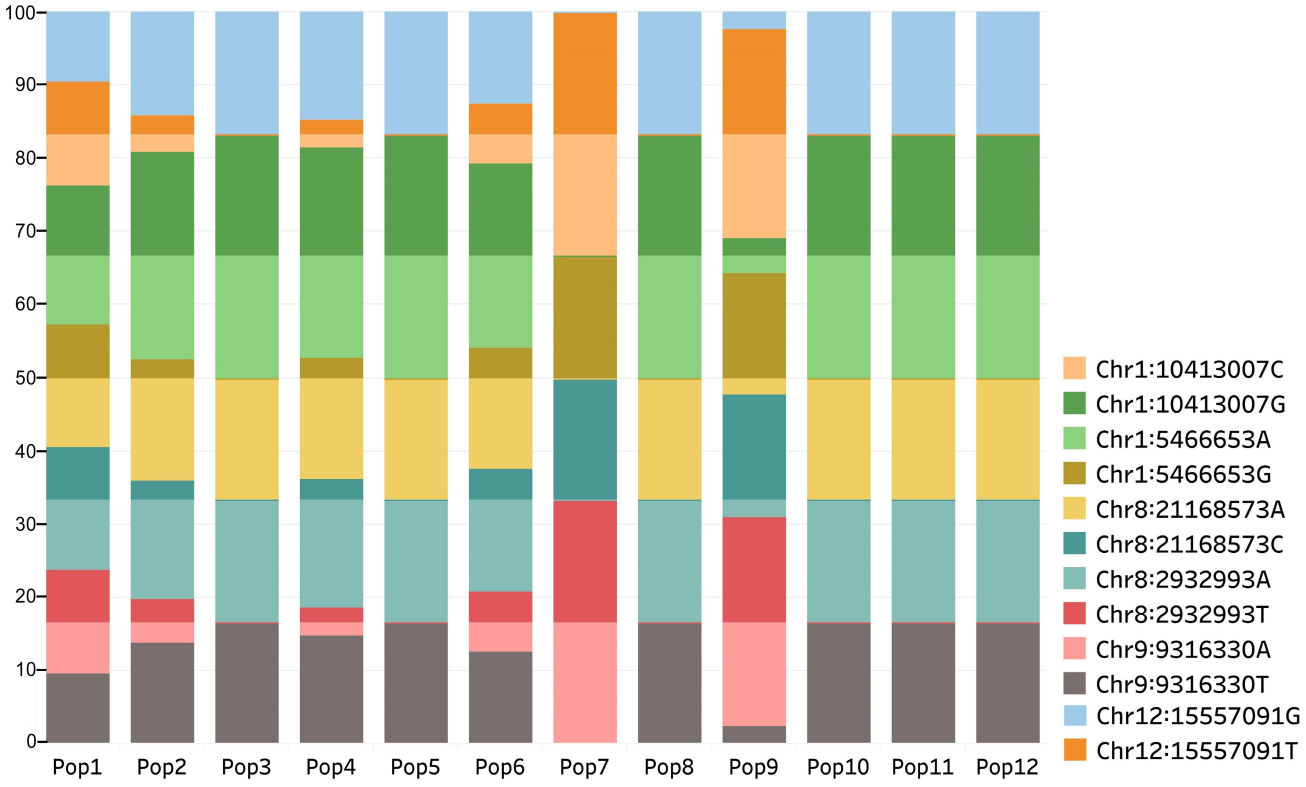

Fig. 5. Distribution of alleles at loci detected in the three SNP datasets among the populations. Populations from old flax growing region (Pop3, 5, 7–12) displayed a higher frequency of one of the alleles at each locus than the remaining populations from relatively newer flax growing regions.

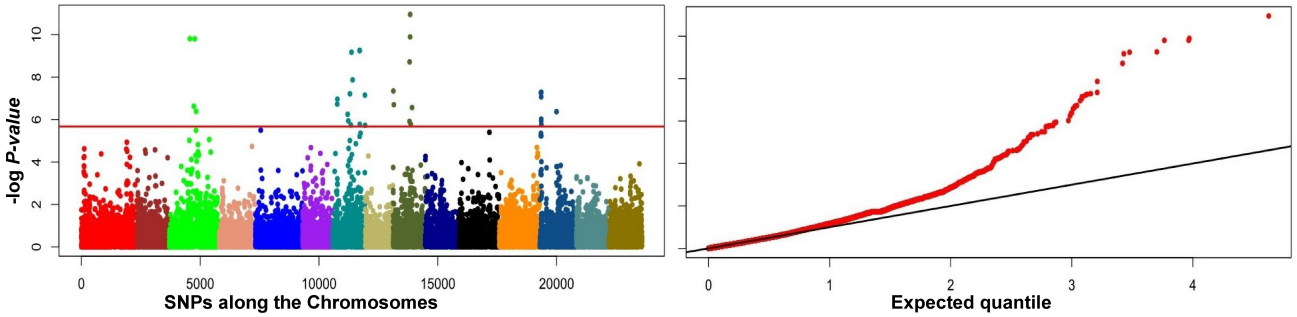

Fig. 6. Manhattan and quantile-quantile (Q-Q) plots of the first principal component (PC) for latitude, day-length and temperature of the cropping seasons using the genome-wide 23K SNP dataset showing outliers on four of the 15 flax chromosomes.

volved in plant adaptation to osmotic stresses, including drought and salinity, via regulation of multiple genes mediating responses to these stresses [82]. The Chr1:3612570-3671655 locus harbors multiple copies of the *UGT87A2* gene which may suggest its involvement in flax adaptation through regulation of stress responses [81, 82]. Other outlier loci also harbored genes of importance for local adaptation. For example, *Lus10042995* and *Lus10042996* were predicted to be orthologous to ethylene response factors *ERF106 (AT5G07580)* and *ERF105* (*AT5G51190*), respectively. The former is *DECREASE WAX BIOSYNTHESIS2 (DEWAX2)* that negatively regulates cuticular biosynthesis [83], and, as such, adversely impacts cuticular wax-related tolerance to abiotic stresses such as drought [84]. In contrast, the latter (ERF105) is a transcription factor involved in freezing tolerance [85]. Moisture availability and temperature are among the major factors that shaped the genetic structure of flax, and they constitute major determinants of the success of the crop in its current eco-geographic regions [31], which span the warm semi-arid Indian subcontinent [55, 86] to the relatively cool and humid temperate Eurasian regions [87].

Most of the outlier SNP loci that are associated with the PC1 of major environmental factors such as cropping season day-length, temperature and latitude, also harbored important genes with known roles in adaptive di-

vergence of plants along eco-geographic gradients. The *ELF4-L1* orthologous gene at Chr3:6169770 locus is predicted to regulate circadian rhythm and flowering time [88, 89], suggesting the adaptive signature of this locus in flax. Chr7:16799360, which marks the locus harboring the predicted *FLOWERING LOCUS C EXPRESSOR-LIKE 2* (*FLL2*) gene, is another crucial latitudinal gradient adaptation signature. *FLLs* have been reported for their role in the regulation of *FLOWERING LOCUS C* (*FLC*) and vernalisation [90, 91]. The *AT2G02540* orthologous genes *Lus10010693* and *Lus10010694* at the Chr13:10377859 locus also mediate flowering time via positive regulation of *FLC* [92, 93]. Flax is a "long-day" plant whose flowering time can be determined by both photoperiod and vernalisation [94]. The strong association between loci harboring genes that regulate flowering time and vernalisation suggests the importance of these loci for genetic divergence, which allowed flax to expand to vast geographic regions. Since GEA was performed based on representative putative locations, validation using materials of known and precise geographic coordinates and associated site factors such as climatic, edaphic and biodiversity records, remains warranted.

6. Conclusions

The F_{ST}- and PC-based genome scans and GEA have captured important genetic signatures of eco-geographic adaptations of flax to abiotic and biotic factors. Genome regions that harbor genes responding to light and important flax diseases such as *Fusarium* wilt have contributed in shaping the genetic structure and successes of the crop into its current diverse eco-geographic regions. However, given the putative nature of the genes discussed herein, further investigation is warranted to validate them. Precise original collection site information of each accession, including the geographic coordinates of the sampling sites, would strengthen the GEA analyses. The inclusion of additional local landraces would also be beneficial because it would increase the number of individuals in small populations and because landraces are good representatives of local adaptation. However, some limitations to the study need to be mentioned in view of the interpretation of the data and for consideration in future research avenues. Here, we use PC as a surrogate or proxy variable for latitude, temperature and day-length. This was justified because of the high correlation between them but, as such, their effects ended up being confounded. Plants' mechanisms of recognition of the photoperiod and temperature environmental cues can differ [95–97] but there is mounting evidence of complex interactions among them. Indeed, photoperiod sensitivity genes that may trigger flowering response can be intricately-linked to temperature shifts, such as in winter wheat, where they work in concert with vernalization (cold response) genes [98]. In *Arabidopsis*, photoperiod and temperature synchronize flowering [99]. Because of these complex interactions, it is difficult to tease apart the role(s) of temperature versus day-length, even in controlled experiments. Here, we could only infer the cues based on the known role(s) of the putative genes which was somewhat restricted to knowledge from the *Arabidopsis* orthologs. This was beyond the scope of the research here but may have implications for breeders attempting to introduce foreign germplasm into their breeding program because the foreign germplasm may be poorly adapted to the different photoperiod and temperature regime. In brief, genetic signatures captured using genome scan and GEA may help to select pre-breeding materials potentially adapted to specific growing niches without prior field performance trials. With the current low genotyping cost and freely available environmental data, this approach can readily provide predictions regarding the suitability of large flax collections to various environments.

7. Author contributions

DS and SC conceived the project. DS and SC designed and conducted the research. SC and FY generated the SNP. DS performed the data analysis. DS prepared the manuscript. SC and FY revised the manuscript.

8. Ethics approval and consent to participate

Not applicable.

9. Acknowledgment

We thank Nick Manseau, Madeleine Lévesque-Lemay and Tara Edwards for editorial inputs.

10. Funding

This Agri-Innovation Project (J-000279) was funded by the Flax Council of Canada and Agriculture and Agri-Food Canada.

11. Conflict of interest

The authors declare no conflict of interest.

12. References

[1] Barrett RD, Schluter D. Adaptation from standing genetic variation. Trends in Ecology and Evolution. 2008; 23: 38–44.

[2] Kelly M. Adaptation to climate change through genetic accommodation and assimilation of plastic phenotypes. Philosophical Transactions of the Royal Society B: Biological Sciences. 2019; 374: 20180176.

[3] Price TD, Qvarnström A, Irwin DE. The role of phenotypic plasticity in driving genetic evolution. Proceedings of the Royal

Society of London. Series B: Biological Sciences. 2003; 270: 1433–1440.
[4] Schluter D. Ecology and the origin of species. Trends in Ecology and Evolution. 2001; 16: 372–380.
[5] Nosil P, Vines TH, Funk DJ. Perspective: Reproductive isolation caused by natural selection against immigrants from divergent habitats. Evolution. 2005; 59: 705–719.
[6] Nakamichi N. Adaptation to the local environment by modifications of the photoperiod response in crops. Plant and Cell Physiology. 2015; 56: 594–604.
[7] Abbo S, Pinhasi van-Oss R, Gopher A, Saranga Y, Ofner I, Peleg Z. Plant domestication versus crop evolution: a conceptual framework for cereals and grain legumes. Trends in Plant Science. 2014; 19: 351–360.
[8] Storz JF. Using genome scans of DNA polymorphism to infer adaptive population divergence. Molecular Ecology. 2005; 14: 671–688.
[9] Wang J, Ding J, Tan B, Robinson KM, Michelson IH, Johansson A, *et al*. A major locus controls local adaptation and adaptive life history variation in a perennial plant. Genome Biology. 2018; 19: 72.
[10] Gallego García N, Forero Medina G, Vargas Ramírez M, Caballero S, Shaffer HB. Landscape genomic signatures indicate reduced gene flow and forest-associated adaptive divergence in an endangered neotropical turtle. Molecular Ecology. 2019; 28: 2757–2771.
[11] Capblancq T, Luu K, Blum MGB, Bazin E. Evaluation of redundancy analysis to identify signatures of local adaptation. Molecular Ecology Resources. 2018; 18: 1223–1233.
[12] Feng X, Jiang G, Fan Z. Identification of outliers in a genomic scan for selection along environmental gradients in the bamboo locust, Ceracris kiangsu. Scientific Reports. 2015; 5: 13758.
[13] Bay RA, Rose N, Barrett R, Bernatchez L, Ghalambor CK, Lasky JR, *et al*. Predicting Responses to Contemporary Environmental Change Using Evolutionary Response Architectures. The American Naturalist. 2017; 189: 463–473.
[14] Hoban S, Kelley JL, Lotterhos KE, Antolin MF, Bradburd G, Lowry DB, *et al*. Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. The American Naturalist. 2016; 188: 379–397.
[15] Yoder JB, Stanton-Geddes J, Zhou P, Briskine R, Young ND, Tiffin P. Genomic signature of adaptation to climate in *Medicago truncatula*. Genetics. 2014; 196: 1263–1275.
[16] Lasky JR, Upadhyaya HD, Ramu P, Deshpande S, Hash CT, Bonnette J, *et al*. Genome-environment associations in sorghum landraces predict adaptive traits. Science Advances. 2015; 1: e1400218.
[17] Abebe TD, Naz AA, Léon J. Landscape genomics reveal signatures of local adaptation in barley (*Hordeum vulgare* L.). Frontiers in Plant Science. 2015; 6: 813.
[18] Navarro JAR, Wilcox M, Burgueño J, Romay C, Swarts K, Trachsel S, *et al*. Corrigendum: A study of allelic diversity underlying flowering-time adaptation in maize landraces. Nature Genetics. 2017; 49: 970.
[19] Bekele WA, Wight CP, Chao S, Howarth CJ, Tinker NA. Haplotype-based genotyping-by-sequencing in oat genome research. Plant Biotechnology Journal. 2018; 16: 1452–1463.
[20] Berny Mier y Teran JC, Konzen ER, Medina V, Palkovic A, Ariani A, Tsai SM, *et al*. Root and shoot variation in relation to potential intermittent drought adaptation of Mesoamerican wild common bean (*Phaseolus vulgaris* L.). Annals of Botany. 2018; 124: 917–932.
[21] He F, Pasam R, Shi F, Kant S, Keeble-Gagnere G, Kay P, *et al*. Exome sequencing highlights the role of wild-relative introgression in shaping the adaptive landscape of the wheat genome. Nature Genetics. 2019; 51: 896–904.
[22] Mousavi-Derazmahalleh M, Bayer PE, Nevado B, Hurgobin B, Filatov D, Kilian A, *et al*. Exploring the genetic and adaptive diversity of a pan-Mediterranean crop wild relative: narrow-leafed lupin. Theoretical and Applied Genetics. 2018; 131: 887–901.
[23] Brunazzi A, Scaglione D, Talini RF, Miculan M, Magni F, Poland J, *et al*. Molecular diversity and landscape genomics of the crop wild relative *Triticum urartu* across the Fertile Crescent. The Plant Journal. 2018; 94: 670–684.
[24] Frachon L, Bartoli C, Carrère S, Bouchez O, Chaubet A, Gautier M, *et al*. A Genomic map of climate adaptation in *Arabidopsis thaliana* at a micro-geographic scale. Frontiers in Plant Science. 2018; 9: 967.
[25] Zohary D, Hopf M, Weiss E. Domestication of Plants in the Old World: The origin and spread of domesticated plants in Southwest Asia, Europe, and the Mediterranean Basin. 4th edn. Oxford University Press: Oxford. 2012.
[26] Van Zeist W, Bakker-Heeres JAH. Evidence for linseed cultivation before 6000 BC. Journal of Archaeological Science. 1975; 2: 215–219.
[27] Valamoti SM. Flax in Neolithic and Bronze Age Greece: archaeobotanical evidence. Vegetation History and Archaeobotany. 2011; 20: 549–560.
[28] Herbig C, Maier U. Flax for oil or fibre? Morphometric analysis of flax seeds and new aspects of flax cultivation in Late Neolithic wetland settlements in southwest Germany. Vegetation History and Archaeobotany. 2011; 20: 527–533.
[29] Wang Y, Jankauskiene Z, Qiu C, Gruzdeviene E, Long S, Alexopoulou E, *et al*. Fiber flax breeding in China and Europe. Journal of Natural Fibers. 2018; 15: 309–324.
[30] Vavilov NI. The origin, variation, immunity and breeding of cultivated plants (pp. 387). The Chronica Botanica Co: New York. 1951.
[31] Sertse D, You FM, Ravichandran S, Cloutier S. The genetic structure of flax illustrates environmental and anthropogenic selections that gave rise to its eco-geographical adaptation. Molecular Phylogenetics and Evolution. 2019; 137: 22–32.
[32] Gutaker RM, Zaidem M, Fu Y, Diederichsen A, Smith O, Ware R, *et al*. Flax latitudinal adaptation at LuTFL1 altered architecture and promoted fiber production. Scientific Reports. 2019; 9: 976.
[33] You FM, Xiao J, Li P, Yao Z, Jia G, He L, *et al*. Chromosome-scale pseudomolecules refined by optical, physical and genetic maps in flax. The Plant Journal. 2018; 95: 371–384.
[34] Money D, Gardner K, Migicovsky Z, Schwaninger H, Zhong G, Myles S. LinkImpute: fast and accurate genotype imputation for nonmodel organisms. G3: Genes, Genomes, Genetics. 2015; 5: 2383–2390.
[35] Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics. 2007; 23: 2633–2635.
[36] Luu K, Bazin E, Blum MGB. Pcadapt: an R package to perform genome scans for selection based on principal component analysis. Molecular Ecology Resources. 2017; 17: 67–77.
[37] Cattell RB. The Scree Test for the Number of Factors. Multivariate Behavioral Research. 1966; 1: 245–276.
[38] Alexander DH, Lange K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics. 2011; 12: 246.
[39] Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Research. 2009; 19: 1655–1664.
[40] Wickham H. ggplot2: elegant graphics for data analysis. 2nd edn. Springer: New York. 2016.
[41] Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Research. 2016; 44: W242–W245.
[42] Excoffier L, Lischer HEL. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources. 2010; 10: 564–567.

[43] Frichot E, François O. LEA: an R package for landscape and ecological association studies. Methods in Ecology and Evolution. 2015; 6: 925–929.
[44] Dereeper A, Homa F, Andres G, Sempere G, Sarah G, Hueber Y, *et al*. SNiPlay3: a web-based application for exploration and large scale analyses of genomic variations. Nucleic Acids Research. 2015; 43: W295–W300.
[45] Rousset F. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Molecular Ecology Resources. 2008; 8: 103–106.
[46] Rousset F, Lopez J, Belkhir K. Package 'genepop'. 2020; 1: 7.
[47] Hijmans R, Guarino L, Mathur P. DIVA-GIS Version 7.5, A geographic information system for the analysis of species distribution data. 2012. Available at: http://www.diva-gis.org (Accessed: 30 November 2021).
[48] Caye K, Jumentier B, Lepeule J, François O. LFMM 2: Fast and Accurate Inference of Gene-Environment Associations in Genome-Wide Studies. Molecular Biology and Evolution. 2019; 36: 852–860.
[49] Kim SA, Brossard M, Roshandel D, Paterson AD, Bull SB, Yoo YJ. Gpart: human genome partitioning and visualization of high-density SNP data by identifying haplotype blocks. Bioinformatics. 2019; 35: 4419–4421.
[50] Howden SM, Soussana J, Tubiello FN, Chhetri N, Dunlop M, Meinke H. Adapting agriculture to climate change. Proceedings of the National Academy of Sciences of the United States of America. 2007; 104: 19691–19696.
[51] Savolainen O, Lascoux M, Merilä J. Ecological genomics of local adaptation. Nature Reviews. Genetics. 2013; 14: 807–820.
[52] Eo SH, Wares JP, Carroll JP. Population divergence in plant species reflects latitudinal biodiversity gradients. Biology Letters. 2008; 4: 382–384.
[53] Parmesan C. Influences of species, latitudes and methodologies on estimates of phenological response to global warming. Global Change Biology. 2007; 13: 1860–1872.
[54] Soto-Cerda BJ, Diederichsen A, Ragupathy R, Cloutier S. Genetic characterization of a core collection of flax (*Linum usitatissimum* L.) suitable for association mapping studies and evidence of divergent selection between fiber and linseed types. BMC Plant Biology. 2013; 13: 78.
[55] Rajwade AV, Arora RS, Kadoo NY, Harsulkar AM, Ghorpade PB, Gupta VS. Relatedness of Indian flax genotypes (*Linum usitatissimum* L.): an inter-simple sequence repeat (ISSR) primer assay. Molecular Biotechnology. 2010; 45: 161–170.
[56] Hoque A, Fiedler JD, Rahman M. Genetic diversity analysis of a flax (*Linum usitatissimum* L.) global collection. BMC Genomics. 2020; 21: 557.
[57] Casa R, Russell G, Lo Cascio B, Rossini F. Environmental effects on linseed (*Linum usitatissimum* L.) yield and growth of flax at different stand densities. European Journal of Agronomy. 1999; 11: 267–278.
[58] Domantovich AV, Koshkin VA, Brutch NB, Matvienko II. Investigation of photoperiod sensitivity of *Linum usitatissimum* L. Lines and effect of short-day conditions on their economically valuable traits. Russian Agricultural Sciences. 2012; 38: 173–177.
[59] Vaisey-Genser M, Morris DH. History of the cultivation and uses of flaxseed. In Muir A, Westcott N (eds.) Flax: the genus *Linum* (pp.1–21). Taylor & Francis: London. 2003.
[60] LACY RC. Loss of Genetic Diversity from Managed Populations: Interacting Effects of Drift, Mutation, Immigration, Selection, and Population Subdivision. Conservation Biology. 1987; 1: 143–158.
[61] You FM, Jia G, Xiao J, Duguid SD, Rashid KY, Booker HM, *et al.* Genetic variability of 27 traits in a core collection of flax (*Linum usitatissimum* L.). Frontiers in Plant Science. 2017; 8: 1636.
[62] Galindo-González L, Deyholos MK. RNA-seq Transcriptome Response of Flax (*Linum usitatissimum* L.) to the Pathogenic *Fungus Fusarium oxysporum* f. sp. *lini*. Frontiers in Plant Science. 2016; 7: 1766.
[63] Lorenc-Kukuła K, Zuk M, Kulma A, Czemplik M, Kostyn K, Skala J, *et al*. Engineering flax with the GT family 1 *Solanum sogarandinum* glycosyltransferase SsGT1 confers increased resistance to Fusarium infection. Journal of Agricultural and Food Chemistry. 2009; 57: 6698–6705.
[64] Kostyn K, Czemplik M, Kulma A, Bortniczuk M, Skała J, Szopa J. Genes of phenylpropanoid pathway are activated in early response to Fusarium attack in flax plants. Plant Science. 2012; 190: 103–115.
[65] Torres-Trenas A, Cañizares MC, García-Pedrajas MD, Pérez-Artés E. Molecular and biological characterization of the first hypovirus identified in *Fusarium oxysporum*. Frontiers in Microbiology. 2020; 10: 3131.
[66] Zhang Q, Gao M, Wu L, Wu H, Chen Y, Wang Y. Expression network of transcription factors in resistant and susceptible tung trees responding to Fusarium wilt disease. Industrial Crops and Products. 2018; 122: 716–725.
[67] Xing M, Lv H, Ma J, Xu D, Li H, Yang L, *et al*. Transcriptome profiling of resistance to *Fusarium oxysporum* f. sp. *conglutinans* in cabbage (*Brassica oleracea*) roots. PLoS ONE. 2016; 11: e0148048.
[68] Dmitriev AA, Krasnov GS, Rozhmina TA, Novakovskiy RO, Snezhkina AV, Fedorova MS, *et al*. Differential gene expression in response to *Fusarium oxysporum* infection in resistant and susceptible genotypes of flax (*Linum usitatissimum* L.). BMC Plant Biology. 2017; 17: 253.
[69] Abbasi S, Safaie N, Sadeghi A, Shamsbakhsh M. *Streptomyces* strains induce resistance to *Fusarium oxysporum* f. sp. *lycopersici* race 3 in tomato through different molecular mechanisms. Frontiers in Microbiology. 2019; 10: 1505.
[70] Rashid K Y. Principal diseases of flax. In: Muir A, Weir BS (eds) Flax: the genus *Linum*. pp 104–135. Taylor & Francis: London. 2003.
[71] Kelly AC, Ward TJ. Population genomics of *Fusarium graminearum* reveals signatures of divergent evolution within a major cereal pathogen. PLoS ONE. 2018; 13: e0194616.
[72] Diederichsen A, Fu Y-B. Flax genetic diversity as the raw material for future success. Genus. 2008; 32: 33.
[73] Heimes C, Agerbirk N, Sørensen H, van Mölken T, Hauser TP. Ecotypic differentiation of two sympatric chemotypes of *Barbarea vulgaris* (*Brassicaceae*) with different biotic resistances. Plant Ecology. 2016; 217: 1055–1068.
[74] Augustin JM, Drok S, Shinoda T, Sanmiya K, Nielsen JK, Khakimov B, *et al*. UDP-glycosyltransferases from the UGT73C subfamily in *Barbarea vulgaris* catalyze sapogenin 3-O-glucosylation in saponin-mediated insect resistance. Plant Physiology. 2012; 160: 1881–1895.
[75] Dechaine JM, Gardner G, Weinig C. Phytochromes differentially regulate seed germination responses to light quality and temperature cues during seed maturation. Plant, Cell and Environment. 2009; 32: 1297–1309.
[76] Boxall SF, Foster JM, Bohnert HJ, Cushman JC, Nimmo HG, Hartwell J. Conservation and divergence of circadian clock operation in a stress-inducible crassulacean acid metabolism species reveals clock compensation against stress. Plant Physiology. 2005; 137: 969–982.
[77] He T, Hill CB, Angessa TT, Zhang X, Chen K, Moody D, *et al*. Gene-set association and epistatic analyses reveal complex gene interaction networks affecting flowering time in a worldwide barley collection. Journal of Experimental Botany. 2019; 70: 5603–5616.
[78] Jain M, Nijhawan A, Arora R, Agarwal P, Ray S, Sharma P, *et al*. F-Box Proteins in Rice. Genome-wide analysis, classification, temporal and spatial gene expression during panicle and seed development, and regulation by light and abiotic stress. Plant Physiology. 2007; 143: 1467–1483.
[79] Baute J, Polyn S, De Block J, Blomme J, Van Lijsebettens M,

Inzé D. F-Box Protein FBX92 Affects leaf size in *Arabidopsis thaliana*. Plant and Cell Physiology. 2017; 58: 962–975.
[80] Wright IJ, Dong N, Maire V, Prentice IC, Westoby M, Díaz S, *et al*. Global climatic drivers of leaf size. Science. 2017; 357: 917–921.
[81] Wang B, Jin S, Hu H, Sun Y, Wang Y, Han P, *et al*. UGT87a2, an *Arabidopsis* glycosyltransferase, regulates flowering time via FLOWERING LOCUS C. The New Phytologist. 2012; 194: 666–675.
[82] Li P, Li Y, Wang B, Yu H, Li Q, Hou B. The *Arabidopsis* UGT87a2, a stress-inducible family 1 glycosyltransferase, is involved in the plant adaptation to abiotic stresses. Physiologia Plantarum. 2017; 159: 416–432.
[83] Kim H, Go YS, Suh MC. DEWAX2 transcription gactor negatively tegulates cuticular eax biosynthesis in *Arabidopsis* leaves. Plant and Cell Physiology. 2018; 59: 966–977.
[84] Aharoni A, Dixit S, Jetter R, Thoenes E, van Arkel G, Pereira A. The SHINE clade of AP2 domain transcription factors activates wax biosynthesis, alters cuticle properties, and confers drought tolerance when overexpressed in *Arabidopsis*. The Plant Cell. 2004; 16: 2463–2480.
[85] Bolt S, Zuther E, Zintl S, Hincha DK, Schmülling T. ERF105 is a transcription factor gene of *Arabidopsis thaliana* required for freezing tolerance and cold acclimation. Plant, Cell and Environment. 2017; 40: 108–120.
[86] Singh DP, Singh K, Sharma J. Effect of salinity on flowering ability in linseed under experimental conditions. Journal of Advanced Laboratory Research in Biology. 2016; 7: 94–98.
[87] Hall LM, Booker H, Siloto RMP, Jhala AJ, Weselake RJ. Flax (*Linum usitatissimum* L.) Industrial Oil Crops. 2016; 135: 157–194.
[88] Khanna R, Kikis EA, Quail PH. EARLY FLOWERING 4 functions in phytochrome B-regulated seedling de-etiolation. Plant Physiology. 2003; 133: 1530–1538.
[89] Marcolino-Gomes J, Nakayama TJ, Molinari HBC, Basso MF, Henning LMM, Fuganti-Pagliarini R, *et al*. Functional characterization of a putative Glycine max ELF4 in transgenic *Arabidopsis* and its role during flowering control. Frontiers in Plant Science. 2017; 8: 618.
[90] Choi K, Kim J, Hwang H, Kim S, Park C, Kim SY, *et al*. The FRIGIDA complex activates transcription of FLC, a strong flowering repressor in *Arabidopsis*, by recruiting chromatin modification factors. The Plant Cell. 2011; 23: 289–303.
[91] Lee J, Amasino RM. Two FLX family members are non-redundantly required to establish the vernalization requirement in *Arabidopsis*. Nature Communications. 2014; 4: 2186.
[92] Li Y, Li C, Bradbury PJ, Liu X, Lu F, Romay CM, *et al*. Identification of genetic variants associated with maize flowering time using an extremely large multi-genetic background population. The Plant Journal. 2016; 86: 391–402.
[93] Proveniers M, Rutjens B, Brand M, Smeekens S. The *Arabidopsis* TALE homeobox gene ATH1 controls floral competency through positive regulation of FLC. The Plant Journal. 2007; 52: 899–913.
[94] Brutch N, Koshkin V, Matvienko I, Pookhovinova E, Tavares de Sousa M, Domantovich A. Influence of low temperatures and short photoperiod on the time of flowering in flax. In Pookhovinova E, Tavares de Sousa M, Domantovich A (eds.) Fiber foundations – transportation, clothing and shelter in the bioeconomy. Proceedings of the International Conference on Flax and Other Bast Plants (pp. 81–91). SaskFlax: Saskatoon. 2008.
[95] Wang H, Wang H, Ge Q, Dai J. The interactive effects of chilling, photoperiod, and forcing temperature on flowering phenology of temperate woody plants. Frontiers in Plant Science. 2020; 11: 443.
[96] Ream TS, Woods DP, Schwartz CJ, Sanabria CP, Mahoy JA, Walters EM, *et al*. Interaction of photoperiod and vernalization determines flowering time of *Brachypodium distachyon*. Plant Physiology. 2014; 164: 694–709.
[97] Singh RK, Svystun T, AlDahmash B, Jönsson AM, Bhalerao RP. Photoperiod- and temperature-mediated control of phenology in trees - a molecular perspective. The New Phytologist. 2017; 213: 511–524.
[98] Distelfeld A, Li C, Dubcovsky J. Regulation of flowering in temperate cereals. Current Opinion in Plant Biology. 2009; 12: 178–184.
[99] Song YH, Ito S, Imaizumi T. Flowering time regulation: photoperiod-and temperature-sensing in leaves. Trends in Plant Science. 2013; 18: 575–583.

Supplementary material: Supplementary material associated with this article can be found, in the online version, at https://www.imrpress.com/journal/FBL/26/12/10.52586/5049.

Keywords: GWAS; F_{ST}; Principal component; Adaptive loci; Genetic signatures; Candidate genes

Send correspondence to: Sylvie Cloutier, Department of Biology, University of Ottawa, Ottawa, ON K1N 6N5, Canada, Department of Agriculture and Agri-Food, Ottawa Research and Development Center, Ottawa, ON K1A 0C6, Canada, E-mail: Sylviej.Cloutier@agr.gc.ca