

# The potential of pale flax as a source of useful genetic variation for cultivated flax revealed through molecular diversity and association analyses

Braulio J. Soto-Cerda · Axel Diederichsen ·  
Scott Duguid · Helen Booker · Gordon Rowland ·  
Sylvie Cloutier

Received: 25 March 2014 / Accepted: 29 July 2014 / Published online: 12 August 2014  
© The Author(s) 2014. This article is published with open access at Springerlink.com

**Abstract** Pale flax (*Linum bienne* Mill.) is the wild progenitor of cultivated flax (*Linum usitatissimum* L.) and represents the primary gene pool to broaden its genetic base. Here, a collection of 125 pale flax accessions and the Canadian flax core collection of 407 accessions were genotyped using 112 genome-wide simple sequence repeat markers and phenotyped for nine traits with the aim of conducting population structure, molecular diversity and association mapping analyses. The combined population structure analysis identified two well-supported major groups

corresponding to pale and cultivated flax. The *L. usitatissimum* convar. *crepitans* accessions most closely resembled its wild progenitor, both having dehiscent capsules. The unbiased Nei's genetic distance (0.65) confirmed the strong genetic differentiation between cultivated and pale flax. Similar levels of genetic diversity were observed in both species, albeit 430 (48 %) of pale flax alleles were unique, in agreement with their high genetic differentiation. Significant associations were identified for seven and four traits in pale and cultivated flax, respectively. Favorable alleles with potentially positive effect to improve yield through yield components were identified in pale flax. The allelic frequencies of markers associated with domestication-related traits such as

**Electronic supplementary material** The online version of this article (doi:10.1007/s11032-014-0165-5) contains supplementary material, which is available to authorized users.

B. J. Soto-Cerda · S. Cloutier  
Department of Plant Science, University of Manitoba, 66  
Dafoe Road, Winnipeg, MB R3T 2N2, Canada

B. J. Soto-Cerda · S. Cloutier (✉)  
Cereal Research Centre, Agriculture and Agri-Food  
Canada, 195 Dafoe Rd, Winnipeg, MB R3T 2M9, Canada  
e-mail: Sylvie.J.Cloutier@agr.gc.ca

*Present Address:*

B. J. Soto-Cerda  
Agriaquaculture Nutritional Genomic Center, CGNA,  
Genomics and Bioinformatics Unit, 4791057 Temuco,  
Chile

A. Diederichsen  
Plant Gene Resources of Canada, Agriculture and Agri-  
Food Canada, 107 Science Place, Saskatoon,  
SK S7N 0X2, Canada

S. Duguid  
Morden Research Station, Agriculture and Agri-Food  
Canada, 101 Route 100, Unit 100, Morden,  
MB R6M 1Y5, Canada

H. Booker · G. Rowland  
Department of Plant Sciences, College of Agriculture and  
Bioresources, University of Saskatchewan, 51 Campus  
Drive, Saskatoon, SK S7N 5A8, Canada

*Present Address:*

S. Cloutier  
Eastern Cereal and Oilseed Research Centre, Agriculture  
and Agri-Food Canada, 960 Carling Avenue,  
Ottawa ON K1A 0C6, Canada

capsular dehiscence indicated directional selection with the most common alleles in pale flax being absent or rare in cultivated flax and vice versa. Our results demonstrated that pale flax is a potential source of novel variation to improve multiple traits in cultivated flax and that association mapping is a suitable approach to screening pale flax germplasm to identify favorable quantitative trait locus alleles.

**Keywords** Pale flax · Cultivated flax · Genetic diversity · Association mapping · Wild ancestor · Exotic alleles

## Introduction

Plant breeders select favorable alleles to introduce new variation (Sim et al. 2011). Conversely, as the genetic base of breeding populations narrows through selection and fixation of specific alleles, breeding progress is hampered (Sim et al. 2011). Shrinking of the genetic diversity consequently reduces options to ensure diverse nutrition, to enhance food production and to face climate change. Crop wild relatives hosted in gene bank collections have received some attention because they harbor untapped genetic variation available for domesticated crops (McCouch et al. 2013; Xiao et al. 1998).

Pale flax (*Linum bienne* Mill.) is the wild progenitor of cultivated flax (*Linum usitatissimum* L.) as supported by morphological, cytological and molecular characterizations (Diederichsen and Hammer 1995; Fu and Allaby 2010; Fu and Peterson 2010; Gill and Yermanos 1967; Tammes 1928; Uysal et al. 2010, 2012). Pale flax is a winter annual or perennial plant with narrow leaves and dehiscent capsules and, usually displays large variation in the vegetative plant parts and variable growth habit (Diederichsen and Hammer 1995; Uysal et al. 2010, 2012). On the other hand, cultivated flax is mostly a spring annual plant that has variable seed dormancy, grows fast, flowers early, with almost indehiscent capsules and large seeds (Diederichsen and Richards 2003). Because both species share the same chromosome number ( $n = 15$ ) and produce fully fertile offspring, pale flax represents cultivated flax's primary gene pool (Diederichsen 2007), which can be accessed to broaden its generally narrow genetic bases (Cloutier et al. 2009; Diederichsen and Fu 2006; Fu et al. 2002, 2003; Smýkal et al. 2011; Soto-Cerda et al. 2012). However, there are no

studies to date reporting the deployment of pale flax into cultivated flax breeding (Diederichsen 2007). In total, 279 pale flax accessions have been stored in gene banks around the world, the largest collection being at the German Institute of Plant Genetics and Crop Plant Research (IPK) (Diederichsen 2007). Recent efforts have expanded the pale flax germplasm to add 34 accessions collected in Turkey (Uysal et al. 2010).

Molecular markers have facilitated the study and comparison of the genetic diversity between cultivated crops and their wild relatives. Molecular diversity studies conducted between soybean (*Glycine max*) and *Glycine soja* (Lam et al. 2010), barley (*Hordeum vulgare*) and *Hordeum vulgare* ssp. *spontaneum* (Russell et al. 2011), maize (*Zea mays*) and *Zea mays* ssp. *parviglumis* (Hufford et al. 2012), and rice (*Oryza sativa*) and *Oryza rufipogon* and *Oriza nivara* (Xu et al. 2013) have provided insights into the amount, distribution and erosion of the genetic variation present in cultivated crops relative to their wild ancestors. Nevertheless, these studies have mostly revealed patterns of variation at neutral loci, including allelic diversity, population structure and shared ancestry. Combined analysis of neutral and fitness-related loci between cultivated crops and wild relatives can assist in the identification and deployment of novel or eroded variation. Tanksley and Nelson (1996) and Zamir (2001) proposed the advanced backcross quantitative trait loci (AB-QTLs) and the introgression lines (ILs) methods, respectively, for the simultaneous detection of QTL and the development of improved varieties (Xiao et al. 1998). Although these approaches have proven successful for several crops (Liu et al. 2006; Schmalenbach et al. 2009; Tanksley 1993; Xiao et al. 1998), they screen a limited number of alleles per locus because they are based on biparental populations. Alternatively, association mapping (AM) utilizes a population of individuals representing a higher level of allelic diversity that improves the probability of QTL detection and the mapping resolution (Flint-Garcia et al. 2003). AM has identified favorable exotic alleles in *Solanum lycopersicum* var. *cerasiforme* (Ranc et al. 2012), *Hordeum vulgare* ssp. *spontaneum* (Matus et al. 2003) and *Zea mays* ssp. *parviglumis* (Weber et al. 2009).

In the present study, we conducted population structure, molecular diversity and AM analyses for 125 pale flax accessions and 407 cultivated flax accessions from the Canadian flax core collection

(Diederichsen et al. 2013). The aims of our study were: (1) to understand the population structure and genetic relationships between pale and cultivated flax (2) to quantify their genetic diversity and (3) to identify QTL alleles in pale flax that could be used to broaden the allelic richness of cultivated flax for nine agronomic, phenological and morphological traits.

## Materials and methods

### Plant material, genotyping and field trials

One hundred and twenty-five pale flax accessions were provided by the Institute of Plant Genetics and Crop Plant Research (IPK), Plant Gene Resources of Canada (PGRC) and the US Department of Agriculture (USDA) (Online Resource 1). Among the pale flax accessions, those recently collected from Turkey were not included (Uysal et al. 2010). The Canadian flax core collection contains 381 accessions selected by Diederichsen et al. (2013), and twenty-six accessions of relevance to recent Canadian flax breeding programs. Both collections were genotyped with 112 simple sequence repeat (SSR) markers distributed across the 15 linkage groups of flax (Cloutier et al. 2012) using a single plant for each accession. The 112 SSRs were selected based on their polymorphic information content (PIC). The amplification products were resolved on an ABI 3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). Output files were analyzed by GeneScan (Applied Biosystems) and subsequently imported into Genographer. Fragment sizes were estimated using GeneScan ROX-500 (Applied Biosystems) and MapMarker<sup>®</sup> 1000 (BioVentures Inc., Murfreesboro, TN) internal size standards. The genotype of each locus was encoded based on its allele size in bp or as a null allele for dominant markers.

For pale flax, sixty-five accessions (Diederichsen and Hammer 1995) were phenotypically evaluated between 1983 and 1992. Every accession was cultivated for three years with two observations of ten plants every year (Diederichsen and Hammer 1995). For cultivated flax, 390 accessions from the Canadian flax core collection were evaluated during 4 years (2009, 2010, 2011 and 2012) at the Morden Research Station, Morden, Manitoba and at the Kernen Crop Research Farm located near Saskatoon, Saskatchewan, Canada. A type 2 modified augmented design

(MAD) (Lin and Poushinsky 1985) was used for the field experiments from which phenotypic data were collected for nine traits. Main plots were arranged in grids of ten rows and ten columns. Each main plot was divided into five paralleled subplots (2 m × 2 m with 20 cm between rows) with a plot control (CDC Bethune) located in the center. Additional subplot controls (Hanley and Macbeth) were assigned to five randomly selected main plots.

### Phenotypic data

For pale flax, phenotypic data for thousand seed weight (TSW), seeds per boll (SPB), plant height (PH), capsular dehiscence (DEH), start of flowering (FL5%), petal color (PC), petal overlap (PO), flower shape (FS) and seed color (SC) were as described in Diederichsen and Hammer (1995). For cultivated flax, TSW and SPB were obtained by harvesting two half meter sections from the central part of each subplot. PH (cm) was recorded at maturity using the average of ten plants located in the center of the subplots. DEH was scored as follows: 1 = dehiscent, 2 = medium opened, 5 = slightly opened, 7 = weak, 9 = indehiscent. FL5 % was measured as the number of days between sowing and when 5 % of the flowers had opened. PC was scored using the following scale: 1 = white, 2 = light blue, 3 = blue, 4 = dark blue, 5 = pink, 6 = violet. PO was measured as follows: 1 = petal overlap more than 50 % of length, 2 = petal overlap less than 50 % of length. FS was recorded using the following scale: 1 = tube, 2 = funnel, 3 = bowl. SC was scored as follows: 1 = yellow, 2 = yellowish brown, 3 = olive, 4 = light brown, 5 = medium brown, 6 = dark brown.

### Statistical analysis

Pale flax data were statistically summarized (Diederichsen and Hammer 1995) and the mean for each trait was used for posterior analyses. For cultivated flax, adjusted data were obtained for each trait as previously described based on the MAD (You et al. 2013). Normal distribution of the mean adjusted traits was tested using the Shapiro–Wilk test (Shapiro and Wilk 1965). Traits with significant deviation from a normal distribution were log-transformed in pale and cultivated flax prior to AM analysis and that included PC, PO, FS and SC. For cultivated flax, an analysis of

variance for the different sources of variation including year, location, genotype and their interactions was reported (Soto-Cerda et al. 2014).

### Population structure and phylogenetic analyses

To investigate the patterns of population structure, Bayesian-based and principal coordinate (PCoA) analyses based on the 112 SSRs including minor allele frequency (MAF)  $<0.05$  were conducted. Population structure analysis was first carried out including both pale and cultivated flax accessions using STRUCTURE 2.3.3 (Pritchard et al. 2000). The admixture model with correlated allele frequencies was used with a burn in of 10,000 and 100,000 iterations for  $K$  populations ranging from 1 to 6. Population structure analysis was conducted separately for  $K$  populations ranging from 1 to 7. Ten runs for each  $K$  value were performed, and the ad hoc statistic  $\Delta k$  was used to determine the optimum number of populations (Evanno et al. 2005). The membership coefficient estimate ( $Q$ ) for each accession was calculated by averaging the ten runs of the best  $K$ . The inferred populations were merged post-analysis and visualized in Distruct (Rosenberg 2004). PCoA was performed in a multidimensional space with data standardization using GENALEX version 6.51 (Peakall and Smouse 2012). In addition, the 407 cultivated flax accessions were divided in four convarieties including *L. usitatissimum* convar. *mediterraneum* (15), *L. usitatissimum* convar. *usitatissimum* (304), *L. usitatissimum* convar. *elongatum* (83) and *L. usitatissimum* convar. *crepitans* (5), with the aim of identifying which convariety was the most closely related to pale flax. Pairwise  $F_{ST}$  comparisons and the unbiased Nei's genetic distance parameter were calculated using GENALEX version 6.51 (Peakall and Smouse 2012) to determine the genetic differentiation between pale and cultivated flax. The partitioning of the genetic variation across populations identified by STRUCTURE was also performed using AMOVA, and the significance of each variation was estimated using 1,000 permutations in GENALEX version 6.51 (Peakall and Smouse 2012). Furthermore, the genetic differentiation among these populations was calculated as the pairwise  $\Phi_{ST}$ . To assess the genetic relationships between the wild progenitor and its cultivated form, a dendrogram was generated using the neighbor-joining (NJ) algorithm based on the Nei

(1973) minimum genetic distance method implemented in PowerMarker version 3.25 (Liu and Muse 2005) and displayed by MEGA 5 (Tamura et al. 2011).

### Genetic diversity, relatedness and gain/loss of diversity

Genetic diversity parameters were estimated for the 125 pale flax and 407 cultivated flax accessions, and across the populations identified by STRUCTURE. The unbiased gene diversity ( $UH_e$ ), observed heterozygosity ( $H_o$ ), total number of alleles ( $N_a$ ), inbreeding coefficient ( $F_{IS}$ ) and polymorphic loci (%) were calculated in GENALEX version 6.51 (Peakall and Smouse 2012). Allelic richness ( $R_s$ ) and private alleles ( $\Pi$ ) were corrected for sample size differences and estimated using the rarefaction method implemented in HP-RARE version 1.2 (Kalinowski 2005). The number of rare alleles (MAF  $< 0.05$ ) and the PIC values were calculated in PowerMarker version 3.25 (Liu and Muse 2005). Significant differences between pale flax and cultivated flax genetic diversity parameters were estimated with the Kruskal–Wallis nonparametric test (Kruskal and Wallis 1952).

Relatedness was estimated using the molecular coancestry parameter ( $f_{ij}$ ) according to Caballero and Toro (2002). The molecular coancestry between two individuals  $i$  and  $j$  is the probability that two randomly sampled alleles from the same locus in two individuals are identical by state (Caballero and Toro 2002). Molecular coancestry matrices comparing all pairs of individuals within the pale and the cultivated flax collections, as well as within the populations identified by STRUCTURE were calculated using all 112 SSRs including MAF  $< 0.05$  in MolKin version 3.0 (Gutiérrez et al. 2005).

The gain/loss of genetic diversity from the entire data set (532 accessions) when each population is removed was assessed with replacement of each population identified by STRUCTURE using the molecular kinship analysis proposed by Caballero and Toro (2002) implemented in MolKin version 3.0 (Gutiérrez et al. 2005). Caballero and Toro (2002) proposed setting priorities for conservation using the maintenance of the maximum overall Nei's (1987) gene diversity (GD) in the preserved set of individuals. This is equivalent to the minimization of the overall molecular coancestry ( $f$ ) because  $GD = 1 - f$  mean.

## Linkage disequilibrium in pale flax

In our previous study, we reported an average genome-wide LD ( $r^2$ ) of 0.036, with a relatively fast decay of  $\sim 1$  cM in the Canadian flax core collection (Soto-Cerda et al. 2013). In pale flax, LD was estimated based on the 65 accessions with available phenotypic data and used for AM by calculating  $r^2$  in GGT 2.0 (van Berloo 2008). The absence of a pale flax linkage map imposed the use of the surrogate cultivated flax linkage map. This assumption is based on the morphological, cytological, molecular characterizations and SSR markers transferability between both species (Diederichsen and Hammer 1995; Fu and Allaby 2010; Fu and Peterson 2010; Gill and Yermamos 1967; Soto-Cerda et al. 2011; Tammes 1928; Uysal et al. 2010, 2012). Allele's frequency was calculated in PowerMarker version 3.25 (Liu and Muse 2005), and MAF < 0.05 was set to "U" (missing data) and excluded from the LD analysis. Genetic distances between markers were obtained from the SSR consensus linkage map of flax (Cloutier et al. 2012) integrated with the physical map (Ragupathy et al. 2011). In total, 104 SSRs with known chromosome information in the consensus map of flax (Cloutier et al. 2012) were used for LD estimation. SSRs on the same linkage group were considered linked and those on different linkage groups, unlinked. All unlinked LD pairwise comparisons were arbitrarily assigned to 140 cM, because at this distance none of the linked pairwise LD comparisons were located. Average genome-wide LD decay versus genetic distance was estimated as described (Bressegello and Sorrells 2006) and displayed in a LD scatter plot. A statistical cutoff value of  $r^2 = 0.1$  was set to estimate the average genome-wide LD block.

## Association mapping

The mean phenotypic values and the mean of the adjusted phenotypic values of nine traits were used for AM in pale flax and cultivated flax, respectively. Five AM models were tested in TASSEL 2.1 (Bradbury et al. 2007) including two general linear models (GLMs) ( $Q$  and PCA) and three mixed linear models (MLMs) ( $K$ ,  $Q + K$  and PCA +  $K$ ) (Price et al. 2006; Pritchard et al. 2000; Yu et al. 2006). The  $Q$  matrix was estimated using the 112 SSRs in STRUCTURE. The PCA matrix calculated in TASSEL 2.1 retained

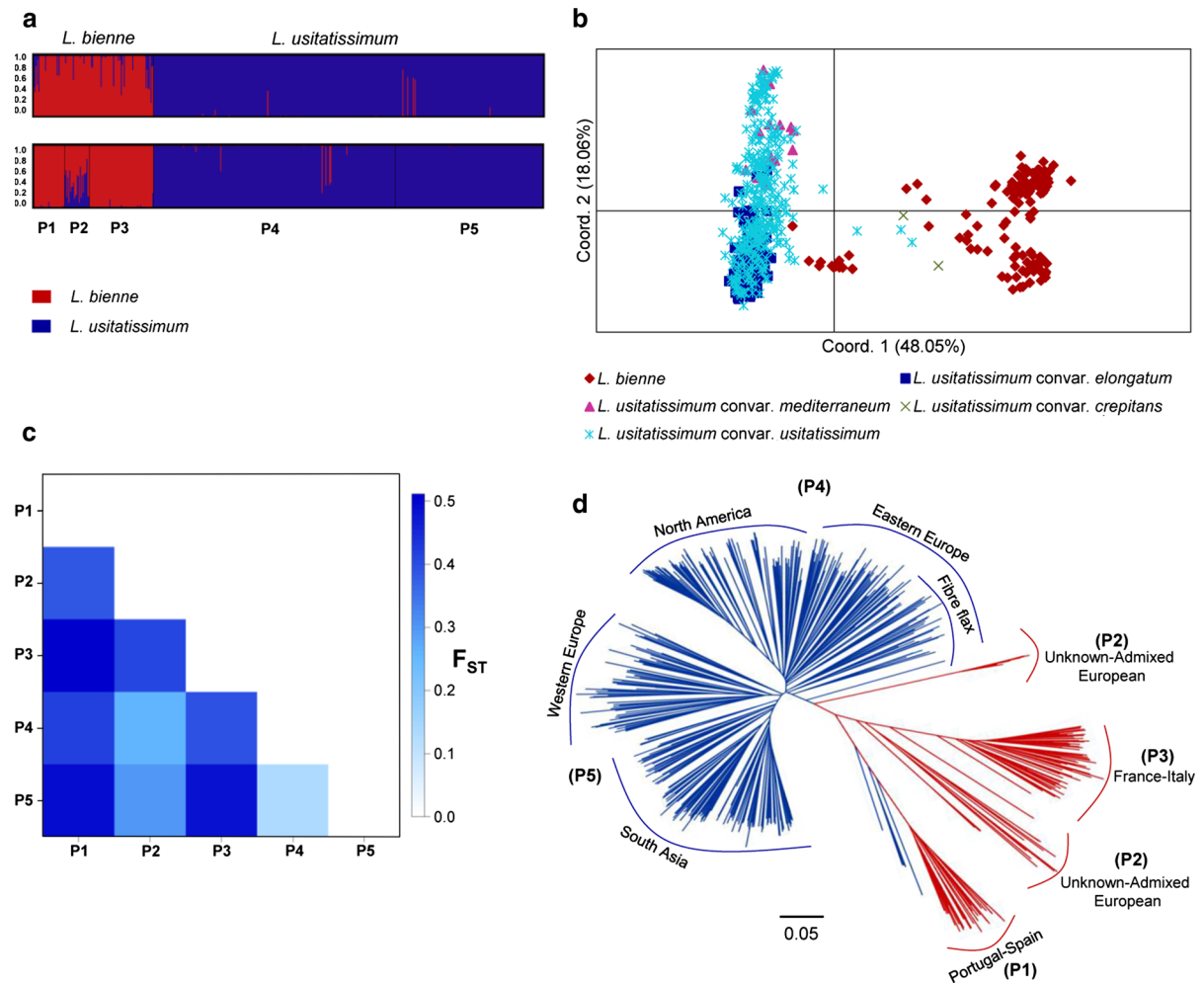
the first three components. The  $K$  matrix was constructed on the basis of 112 SSRs using SPAGeDi (Hardy and Vekemans 2002). All negative values between individuals were set to zero (Yu et al. 2006). The best AM model was selected using probability–probability ( $P$ – $P$ ) plots. For the AM analysis, only minor allele frequency (MAF) > 0.05 was retained (Bressegello and Sorrells 2006).

Correction for multiple testing was performed using the estimated false discovery ( $q$ FDR) values (Benjamini and Hochberg 1995). The  $q$  values were calculated with the QVALUE R package using the smoother method (Storey and Tibshirani 2003). Markers with  $q$ FDR < 0.01 were considered significant. For markers significantly associated with a trait, a GLM with all fixed-effect terms was used to estimate the amount of phenotypic variation explained by each marker ( $R^2$ ). Allelic effects of the significant marker loci were calculated as the difference between the average phenotypic values of the homozygous alleles with MAF > 0.05. The significant differences between the allele means were estimated by the Kruskal–Wallis nonparametric test (Kruskal and Wallis 1952) and visualized as box plots.

## Results

### Population structure and phylogenetic analyses

The Bayesian-based clustering approach implemented in STRUCTURE identified two major groups according to the  $\Delta k$  approach (Fig. 1a, Online Resource 2), which corresponded to the pale and cultivated flax accessions. Separate analysis of the 125 pale flax accessions provided support for the existence of three populations named P1, P2 and P3 (Fig. 1a, Online Resource 2), which clustered accessions from Portugal–Spain, unknown-admixed European and France–Italy, respectively. P2 showed the largest proportion of shared alleles with cultivated flax. Separate analysis of the 407 cultivated flax accessions supported the existence of two major populations, here named P4 and P5 (Fig. 1a), in agreement with our previous study (Soto-Cerda et al. 2013). P4 contained mostly accessions from North America and Eastern Europe with a strong effect of the fiber morphotype from Eastern European countries, while P5 mostly clustered accessions from South Asia and Western Europe. Similar to



**Fig. 1** Population structure and genetic relationships between pale flax (*L. bienne*) and cultivated flax (*L. usitatissimum*). **a** structure analysis for  $K = 2$  (upper panel) and the merged STRUCTURE ( $K = 2$  and  $K = 3$ ) populations (lower panel) of *L. bienne* (red) and *L. usitatissimum* (blue) accessions. P1, P2 and P3 represent populations within *L. bienne* major group and P4 and P5 represent populations within *L. usitatissimum* major group. **b** Principal coordinate analysis (PCoA) of the 532 pale and cultivated flax accessions. Cultivated flax was further divided into four convarieties, namely *L. usitatissimum* convar.

the STRUCTURE analysis, the PCoA revealed the presence of two major groups where coordinates 1 and 2 explained 66 % of the total genetic variation (Fig. 1b). A few pale flax accessions from P2 grouped close to the fiber flax accessions, of which the majority corresponded to the convar. *elongatum* (Fig. 1a, b). Likewise, the convar. *crepitans* accessions clustered close to the pale flax major group, having in common dehiscent capsules (Fig. 1b).

*mediterraneum*, *L. usitatissimum* convar. *usitatissimum*, *L. usitatissimum* convar. *elongatum* and *L. usitatissimum* convar. *crepitans*. **c** Pairwise  $F_{ST}$  comparisons among the five populations inferred by STRUCTURE. **d** Phylogenetic tree created using the Neighbor-joining (NJ) algorithm (Nei 1973). Colored clusters represent pale and cultivated flax STRUCTURE ( $K = 2$ ) major groups. P1–P5 correspond to the merged STRUCTURE ( $K = 2$  and  $K = 3$ ) populations, indicating their geographic distribution. The scale bar indicates the Nei (1973) minimum genetic distance

The coefficient of population differentiation  $F_{ST}$  and the unbiased Nei's genetic distance parameter between pale and cultivated flax was estimated at 0.23 ( $P < 0.001$ ) and 0.65, respectively, indicating a strong population structure. Pairwise  $F_{ST}$  among the five populations identified by STRUCTURE ranged from 0.07 ( $P < 0.001$ ) between P4 and P5 to 0.35 ( $P < 0.001$ ) between P1 and P5 (Fig. 1c). Among the pale flax populations, P2 exhibited the weakest

population structure when compared to the cultivated flax populations (Fig. 1c).

The AMOVA showed highly significant ( $P < 0.001$ ) differentiation between the five STRUCTURE populations, which indicated that 79 % of the observed variation can be explained by the within populations subdivision (Online Resource 3). The  $\Phi_{ST}$  value (0.21) indicated a strong population differentiation similar to the  $F_{ST}$  value. The high  $F_{ST}$ ,  $\Phi_{ST}$  and unbiased Nei's genetic distance values between pale flax and cultivated flax suggest that AM analysis should be conducted separately.

The phylogenetic analysis using the NJ algorithm also partitioned the 532 accessions into two major groups, largely corresponding to the pale and cultivated flax major groups identified by STRUCTURE and PCoA (Fig. 1d). Further, the cultivated flax accessions were divided into two populations resembling the P4 and P5 STRUCTURE populations. Likewise, the pale flax accessions were divided into three populations also resembling but not identical to the P1, P2 and P3 STRUCTURE populations (Fig. 1d). Similarly to the PCoA, a few pale flax accessions from P2 grouped close to the fiber flax cultivars, while the five convar. *crepitans* accessions clustered close to the pale flax accessions from Portugal–Spain.

#### Genetic diversity, relatedness and gain/loss of diversity

In pale flax, the mean number of alleles adjusted for sample size ( $N_{a,adj}$ ) was 8.03, of which, 55.6 % had a MAF  $< 0.05$  and were considered rare alleles ( $R_a$ ) (Table 1). In comparison, the mean  $N_{a,adj}$  was 6.59 in the cultivated flax accessions, of which, 58.5 % were considered rare alleles. The observed heterozygosity ( $H_o$ ), the total unbiased gene diversity ( $U_{H_e}$ ), the inbreeding coefficient ( $F_{IS}$ ), the allelic richness ( $R_s$ ), the number of private alleles ( $\Pi$ ), the level of relatedness ( $f_{ij}$ ) and the gain/loss of genetic diversity parameters for the 125 pale flax accessions, the 407 cultivated flax accessions and the five STRUCTURE populations are summarized (Table 1). The cultivated flax accessions harbored similar genetic diversity to that observed in pale flax, based on the nonsignificant mean differences ( $P > 0.05$ ) for  $N_a$ ,  $H_o$ ,  $R_s$  and  $R_a$  but for  $U_{H_e}$  ( $P = 0.011$ ) and  $\Pi$  ( $P = 1.41E-6$ ). In general, P2 and P4 exhibited superior genetic diversity parameters within their respective species.

The  $f_{ij}$  for pale flax was estimated at 0.33, whereas it was 0.29 for cultivated flax. The intrapopulation  $f_{ij}$  ranged from 0.24 (P2 and P4) to 0.33 (P1 and P3). The coancestry analysis indicated that most of the accessions within the five populations had weak familial relatedness.

The gain/loss of genetic diversity was assessed with replacement using the molecular kinship analysis. The greatest loss of genetic diversity was observed when the 125 pale flax accessions were removed from the data set ( $\sim 15$  %). Thereafter, the 125 pale flax accessions were added in the data set and the 407 cultivated flax accessions removed, generating  $\sim 4.4$  % loss of genetic diversity (Table 1). When any of the pale flax populations were removed, there was a loss of diversity while removing any of the cultivated flax populations caused a gain of diversity. These results suggest that redundant alleles exist within P4 and P5, but unique allelic variation is present in pale flax populations, in agreement with the large number of private alleles.

#### Linkage disequilibrium in pale flax

To analyze LD variation, genetic distances for 104 SSRs were available from the consensus linkage map of flax (Cloutier et al. 2012). In the 65 pale flax accessions assessed, the average  $r^2$  values for linked and unlinked markers were 0.12 and 0.09, respectively. The average genome-wide LD decayed to 0.1 within 1.8 cM (Online Resource 4). Strong LD ( $r^2 > 0.4$ ), suitable for efficient marker assisted selection (MAS), decayed even faster, to less than 0.1 cM.

#### Association mapping

As depicted by the  $P$ – $P$  plots for both pale and cultivated flax, numerous spurious associations for all traits were observed with the GLM ( $Q$ ) (Fig. 2, Online Resource 5). This model was characterized by an excess of small  $P$  values causing spurious associations. On the other hand, the GLM (PCA) over corrected the majority of the small  $P$  values. The MLMs performed differently in both species. For example, in pale flax, the  $K$  and  $Q + K$  models had the smallest deviations from the expected distribution for SPB and PH, while the PCA +  $K$  model was the most suitable for the other seven traits (Fig. 2a, Online Resource 5). In cultivated flax, the PCA +  $K$  model

**Table 1** Genetic diversity in pale flax and cultivated flax STRUCTURE populations

Population	$N^1$	$N_a^2$	$N_{aadj}^3$	$H_o^4$	$UH_c^5$	$F^6$	$R_s^7$	$II^8$	$R_a^9$	$f_{ij}^{10}$	Gain (+)/ loss (-) (%) <sup>11</sup>	Polymorphic loci (%)
Pale flax	125	899	899	0.02	0.59	0.97	7.96	430	500	0.33	-14.7	100
P1	32	387	376	0.01	0.37	0.95	3.36	93	100	0.33	-1.43	79.5
P2	26	529	529	0.02	0.55	0.98	4.67	110	178	0.24	-0.61	98.2
P3	67	541	437	0.02	0.35	0.93	3.90	97	285	0.33	-2.24	92.9
Cultivated flax	407	838	738	0.03	0.53	0.95	6.24	269	490	0.29	-4.37	100
P4	253	790	544	0.03	0.53	0.95	4.86	81	437	0.24	+0.68	99.1
P5	154	601	440	0.03	0.43	0.93	3.93	37	301	0.29	+1.67	99.1

<sup>1</sup> Number of accessions<sup>2</sup> Number of alleles<sup>3</sup> Number of alleles adjusted for sample size<sup>4</sup> Observed heterozygosity<sup>5</sup> Unbiased gene diversity<sup>6</sup> Inbreeding coefficient<sup>7</sup> Allelic richness<sup>8</sup> Number of private alleles estimated on a sample of balanced size using rarefaction method (Kalinowski 2005)<sup>9</sup> Rare alleles <5 %<sup>10</sup> Molecular coancestry<sup>11</sup> Gain/loss of genetic diversity (Caballero and Toro 2002)

had the smallest deviation from the expected distribution for all traits (Fig. 2b, Online Resource 5). Taking into account the  $P$ - $P$  plot results, three different MLMs were used for AM in pale flax, whereas only the PCA +  $K$  model was used in cultivated flax.

#### Marker-trait associations

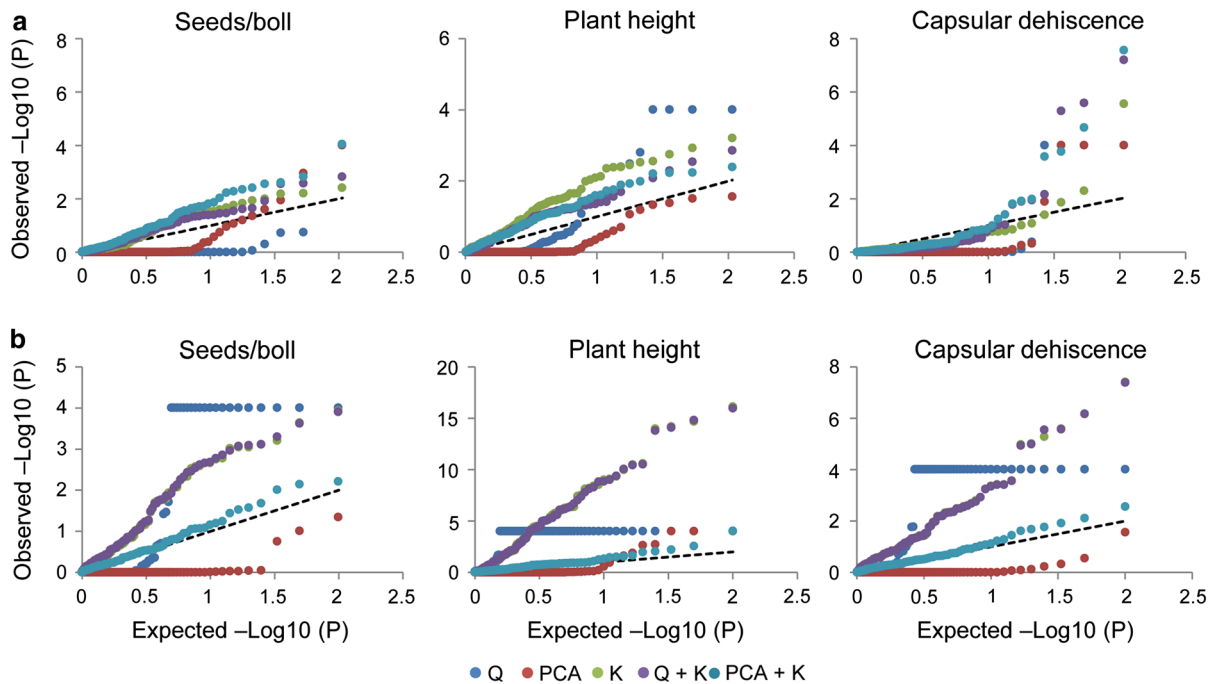
Association mapping was carried out to identify SSRs potentially associated with the nine traits. Significant associations ( $qFDR < 0.01$ ) after Bonferroni correction ( $0.05/112 = 4.46E-4$ ) were identified for seven and four of the nine traits in pale and cultivated flax, respectively. An overview of the associated markers, their map position, correlations with traits ( $R^2$ ) and the effects of favorable alleles is presented in Table 2. A single marker for SPB, PH, FL5 % and PC and two for TSW and DEH were identified in pale flax, while a single marker was identified for TSW, PH and PO and two for FL5 % in cultivated flax but none were common between the two species. In pale flax, the percentage of the total phenotypic variation explained by the associated markers ranged from 1.3 to 97 %, whereas in cultivated flax it ranged from 1.4 to 16.3 % (Table 2).

In pale flax, TSW could be significantly increased by an average of 0.49 g ( $P = 0.002$ ) as an effect of the 274-bp allele of marker Lu652 (Fig. 3a). Similarly, the 280-bp allele of marker Lu1171 had the largest effect on SPB, increasing it by 1.56 seeds ( $P = 0.01$ ). An increase of 0.1 units ( $P = 0.0012$ ) in DEH was associated with the 303-bp allele of Lu2344, but all the genotypes carrying it had dehiscent capsules (Fig. 3a). For cultivated flax, TSW could be improved by an average of 1.64 g ( $P = 6.3E-4$ ) as an effect of the 162-bp allele of marker Lu2042 (Fig. 3b). A reduction of up to 20.14 cm ( $P = 9.2E-4$ ) in PH was associated with the 205-bp allele of Lu2067a compared with the 213-bp allele (Fig. 3b). The 271-bp allele of Lu943 significantly shortened FL5 % by 2.5 days ( $P = 3.65E-5$ ) compared to the other two alleles (Fig. 3b).

#### Allelic abundance and frequency of associated markers

The number of alleles harbored by the associated loci, including  $MAF < 0.05$ , was compared between pale and cultivated flax (Online Resource 6). Taking into account all the markers identified in both species, the observed number of alleles in pale flax was 110 (9.2/





**Fig. 2** Comparisons of five association mapping models in pale and cultivated flax for seeds per boll, plant height and capsular dehiscence. Probability–probability ( $P$ – $P$ ) plots of observed versus expected  $-\log_{10}(P)$  values for **a** Pale flax. **b** Cultivated flax.  $Q$  general linear model using the  $Q$  matrix, PCA general

linear model using the PCA matrix,  $K$  mixed linear model using the kinship matrix,  $Q + K$  mixed linear model using the  $Q$  and  $K$  matrices, PCA +  $K$  mixed linear model using the PCA and  $K$  matrices

locus) while in cultivated flax it was 99 (8.2/locus). A total of 53 common alleles (4.4/locus) were observed across the twelve associated loci. In general, pale flax harbored more alleles than cultivated flax but as indicated by the Kruskal–Wallis test, the means were not statistically different ( $P = 0.321$ ). For the twelve associated loci, pale flax registered 57 private alleles. The 262-, 265- and 277-bp alleles of marker Lu451b and the 208- and 247-bp alleles of marker Lu652 associated with TSW in pale flax were absent in cultivated flax (Online Resource 7). Similarly, marker Lu1171 associated with SPB registered six alleles absent in cultivated flax including the 280-bp allele with the largest effect on SPB (Table 2). For the same marker, however, the most frequent alleles in cultivated flax (265 and 268 bp) were rare or absent in pale flax (Online Resource 7). Marker Lu265 associated with PH in pale flax showed nine alleles absent in cultivated flax although at low frequencies (Online Resource 7). The most common alleles in cultivated flax, 256 and 259 bp (combined frequency of 90 %), were rare or absent in pale flax. Similarly, the null allele with the largest effect and highest frequency in

pale flax was absent in cultivated flax (Table 2, Online Resource 7). For FL5 %, pale flax had one allele that was absent in cultivated flax (Online Resource 7).

Notably, among the 65 pale flax accessions, six had violet petals. All violet accessions carried the 360-bp allele of marker Lu2725 which increased PC by four units (Table 2). Conversely, alleles 330-, 332- and 334-bp alleles (combined frequency of 67 %), associated with light blue petals, were absent (332) and rare (330 and 334) in cultivated flax (Online Resource 7).

Likewise, the alleles of markers Lu2344 and Lu442a, associated with DEH in pale flax, were either absent or rare in cultivated flax. For example, the 284-, 288- and 315-bp alleles of Lu2344 were absent in cultivated flax, while the 285- and 303-bp alleles had frequencies of 1.1 and 3.7 %, respectively (Online Resource 7). The five ( $\sim 1.1$  %) *conv.* *crepitans* accessions characterized by their dehiscent capsules were the only ones carrying the 285-bp allele. In pale flax, however, the 284-, 285-, 288-, 303- and 315-bp alleles of Lu2344 had frequencies in the range of 4.1 to 45 %. Moreover, the two most common alleles in cultivated flax, the 297 (52.2 %) and 306 bp (33.5 %) alleles

**Table 2** Marker loci significantly associated with thousand seed weight (TSW), seeds per boll (SPB), dehiscent capsules (DEH), plant height (PH), start of flowering (FL5 %), flowershape (FS), petal color (PC) and petal overlap (PO), and their explained phenotypic variance ( $R^2$ ), the allelic effects of favorable alleles and their frequencies

Species	Trait	Marker	LG (cM)	<i>P</i> value	$R^2$ (%)	Effect <sup>a</sup>	Favorable allele (bp)	Frequency (%)	
Pale flax	TSW	Lu451b	Unknown	1.42 E−4**	18.9	0.07 g	202	78.2	
		Lu652	5 (79.4)	9.54 E−5**	28.8	0.49 g	274	9.2	
	SPB	Lu1171	8 (27.6)	3.28 E−4**	4.8	1.56 U	280	5.8	
		DEH	Lu2344	2 (89.3)	2.20 E−5**	20.1	0.1 U	303	47.1
			Lu442a	6 (44.3)	2.77 E−8**	9.0	0.08 U	267	14.4
	PH	Lu265	8 (37.1)	5.76 E−4*	14.3	−20.43 cm	null	52.3	
	FL5%	Lu271	15 (53.6)	4.91 E−4*	8.5	−11.74 d	271	92.6	
	FS	Lu2344	2 (89.3)	2.98 E−5**	1.31	0.60 U	315	15.9	
	PC	Lu2725	10 (40.1)	2.00 E−13**	96.9	4 U	360	5.9	
Cultivated flax	TSW	Lu2042	12 (18.2)	7.64 E−5**	6.74	1.46 g	162	7.0	
	PH	Lu2067a	2 (59.7)	9.45 E−5**	16.3	−20.14 cm	205	27.6	
	FL5%	Lu943	1 (150.0)	3.86 E−4**	9.2	−2.5 d	271	60.8	
		Lu2067a	2 (59.7)	2.09 E−5**	2.3	−2.1 d	209	14.7	
	PO	Lu3038	14 (32.2)	7.95 E−4*	1.4	0.1 U	303	6.2	

\*\* Significant at  $qFDR < 0.01$  and after Bonferroni correction ( $0.05/112 = 4.46 E−4$ )

\* Significant at  $qFDR < 0.01$

<sup>a</sup> Effects of favorable alleles are represented in grams (g) for TSW, days (d) for FL5%, centimeters (cm) for PH and units (u) for SPB, DEH, FS, PC and PO

were absent in the first case and rare in the second in pale flax. Because the low marker density might have influenced our ability to detect significant associations for DEH in cultivated flax, we considered 464 genome-wide SSRs from our previous work (Soto-Cerda et al. 2014) to conduct AM. Despite the increase marker density, no associations were detected, likely because of the small number of dehiscent accessions in the Canadian core collection (data not shown).

## Discussion

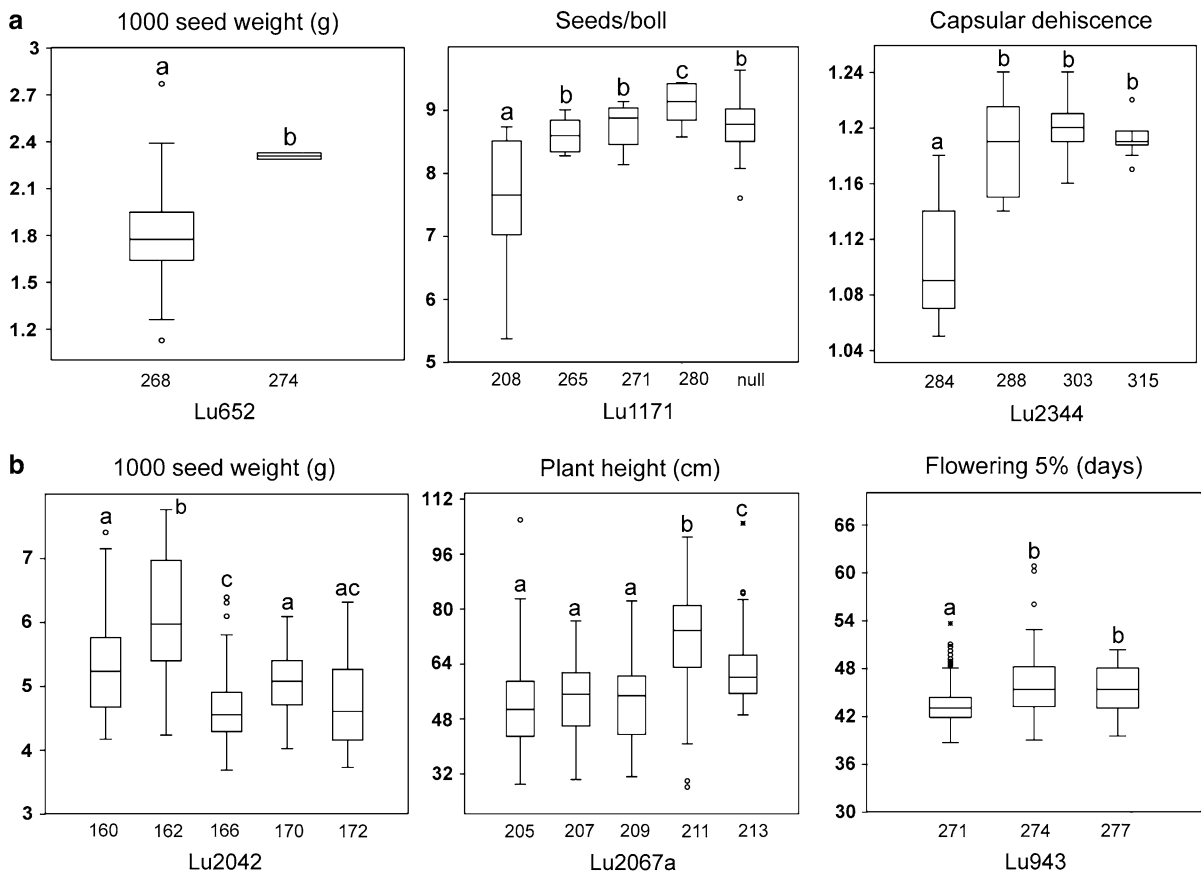
The remarkable diversity of wild relatives offers a reservoir of genetic variation that has the potential to positively impact crop improvement and support our food system in the future (McCouch et al. 2013). The genetic diversity in cultivated flax, one of the oldest oil and fiber crop domesticated by early civilizations, has been generally described as narrow (Cloutier et al. 2009; Diederichsen and Fu 2006; Fu et al. 2002, 2003; Smýkal et al. 2011; Soto-Cerda et al. 2012). Here, we reported on the genetic diversity, population structure and AM of the two largest collections of cultivated flax (Diederichsen and Fu 2006; Everaert et al. 2001; Fu

et al. 2002, 2003; Fu 2005; Rajwade et al. 2010; Smýkal et al. 2011; Soto-Cerda et al. 2012) and its wild progenitor, pale flax (Allaby et al. 2005; Diederichsen and Hammer 1995; Fu 2011; Uysal et al. 2010, 2012). Cultivated and pale flax possess similar levels of genetic diversity, and pale flax could be a source of useful unique alleles for economically important traits in cultivated flax.

## Population structure and phylogenetic analyses

The STRUCTURE, PCoA and phylogenetic analyses were fully consistent with the phenotypic differences exhibited by pale and cultivated flax (Diederichsen and Hammer 1995; Uysal et al. 2012), as both species were separated into two genetically distinct groups. Further, the PCoA and the NJ tree, showed that the dehiscent convar. *crepitans* accessions most closely resembled its wild progenitor.

The  $F_{ST}$  pairwise comparisons indicated stronger population structure between pale flax populations than between cultivated flax populations, a phenomenon that could be explained by limited gene flow or strong local adaptation of pale flax accessions (Uysal et al. 2010). Conversely, the weak  $F_{ST}$  value between cultivated flax



**Fig. 3** Comparisons of allelic effects of associated markers with agronomic and phenological traits. **a** Allelic effects on thousand seed weight, seeds per boll and capsular dehiscence in pale flax. **b** Allelic effects on thousand seed weight, plant height

and 5 % flowering in cultivated flax. Same letters above the box plots indicated values that do not differ statistically according to the Kruskal–Wallis test ( $\alpha = 0.01$ )

populations might reflect the more intensive germplasm exchange between gene banks and breeding programs around the world (Maggioni et al. 2002). Overall, our combined population structure analyses were consistent with clustering patterns obtained in previous molecular and phenotypic studies (Fu 2011; Uysal et al. 2010, 2012), supporting capsular dehiscence as the primary trait modified in early flax domestication, preceding other domestication traits such as oil content and fiber length (Fu 2011; Uysal et al. 2010).

Genetic diversity, relatedness and gain/loss of diversity

The universality of loss of diversity in cultivated crops relative to their wild ancestors is well documented (Dubkovsky and Dvorak 2007; Liu and Burke 2006; Wright and Gaut 2005). However, we observed a

similar level of diversity for most of the parameters between pale and cultivated flax (Table 1). ISSR analysis revealed similar results despite the limited number of accessions (Uysal et al. 2010). These findings suggest that much of the diversity that survived through the stages of domestication has been retained in gene banks and is well represented in the Canadian flax core collection (Diederichsen et al. 2013). Nevertheless, these seemingly similar levels of diversity could be misleading because the pale flax group of only 125 accessions is not a core collection but represent a collection primarily selected based on the availability and accuracy of the passport data and not based on diversity parameters. As a consequence, differences in population size and sampling methods can inflate the diversity harbored by the flax core collection as compared to a random sample of cultivated flax, making our comparisons biased.

The significantly higher percentage of genetic variation contained within populations as indicated by AMOVA, the number of private alleles and the 15 % loss of diversity when the 125 pale flax accessions were removed from the entire data set support the widely held view that pale flax is a potential source of novel alleles for cultivated flax improvement and therefore deserves special conservation efforts (Diederichsen 2007). The weak intra-population relatedness suggests that intra- as well as interpopulation breeding could be fruitful.

#### Linkage disequilibrium in pale flax

Generally, LD decays more slowly in selfed pollinated species as compared to outcrossing ones (Flint-Garcia et al. 2003). Further, in selfed pollinated species, LD declines more rapidly in wild relatives and landraces than in their cultivated counterpart such as *O. rufipogon* (Xu et al. 2013), *G. soja* (Lam et al. 2010) and *H. vulgare* ssp. *spontaneum* (Morrell et al. 2005). In pale flax, the average LD block based on 65 accessions was estimated at  $\sim 1.8$  cM. However, it is expected that using the 125 accessions, LD should decay more rapidly to similar levels or even faster than the  $\sim 1$  cM estimated in the Canadian flax core collection (Soto-Cerda et al. 2013) based on the assumption that the marker order in a linkage map of pale flax would be similar to that observed in the linkage map of cultivated flax. An extreme example where LD remains high over long genetic distance (6–20 cM) is cultivated tomato where several bottlenecks that occurred during domestication hamper the use of AM (van Berloo 2008). Nevertheless, part of the cultivated tomato germplasm displays a genetic admixture with the wild relative *S. lycopersicum* var. *cerasiforme*. Such admixture could be compared with a multi parent advanced generation intercross (MAGIC) population that would provide greater level of genetic diversity and faster LD decay (Ranc et al. 2012). Similarly, pale flax can be used not only as a source of novel alleles for flax breeding but also to design interspecific MAGIC populations harboring superior allelic diversity and smaller LD blocks, both desirable properties for AM.

#### Association mapping

An AM panel of diverse breeding history and/or geographic origin might exhibit population and family

structure, both recognized as sources of spurious associations (Flint-Garcia et al. 2003; Würschum 2012). Several linear and mixed models have been proposed to correct for the effect of both confounding factors (Price et al. 2006; Pritchard et al. 2000; Yu et al. 2006) but the best fitting model will depend on the genetic architecture of the traits and their correlations with population and family structure. In pale flax, the  $K$  and  $Q + K$  models exhibited the best  $P$  value fitting for SPB and PH, although for most of the traits the three MLMs assessed performed similarly. In cultivated flax, however, the  $K$  and  $Q + K$  models tended to generate more false positives as compared to the  $PCA + K$  model, suggesting strong differences in selective and demographic forces shaping allele frequencies and trait architecture in both species. In pale flax, natural selection may have been the main force determining allele frequencies and genetic architecture of traits as an adaptive process, perhaps minimizing the negative effects of population structure on model performance for the traits assessed. On the contrary, the genotypes of the flax core collection have undergone divergent artificial selection for fiber, agronomic and seed quality traits, creating strong allele frequency differences.

#### Marker–trait associations

The aim of the AM analysis was to identify genomic regions in pale flax potentially useful for cultivated flax breeding rather than identifying environment-specific QTL or QTL-by-environment interaction patterns. Significant associations were identified for seven and four traits in pale and cultivated flax, respectively, but none of the associated markers were common. These differences could be accounted by several factors including (1) the downside of unbalanced allele frequencies between flax morphotypes (Soto-Cerda et al. 2014), (2) differences in LD decay between pale and cultivated flax which requires lower marker density in the former to cover more LD blocks across the genome (Würschum 2012), and (3) differences in trait architecture and heritability (Würschum 2012). Despite the few marker-trait associations identified herein, we provided a proof of concept for pale flax functional variants as potentially useful for cultivated flax improvement.

Plant breeders have long recognized the value of wild relatives for the improvement of simply inherited traits, including disease and insect resistance or

cytoplasmic male sterility (Feuillet et al. 2008; Xiao et al. 1998). Despite these successes, their uses in quantitative trait breeding such as yield have been rather limited because agriculturally desirable alleles are present in low frequency and are often masked by the effects of deleterious alleles in wild species (Swamy and Sarla 2008). Nevertheless, successful introgressions of exotic alleles to enhance yield have been reported in cultivated tomato (de Vicente and Tanksley 1993), cultivated oat, wheat, barley and maize (Frey et al. 1983). In pale flax, we identified associations with positive effects for TSW and SPB, two important yield components in flax breeding. Although the overall TSW and SPB values were lower than cultivated flax, pale flax's alleles can improve the traits as demonstrated in cultivated tomato where de Vicente and Tanksley (1993) identified QTL for eleven traits that had an allelic effect opposite of what would be expected based on the poor phenotype of the wild parent. Similar observations were made in barley relatives (Schmalenbach et al. 2009). Recently, AM studies have identified favorable QTL for phenological, agronomical and kernel composition traits in teosinte (Weber et al. 2009), for spot blotch resistance in wild barley (Roy et al. 2010) and for fruit quality traits in wild tomato (Xu et al. 2013). Taken together, these studies demonstrate the potential to improve multiple traits in cultivated crops using wild relatives and AM.

Yield improvement through yield components and related traits such as flowering time and PH could be advantageous because of their simpler genetic architecture and higher stability than yield per se (Peng et al. 2011). PH is an important developmental and yield-related trait, and many genes regulating PH have been shown to affect harvest index and yield in rice (Xue et al. 2008), and yield and flowering time in soybean (Liu et al. 2011). Pale flax possessed five, six, nine and one novel alleles for TSW, SPB, PH and FL5 %, respectively, potentially useful for yield improvement through yield components in cultivated flax. For example, the favorable alleles associated with SPB and PH in pale flax, which were absent in cultivated flax, could be the first targets for the introgression of new functional variation in cultivated flax.

Flower color is determined by various pigments and co-pigments (Grotewold 2006). Flower color protects against UV-light and attracts pollinators, and the evolution of new pollinator associations are often

accompanied by shifts in flower color (Fenster et al. 2004). In pale flax, marker Lu2725 explained ~97 % of PC variation, and based on the analysis of allele frequencies, the most common alleles in pale flax were absent in cultivated flax and vice versa. Comparison of phenotypic characters between both species revealed that most of the darker petal colors are absent in pale flax but they are in high frequency in cultivated flax, suggesting that PC could be another selected trait during domestication by ancient humans (Uysal et al. 2012).

Seed shattering, pod shattering or capsule dehiscence are common adaptive strategies in plants to ensure seed dispersal and survival. However, this commonly found wild species trait is undesirable in their domesticated counterparts because it results in yield losses prior to and during harvest (Lin et al. 2012). In pale flax, we identified markers Lu2344 and Lu442a associated with capsules dehiscence accounting for ~30 % of the variation. As with PC, the most common alleles in pale flax have been selected against in cultivated flax and vice versa. These markers could be exploited to perform early selection of non-dehiscent genotypes in interspecific populations, enabling removal of this undesirable trait while maintaining favorable exotic alleles.

#### Allelic abundance and frequency of associated markers

Among the 12 associated markers identified herein, more than 48 % of the alleles were specific to pale flax, suggesting that these alleles most likely were lost through directional selection during domestication. The introgression of pale flax-specific alleles associated with TSW, SPB, PH and FL5% may be a useful strategy for widening the phenotypic variation in cultivated flax and counteract the inbreeding depression observed in the last decade or two in linseed in Canada. However, wild relatives' introgressions are often associated with undesirable linkage drag (Feuillet et al. 2008). Depressed recombination between homeologous chromosomes is directly related to the evolutionary distance between species (Feuillet et al. 2008). Pale flax has been reported to cytogenetically differ from cultivated flax by a single translocation (Gill and Yermanos 1967), in agreement with morphological, phylogenetic and marker transferability evidence (Diederichsen and Hammer 1995; Fu and

Allaby 2010; Fu and Peterson 2010; Uysal et al. 2010, 2012). Hence, recombination should not be an impediment to fully deploy pale flax in cultivated flax breeding. Current methods to transfer desirable wild alleles into cultivated material rely mostly on advanced backcross quantitative trait loci (AB-QTL) (Tanksley and Nelson 1996) and the introgression lines (ILs) methods (Zamir 2001). These approaches are limited to the evaluation of a single wild genotype at a time and require the need for validating the positive phenotypic effect of an exotic allele across different elite genetic backgrounds (Swamy and Sarla 2008). Alternatively, we propose the construction of interspecific MAGIC populations between pale and cultivated flax. This approach can overcome many of the limitations of conventional AB-QTL, ILs and AM populations (Morrell et al. 2012). For example, combining multiple parents across multiple generations of intermating can increase the rate of effective recombination per generation, improving genetic resolution, and the relatively even contribution of all parents to allele frequencies helps tackling the negative effects of population and family structures and is more effective in sampling rare alleles (Morrell et al. 2012).

In conclusion, our results corroborate the long held view that pale flax is a potential source of novel variation for cultivated flax. Pale flax contains not only high genetic diversity but also unique and rare alleles for agronomically important traits. To fully exploit this wild relative germplasm, new strategies such as MAGIC populations hold promise. However, the construction of a highly saturated linkage map is required for the accurate estimation of LD patterns and comparative mapping. The re-sequencing of the 125 pale flax accessions is underway and, along with the recently identified ~1.8 million single-nucleotide polymorphisms generated from the re-sequencing of the Canadian flax core collection, these resource promise to expand our knowledge of genome evolution, domestication events, LD patterns, QTL variation and architecture for multiple traits in cultivated flax and its wild ancestor, pale flax.

**Acknowledgments** The authors are grateful to Andrzej Walichnowski, Elsa Reimer, Natasa Radovanovic, Evelyn Miranda and the breeding teams at the Morden Research Station and the Crop Development Centre for technical assistance. This work was conducted as part of the Total Utilization Flax Genomics (TUFGEN) project funded by

Genome Canada and co-funded by the Government of Manitoba, the Flax Council of Canada, the Saskatchewan Flax Development Commission, Agricultural Development Fund and the Manitoba Flax Growers Association. Project management and support by Genome Prairie are also gratefully acknowledged. Braulio J. Soto-Cerda was supported by Becas Chile—Comisión Nacional de Investigación Científica y Tecnológica (CONICYT).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

## References

- Allaby RG, Peterson GW, Merriwether DA, Fu Y-B (2005) Evidence of the domestic history of flax (*Linum usitatissimum*) from genetic diversity of the *sad2* locus. *Theor Appl Genet* 112:58–65
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B* 57:289–300
- Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633–2635
- Breseghele F, Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics* 172:1165–1177
- Caballero A, Toro MA (2002) Analysis of genetic diversity for the management of conserved subdivided populations. *Conserv Genet* 3:289–299
- Cloutier S, Niu Z, Datla R, Duguid S (2009) Development and analysis of EST-SSRs for flax (*Linum usitatissimum* L.). *Theor Appl Genet* 119:53–63
- Cloutier S, Ragupathy R, Miranda E, Radovanovic N, Reimer E, Walichnowski A, Ward K, Rowland G, Duguid S, Banik M (2012) Integrated consensus genetic and physical maps of flax (*Linum usitatissimum* L.). *Theor Appl Genet* 125:1783–1795
- de Vicente MC, Tanksley SD (1993) QTL analysis of transgressive segregation in an interspecific tomato cross. *Genetics* 134:585–596
- Diederichsen A (2007) Ex situ collections of cultivated flax (*Linum usitatissimum* L.) and other species of the genus *Linum* L. *Genet Resour Crop Evol* 54:661–678
- Diederichsen A, Fu YB (2006) Phenotypic and molecular (RAPD) differentiation of four infraspecific groups of cultivated flax (*Linum usitatissimum* L. subsp. *usitatissimum*). *Genet Resour Crop Evol* 53:77–90
- Diederichsen A, Hammer K (1995) Variation of cultivated flax (*Linum usitatissimum* L. subsp. *usitatissimum*) and its wild progenitor pale flax (subsp. *angustifolium* (Huds.) Thell.). *Genet Resour Crop Evol* 42:263–272
- Diederichsen A, Richards K (2003) Cultivated flax and the genus *Linum* L. Taxonomy and germplasm conservation. In: Muir AD, Westcott ND (eds) flax. The genus *Linum*, CRC Press, Boca Raton, pp 22–54

- Diederichsen A, Kusters PM, Kessler D, Baines Z, Gugel RK (2013) Assembling a core collection from the flax world collection maintained by Plant Gene Resources of Canada. *Genet Resour Crop Evol* 60:1479–1485
- Dubkovsky J, Dvorak J (2007) Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science* 316:1862–1866
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620
- Everaert I, De Riek J, De Loose M, Van Waes J, Van Bockstaele E (2001) Most similar variety grouping for distinctness evaluation of flax and linseed (*Linum usitatissimum* L.) varieties by means of AFLP and morphological data. *Plant Var Seeds* 14:69–87
- Fenster CB, Armbruster WS, Wilson P, Dudash MR, Thompson JD (2004) Pollination syndromes and floral specialization. *Annu Rev Ecol Syst* 35:375–403
- Feuillet C, Langridge P, Waugh R (2008) Cereal breeding takes a walk on the wild side. *Trends Genet* 24:24–32
- Flint-Garcia S, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. *Annu Rev Plant Biol* 54:357–374
- Frey KJ, Cox TS, Rodgers DM, Bramel-Cox P (1983) Increasing cereal yields with genes from wild and weedy species. In: Chopra VL, Joshi BC, Sharma RP, Bansal HC (eds) Proceedings of the XV international congress of genetics, vol IV, pp 51–68
- Fu YB (2005) Geographic patterns of RAPD variation in cultivated flax. *Crop Sci* 45:1084–1091
- Fu YB (2011) Genetic evidence for early flax domestication with capsular dehiscence. *Genet Resour Crop Evol* 58:1119–1128
- Fu YB, Allaby RG (2010) Phylogenetic network of *Linum* species as revealed by non-coding chloroplast DNA sequences. *Genet Resour Crop Evol* 57:667–677
- Fu YB, Peterson GW (2010) Characterization of expressed sequence tag-derived simple sequence repeat markers for 17 *Linum* species. *Botany* 88:537–543
- Fu YB, Diederichsen A, Richards KW, Peterson G (2002) Genetic diversity within a range of cultivars and landraces of flax (*Linum usitatissimum* L.) as revealed by RAPDs. *Genet Resour Crop Evol* 49:167–174
- Fu YB, Rowland GG, Duguid SD, Richards KW (2003) RAPD analysis of 54 North American flax cultivars. *Crop Sci* 43:1510–1515
- Gill KS, Yermanos DM (1967) Cytogenetic studies on the genus *Linum* I. Hybrids among taxa with 15 as the haploid chromosome number. *Crop Sci* 7:623–627
- Grotewold E (2006) The genetics and biochemistry of floral pigments. *Annu Rev Plant Biol* 57:761–780
- Gutiérrez JP, Royo LJ, Álvarez I, Goyache F (2005) MolKin v2.0: a computer program for genetic analysis of populations using molecular coancestry information. *J Hered* 96:718–721
- Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol Ecol Notes* 2:618–620
- Hufford MB, Xu X, van Heerwaarden J, Pyhäjärvi T, Chia JM, Cartwright RA, Elshire RJ, Glaubitz JC, Guill KE, Kaeppler SM, Lai J, Morrell PL, Shannon LM, Song C, Springer NM, Swanson-Wagner RA, Tiffin P, Wang J, Zhang G, Doebley J, McMullen MD, Ware D, Buckler ES, Yang S, Ross-Ibarra J (2012) Comparative population genomics of maize domestication and improvement. *Nat Genet* 44:808–811
- Kalinowski ST (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Mol Ecol Notes* 5:187–189
- Kruskal WH, Wallis WA (1952) Use of ranks in one-criterion variance analysis. *J Am Stat Assoc* 47:583–621
- Lam HM, Xu X, Liu X, Chen W, Yang G, Wong FL, Li MW, He W, Qin N, Wang B, Li J, Jian M, Wang J, Shao G, Wang J, Sun SS, Zhang G (2010) Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat Genet* 42:1053–1059
- Lin CS, Poushinsky G (1985) A modified augmented design (type 2) for rectangular plots. *Can J Plant Sci* 65:743–749
- Lin Z, Li X, Shannon LM, Yeh CT, Wang ML, Bai G, Peng Z, Li J, Trick HN, Clemente TE, Doebley J, Schnable PS, Tustinstra MR, Tesso TT, White F, Yu J (2012) Parallel domestication of the *Shattering1* genes in cereals. *Nat Genet* 44:720–724
- Liu A, Burke JM (2006) Patterns of nucleotide diversity in wild and cultivated sunflowers. *Genetics* 173:321–330
- Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21:2128–2129
- Liu S, Zhou R, Dong Y, Li P, Jia J (2006) Development, utilization of introgression lines using synthetic wheat as donor. *Theor Appl Genet* 112:1360–1373
- Liu W, Kim MY, Van K, Lee YH, Li H, Liu X, Lee SH (2011) QTL identification of yield-related traits and their association with flowering and maturity in soybean. *J Crop Sci Biotech* 14:65–70
- Maggioni L, Pavelek M, van Soest LJM, Lipman E (2002) Flax Genetic Resources in Europe. Ad hoc meeting: 7–8 December 2001, Prague, Czech Republic. International Plant Genetic Resources Institute, Rome, Italy
- Matus I, Corey A, Filichkin T, Hayes PM, Vales MI, Kling J, Riera-Lizarazu O, Sato K, Powell W, Waugh R (2003) Development and characterization of recombinant chromosome substitution lines (RCSLs) using *Hordeum vulgare* subsp. *spontaneum* as a source of donor alleles in a *Hordeum vulgare* subsp. *vulgare* background. *Genome* 46:1010–1023
- McCouch S, Baute GJ, Bradeen J, Bramel P, Bretting PK, Buckler E, Burke JM, Charest D, Cloutier S, Cole G, Dempewolf H, Dingkuhn M, Feuillet C, Gepts P, Grattapaglia D, Guarino L, Jackson S, Knapp S, Langridge P, Lawton-Rauh A, Lijua Q, Lusty C, Michael T, Myles S, Naito K, Nelson RL, Pontarollo R, Richards CM, Rieseberg L, Ross-Ibarra J, Rounsley S, Hamilton RS, Schurr U, Stein N, Tomooka N, van der Knaap E, van Tassel D, Toll J, Valls J, Varshney RK, Ward J, Waugh R, Wenzl P, Zamir D (2013) Agriculture: feeding the future. *Nature* 499:23–24
- Morrell PL, Toleno DM, Lundy KE, Clegg MT (2005) Low levels of linkage disequilibrium in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite high rates of self-fertilization. *Proc Natl Acad Sci USA* 102:2442–2447
- Morrell PL, Buckler ES, Ross-Ibarra J (2012) Crop genomics: advances and applications. *Nat Rev Genet* 13:85–96

- Nei M (1973) Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA* 70:3321–3323
- Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York
- Peakall R, Smouse P (2012) GenAIEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28:2537–2539
- Peng B, Li Y, Wang Y, Liu C, Liu Z, Tan W, Zhang Y, Wang D, Shi Y, Sun B, Song Y, Wang T, Li Y (2011) QTL analysis for yield components and kernel-related traits in maize across multi-environments. *Theor Appl Genet* 122:1305–1320
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38:904–909
- Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000) Association mapping in structured populations. *Am J Hum Genet* 67:170–181
- Ragupathy R, Rathinavelu R, Cloutier S (2011) Physical mapping and BAC-end sequence analysis provide initial insights into the flax (*Linum usitatissimum* L.) genome. *BMC Genom* 12:217
- Rajwade AV, Arora RS, Kadoo NY, Harsulkar AM, Ghorpade PB, Gupta VS (2010) Relatedness of Indian flax genotypes (*Linum usitatissimum* L.): an inter-simple sequence repeat (ISSR) primer assay. *Mol Biotechnol* 45:161–170
- Ranc N, Muñoz S, Xu J, Le Paslier MC, Chauveau A, Bounon R, Rolland S, Bouchet JP, Brunel D, Causse M (2012) Genome-wide association mapping in tomato (*Solanum lycopersicum*) is possible using genome admixture of *Solanum lycopersicum* var. *cerasiforme*. *G3* 2:853–864
- Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Mol Ecol Notes* 4:137–138
- Roy KJ, Smith KP, Muehlbauer GJ, Chao S, Close TJ, Steffenson BJ (2010) Association mapping of spot blotch resistance in wild barley. *Mol Breed* 26:243–256
- Russell J, Dawson IK, Flavell AJ, Steffenson B, Weltzien E, Booth A, Ceccarelli S, Grando S, Waugh R (2011) Analysis of >1000 single nucleotide polymorphisms in geographically matched samples of landrace and wild barley indicates secondary contact and chromosome-level differences in diversity around domestication genes. *New Phytol* 191:564–578
- Schmalenbach I, Léon J, Pillen K (2009) Identification and verification of QTLs for agronomic traits using wild barley introgression lines. *Theor Appl Genet* 118:483–497
- Shapiro SS, Wilk MB (1965) An analysis of variance test for normality (complete samples). *Biometrika* 52:591–611
- Sim SC, Robbins MD, Van Deynze A, Michel AP, Francis DM (2011) Population structure and genetic differentiation associated with breeding history and selection in tomato (*Solanum lycopersicum* L.). *Heredity* 106:927–935
- Smykal P, Bačová-Kertesová N, Kalendar R, Corander J, Schulman AH, Pavelek M (2011) Genetic diversity of cultivated flax (*Linum usitatissimum* L.) germplasm assessed by retrotransposon-based markers. *Theor Appl Genet* 122:1385–1397
- Soto-Cerda BJ, Urbina Saavedra H, Navarro Navarro C, Mora Ortega P (2011) Characterization of novel genic SSR markers in *Linum usitatissimum* (L.) and their transferability across eleven *Linum* species. *Electron J Biotechnol*. doi:10.2225/vol14-issue2-fulltext-6
- Soto-Cerda BJ, Maureira-Butler I, Muñoz G, Rupayan A, Cloutier S (2012) SSR-based population structure, molecular diversity and linkage disequilibrium analysis of a collection of flax (*Linum usitatissimum* L.) varying for mucilage seed-coat content. *Mol Breed* 30:875–888
- Soto-Cerda BJ, Diederichsen A, Ragupathy R, Cloutier S (2013) 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 Biol* 13:78
- Soto-Cerda BJ, Duguid S, Booker H, Rowland G, Cloutier S (2014) Genomic regions underlying agronomic traits in linseed (*Linum usitatissimum* L.) as revealed by association mapping. *J Integr Plant Biol* 56:75–87
- Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* 100:9440–9445
- Swamy BPM, Sarla N (2008) Yield-enhancing quantitative trait loci (QTLs) from wild species. *Biotechnol Adv* 26:106–120
- Tammes T (1928) The genetics of the genus *Linum*. *Bibliogr Genet* 4:1–36
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28:2731–2739
- Tanksley SD (1993) Mapping polygenes. *Annu Rev Genet* 27:205–233
- Tanksley SD, Nelson JC (1996) Advanced backcross QTL analysis: a method for simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *Theor Appl Genet* 92:191–203
- Uysal H, Fu YB, Kurt O, Peterson GW, Diederichsen A, Kusters P (2010) Genetic diversity of cultivated flax (*Linum usitatissimum* L.) and its wild progenitor pale flax (*Linum bienne* Mill.) as revealed by ISSR markers. *Genet Resour Crop Evol* 57:1109–1119
- Uysal H, Kurt O, Fu YB, Diederichsen A, Kusters P (2012) Variation in phenotypic characters of pale flax (*Linum bienne* Mill.) from Turkey. *Genet Resour Crop Evol* 59:19–30
- van Berloo R (2008) GGT 2.0: versatile software for visualization and analysis of genetic data. *J Hered* 99:232–236
- Weber AL, Zhao Q, McMullen MD, Doebley JF (2009) Using association mapping in teosinte to investigate the function of maize selection-candidate genes. *PLoS ONE* 4:e8227
- Wright SI, Gaut BS (2005) Molecular population genetics and the search for adaptive evolution in plants. *Mol Biol Evol* 22:506–519
- Würschum T (2012) Mapping QTL for agronomic traits in breeding populations. *Theor Appl Genet* 125:201–210
- Xiao J, Li J, Grandillo S, Ahn SN, Yuan L, Tanksley SD, McCouch SR (1998) Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. *Genetics* 150:899–909
- Xu J, Ranc N, Muñoz S, Rolland S, Bouchet JP, Desplant N, Le Paslier MC, Liang Y, Brunel D, Causse M (2013) Phenotypic diversity and association mapping for fruit quality traits in cultivated tomato and related species. *Theor Appl Genet* 126:567–581



- Xue W, Xing Y, Weng X, Zhao Y, Tang W, Wang L, Zhou H, Yu S, Xu C, Li X, Zhang Q (2008) Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice. *Nat Genet* 143:1–7
- You FM, Duguid SD, Thambugala D, Cloutier S (2013) Statistical analysis and field evaluation of the type 2 modified augmented design (MAD) in phenotyping of flax (*Linum usitatissimum*) germplasms in multiple environments. *Aust J Crop Sci* 7:1789–1800
- Yu J, Pressoir G, Briggs W, Vroh Bi I, Yamasaki M, Doebley J, McMullen M, Gaut B, Nielsen D, Holland J, Kresovich S, Buckler E (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet* 38:203–208
- Zamir D (2001) Improving plant breeding with exotic genetic libraries. *Nat Rev Genet* 2:983–989